AGT

Algebraic & Geometric Topology

Volume 25 (2025)

Issue 6 (pages 3145–3787)

msp

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	Thomas Koberda	University of Virginia thomas.koberda@virginia.edu
Steven Boyer	Université du Québec à Montréal cohf@math.rochester.edu	Markus Land	LMU München markus.land@math.lmu.de
Tara E Brendle	University of Glasgow tara.brendle@glasgow.ac.uk	Christine Lescop	Université Joseph Fourier lescop@ujf-grenoble.fr
Indira Chatterji	CNRS & Univ. Côte d'Azur (Nice) indira.chatterji@math.cnrs.fr	Norihiko Minami	Yamato University minami.norihiko@yamato-u.ac.jp
Octav Cornea	Université' de Montreal cornea@dms.umontreal.ca	Andrés Navas	Universidad de Santiago de Chile andres.navas@usach.cl
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
Matthew Hedden	Michigan State University mhedden@math.msu.edu	Zoltán Szabó	Princeton University szabo@math.princeton.edu
Kristen Hendricks	Rutgers University kristen.hendricks@rutgers.edu	Maggy Tomova	University of Iowa maggy-tomova@uiowa.edu
Hans-Werner Henn	Université Louis Pasteur henn@math.u-strasbg.fr	Daniel T Wise	McGill University, Canada daniel.wise@mcgill.ca
Daniel Isaksen	Wayne State University isaksen@math.wayne.edu	Lior Yanovski	Hebrew University of Jerusalem lior.yanovski@gmail.com

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

The subscription price for 2025 is US $760/year for the electronic version, and $1110/year (+$75, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Algebraic & Geometric Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

https://msp.org/

© 2025 Mathematical Sciences Publishers

msp

Algebraic & Geometric Topology **25**:6 (2025) 3145–3224
DOI: 10.2140/agt.2025.25.3145
Published: 1 October 2025

Holomorphic polygons and the bordered Heegaard Floer homology of link complements

THOMAS HOCKENHULL

We describe the construction of an $\mathcal{A}_\infty$ multimodule in terms of counts of holomorphic polygons in a series of Heegaard multidiagrams. We show that this is quasi-isomorphic to the type-A bordered-sutured invariant of a link complement with a view to calculating, in the sequel, these invariants in terms of the link Floer homology of the corresponding link.

57M27; 57M25

1 Introduction

Suppose that $L \subset S^3$ is a link of h components, and Λ is a choice of framing for L. The purpose of this paper is to introduce an $\mathcal{A}_\infty$ multimodule associated to (L, Λ) over $\mathcal{T}^h$, where $\mathcal{T}$ is the algebra associated to a parameterised torus, from Lipshitz, Ozsváth and Thurston [14]: we call this $\mathrm{Poly}(L, \Lambda)$. The $\mathcal{A}_\infty$ quasi-isomorphism type of $\mathrm{Poly}(L, \Lambda)$ is an invariant of the pair (L, Λ), or, indeed, the three-manifold with parameterised boundary $(S^3 - L, \Lambda)$ — and is closely related to the bordered-sutured Heegaard Floer invariants of this three-manifold.

$\mathrm{Poly}(L, \Lambda)$ is defined in terms of *holomorphic polygons*, and is closely related to the invariants $\widehat{\mathrm{CF}}$ for closed three-manifolds, as defined by Ozsváth and Szabó [18]. At its root, the chain complex $\widehat{\mathrm{CF}}(Y)$ is defined as a chain complex arising from counting holomorphic discs in a symplectic manifold associated to a *Heegaard diagram* corresponding to Y; that is, $\widehat{\mathrm{CF}}(Y)$ is $\widehat{\mathrm{CF}}(\mathcal{D})$ for some Heegaard diagram $\mathcal{D}$ representing Y. More generally, one may consider a *Heegaard n-diagram* for any $n \geq 2$, and count corresponding n-gons in an associated symplectic manifold. A Heegaard n-diagram corresponds with a sequence $Y^{0,1}, Y^{1,2}, \dots, Y^{n-2,n-1}, Y^{n-1,0}$ of three-manifolds, and counting holomorphic n-gons yields maps

$$m_n \colon \widehat{\mathrm{CF}}(Y^{0,1}) \otimes \cdots \otimes \widehat{\mathrm{CF}}(Y^{n-2,n-1}) \to \widehat{\mathrm{CF}}(Y^{n-1,0}).$$

These are well known to satisfy an $\mathcal{A}_\infty$ relation, which is shown by counting the number of ends of one-dimensional moduli spaces and showing that they are of even parity. These moduli spaces decompose by *homology classes*, and, in fact, the parity of the ends of a one-dimensional moduli space remains even when restricted to curves of a given homology class. We exploit this extra information to define $\mathrm{Poly}(L, \Lambda)$.

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Broadly speaking, this multimodule is defined by associating a series of h sequences of elements of the torus algebra $\mathcal{T}$ to each fixed homology class, in such a way that the $\mathcal{A}_\infty$ relations for each homology class are compatible with the algebra — yielding an $\mathcal{A}_\infty$ multimodule over the algebra $\mathcal{T}^h$.

The main theorem we show is:

Theorem 1.1 *The invariant* $\mathrm{Poly}(L,\Lambda)$ *is quasi-isomorphic to the bordered-sutured multimodule* $\mathrm{BSA}(S^3-L,\Lambda)$.

The drawback of the invariants BSA is that they are often rather tricky to compute (see eg Hanselman [6], Levine [9], and Lipshitz, Ozsváth and Thurston [11; 14]). The module $\mathrm{Poly}(L,\Lambda)$, whilst somewhat involved in its definition, is built from objects which have in one guise or other been present from the advent of Heegaard Floer homology for closed three-manifolds. There is a close relationship between the types of polygon maps used to build $\mathrm{Poly}(L,\Lambda)$ and the invariants $\mathrm{CFL}^-(L)$ — see, for instance, Manolescu and Ozsváth [15], Ozsváth and Szabó [17; 21; 22], and Rasmussen [23], or see Section 1.1 below.

Conceptually, our approach is perhaps advantageous to that of [14] in that — besides generalising directly to the case of general link complements — its connections with preexisting gluing type results in closed Heegaard Floer homology are clear; see the discussion in the next section.

1.1 Motivation

The invariant $\mathrm{Poly}(L,\Lambda)$ is rather daunting at first, but has many precursors in Heegaard Floer homology that make it a natural object to study. To orient the reader, we touch upon some of these here — but leave the explicit details to future work.

1.1.1 The mapping cone theorem By the pairing theorems of Lipshitz, Ozsváth and Thurston [14] and Zarev [26], for some choice of framing $\Lambda=(\lambda_1,\dots,\lambda_{|L|})$ the invariant $\mathrm{CFA}(E(L),\Lambda)$ determines the invariant $\widehat{\mathrm{CF}}(S^3_{\Lambda'}(L))\otimes\mathbb{F}_2^{\otimes|L|-1}$ for any choice of framing $\Lambda'=(\lambda'_1,\dots,\lambda'_{|L|})$. More specifically, beginning with the invariant $\mathrm{CFA}(E(L),\Lambda)$, one takes a series of box tensor products with type D invariants of appropriately framed solid tori, such that the sutured three-manifold formed by gluing these solid tori to $E(L)$ according to the framing Λ and the framings on the solid tori is the sutured manifold $S^3_{\Lambda'}$ less $|L|-1$ spheres with standard meridional sutures.

Historically, the invariants $\widehat{\mathrm{CF}}(S^3_{\Lambda'})$ have been computed using a sort of proto-gluing theorem, the *mapping cone theorem* — see [21] for the case of $L=K$ a knot, and [15] for the more general case where L is a general link. This takes the input data of the associated invariants $\mathrm{CFL}^-(L)$, together with some "flip" maps (see [15] for more detail of the precise data required), and writes down the invariants $\widehat{\mathrm{CF}}(S^3_{\Lambda'})$ in terms of a complex built as a mapping cone of associated complexes.

We thus have two ways to compute the same quantity. The optimistic question is: to what extent do the quantities $\mathrm{CFA}(E(L),\Lambda)$ and $\mathrm{CFL}^-(L)$ (together with the "flip" maps) determine one another? More heuristically, does the recipe for $\widehat{\mathrm{CF}}(S^3_{\Lambda'})$ in terms of $\mathrm{CFL}^-(L)$ factor through $\mathrm{CFA}(E(L),\Lambda)$?

The case where $L=K$ is a knot is the most well understood, and, indeed, it turns out not so optimistic:

Theorem 1.2 [14, Chapter 11] *The invariant* $\mathrm{CFK}^-(K)$ *together with the extra information of the "flip" map determines the invariant* $\mathrm{CFD}(E(K),n)$ *for any framing* n.

By taking a box tensor product with the invariant $\mathrm{CFAA}(\mathbb{I})$, this gives the desired relationship.

The results of this paper give a concrete reason to expect this to generalise to links. In particular, the proof of the mapping cone theorem hinges upon a detailed understanding of the relationship between $\mathrm{CFL}^-(L)$ and cobordism maps between three-manifolds $S^3_{\Lambda_0}$ and $S^3_{\Lambda_1}$, for $\Lambda_0\subset\Lambda_1\subset\Lambda$. These cobordism maps are defined in terms of the holomorphic polygons used in the definition of the invariant $\mathrm{Poly}(L,\Lambda)$ in this paper; thus we have a first step towards associating $\mathrm{CFL}^-(L)$ and $\mathrm{Poly}(L,\Lambda)\cong\mathrm{CFA}(E(L),\Lambda)$.

The invariant $\mathrm{Poly}(L,\Lambda)$, however, tracks extra information about these holomorphic polygons, in the form of an associated set of sequences of algebra elements, which are unseen by the corresponding cobordism maps and higher polygon counting maps in the classic set-up. It transpires, however, that much of this extra information can be packaged in terms of gradings on the chain complexes involved. More about this, and the relationship between the invariants $\mathrm{CFL}^-(L)$, "flip maps", and bordered Heegaard Floer homology is to be made explicit in [7].

1.1.2 The spectral sequence of the branched double-cover, and an alternative proof In [19, Section 4.2], a chain complex is associated to a framed h-component link. As a set, this complex is given by $X=\bigoplus_{I\in\{m,l\}^h}\widehat{\mathrm{CF}}(Y(I))$, where $Y(I)$ is the manifold formed by surgering Y along L according to the framing induced by Λ and I. The differential D counts a sum of polygon maps that respect a fixed ordering of the vertices of I^h.

The invariant $\mathrm{Poly}(L,\Lambda)$ is defined in a very similar manner, but with two main differences. First, polygon maps in $\mathrm{Poly}(L,\Lambda)$ do not obey the ordering of the vertices: instead, we must consider all possible polygon maps — ie the maps in $\mathrm{Poly}(L,\Lambda)$ split as sets of maps, where each set respects a fixed ordering of the vertices of I^h. The other clear difference is that $\mathrm{Poly}(L,\Lambda)$ is an $\mathcal{A}_\infty$-module, whereas $X(L,\Lambda)$ is a chain complex. This can be reconciled by considering instead the differential graded module $(P,\partial):=\mathrm{Poly}(L,\Lambda)\otimes\mathcal{T}^*$ (here, we write $\mathcal{T}^*$ to denote the tensor algebra associated to the torus algebra; see [14, page 10]). Both P and $X(L,\Lambda)$ split as a direct sum over idempotents of the torus algebra, and the provision of an extra basepoint for each component of L in the corresponding Heegaard multidiagram gives a filtration upon the differential of $X(L,\Lambda)$, which corresponds with the splitting of ∂ according to corresponding sequences of algebra elements. When L is a knot, this should be compared to Eftekhary [5].

More recently, the complex $X(L,\Lambda)$ has been studied from the perspective of bordered Heegaard Floer homology, to understand the difference between two spectral sequences between the reduced Khovanov homology of the mirror of a link L' and the Heegaard Floer homology $\widehat{\mathrm{HF}}(\Sigma(L'))$, where $\Sigma(L')$ is the branched double cover of L': the one given by Ozsváth and Szabó [19], and another defined in terms of bordered Heegaard Floer homology by Lipshitz, Ozsváth and Thurston [13]. This relationship follows naturally from Theorem 1.1, but, perhaps more interestingly, the methods of [13] can be applied, together with a model computation, to give a different proof of Theorem 1.1 to the one in the present paper.

We sketch the main idea. The main result of [13] is a pairing theorem for holomorphic polygons in *bordered Heegaard multidiagrams* (see Section 4.6 below). Approximately, this states that if $\mathcal{D}_1$ and $\mathcal{D}_2$ are bordered Heegaard k- and l-diagrams respectively (ie $\mathcal{D}_1$ has $\boldsymbol{\beta}^1,\dots,\boldsymbol{\beta}^k$ sets of attaching circles and one set $\boldsymbol{\alpha}$ of attaching arcs and circles, and similar for $\mathcal{D}_2$), a pair of holomorphic polygons u_1 and u_2 with appropriate boundary conditions (encoded by elements of the torus algebra) corresponds to a holomorphic polygon in a diagram $\mathcal{D}$. Here, $\mathcal{D}$ corresponds roughly with the diagram formed by gluing $\mathcal{D}_1$ and $\mathcal{D}_2$ together along a common boundary; but in order to make sense of attaching circles, it is often necessary to add a series of perturbed "copies" of some $\boldsymbol{\beta}^i$ to $\mathcal{D}_1$ or $\mathcal{D}_2$, to promote $\mathcal{D}_1$ to a bordered Heegaard l-diagram (if $k<l$, say).

Consider a cube of framings with vertices given by elements $I\in\{m,l\}^2$. To this we may associate a type D structure given by the sum of the invariants $\mathrm{CFD}(\mathcal{D}(I))$, where each $\mathcal{D}(I)$ is a bordered Heegaard multidiagram for the torus, with attaching circles specified by I. We may promote this type D structure to a type DA-bimodule by assigning to each domain B for $\mathcal{D}(I)$ a sequence of elements of the torus algebra $\vec{\rho}(B)$, and defining $m_{1,k}(\boldsymbol{x},\vec{\rho})$ to be a map counting holomorphic curves in homology class B with $\vec{\rho}(B)=\vec{\rho}$, such that summing these $m_{1,k}(\boldsymbol{x},\vec{\rho})$ over all sequences of $\vec{\rho}$ gives the structure map δ^1 for the invariant $\mathrm{CFD}(\mathcal{D}(I))$. By direct calculation, one can see that this DA-bimodule is equivalent to the identity DA-bimodule for the torus, $\mathrm{CFDA}(\mathbb{I})$. As such,

$$\mathrm{CFA}(E(L))\cong\mathrm{CFA}(E(L))\boxtimes\mathrm{CFDA}(\mathbb{I}),$$

but, by the pairing theorem and the above discussion, we know that this is equal to the sum

$$\bigoplus_I \mathrm{CFA}(E(L))\boxtimes\mathrm{CFD}(\mathcal{D}(I)),$$

which is a sum of polygon maps in a bordered Heegaard multidiagram with $|L|-1$ boundary components, splitting by the algebra element associated to the restriction of their domain to the portion $\mathcal{D}(I)$. Iterating this for all boundary components of $E(L)$ gives an identification of $\mathrm{CFA}(E(L))$ with a module defined in terms of polygon maps in closed Heegaard diagrams, with extra algebra information stemming from how the associated domain behaves in the regions corresponding to $\mathcal{D}(I)$ — this is precisely the way the object $\mathrm{Poly}(L,\Lambda)$ is defined.

This should be compared with the results of Auroux [2, Section 6.2]; in particular, the sketched argument above is approximately equivalent to the sketch proof of Proposition 6.5 given in that paper, but recast in

a setting and language that is perhaps more accessible to the reader with background in Heegaard Floer homology, rather than more broad symplectic geometry.[1]

1.2 Organisation

In Section 2 we recall the algebraic definitions of $\mathcal{A}_\infty$ multimodules necessary for the rest of the paper, and some relevant constructions and properties of the torus algebra $\mathcal{T}$. Then, in Section 3, we consider the close relationship between bordered-sutured Heegaard diagrams which represent the complement of a link, and a corresponding family of "partially splayed" and "splayed" Heegaard diagrams. The invariant $\mathrm{Poly}(L,\Lambda)$ is defined in terms of the splayed diagrams, which are a family of closed Heegaard multidiagrams — but we will need to consider a series of auxiliary modules $\mathrm{Poly}_k(L,\Lambda)$ defined in terms of partially splayed diagrams, which are bordered-sutured k-multidiagrams. These correspond roughly with the final k algebra inputs into a structure map m_n being contributed not by asymptotics at a puncture in the Heegaard diagram, but instead by algebra elements associated to the homology class of a holomorphic curve. In particular, when k is zero, all algebra elements are contributed by asymptotics at a puncture, and so these recover the bordered sutured Floer homology of the link complement; when k is large enough, these modules stabilise and there is no interesting behaviour near the puncture — so these modules may be identified with $\mathrm{Poly}(L,\Lambda)$.

In Section 4, we define a series of moduli spaces of holomorphic polygons which will be instrumental in our definition of the invariants $\mathrm{Poly}(L,\Lambda)$, and the auxiliary modules $\mathrm{Poly}_k(L,\Lambda)$. We carefully analyse the possible boundary of these moduli spaces, thus ensuring that the $\mathcal{A}_\infty$ relations hold for these multimodules. We also lay the foundations of a neck-stretching argument which will allow us, in the final section, to identify all of the modules $\mathrm{Poly}_k(L,\Lambda)$ for each k, thus proving Theorem 1.1.

Acknowledgements

I owe the Heegaard Floer community at large a debt of gratitude for their support and encouragement. In particular, I would like to thank Marco Marengon for many useful and interesting conversations. I am also grateful to Peter Ozsváth for hosting me at Princeton, and the helpful conversations that took place there. I am also grateful to Tom Coates and András Stipsicz for their comments and corrections on the version of this paper which constituted my PhD thesis, along with the referee for such a careful reading and their corrections and comments.

Robert Lipshitz has gone far and beyond in his encouragement and support, and in his tirelessness in answering my questions about the "LOT" oeuvre — he and the University of Oregon were also very gracious hosts to me in the spring of 2016. Richard Thomas has been invaluable in his support, encouragement and cajoling over the past year or so.

[1] For full transparency, I learned of Auroux's results roughly two-thirds of the way through writing this paper.

Finally, I have been constantly humbled by the generosity and kindness of my supervisor Jacob Rasmussen. Without his reassurance, encouragement and patience, I doubt any of this would have been possible.

The author was supported by an EPSRC studentship from 2013–2017, and by Imperial College London mathematics department from 2017–2018. Some of this work was undertaken whilst visiting Princeton University, which was funded by the Doris Chen award.

2 Algebraic background

We spend this section recalling the algebraic background of the bordered sutured invariants — although these are defined in [26, Section 8], the algebraic structure is not discussed at length.

2.1 The torus algebra

We first define the algebras that we will work over for most of this paper. These are the *torus algebra*, $\mathcal{T}$, and the products

$$\mathcal{T}^k := \underbrace{\mathcal{T} \times \cdots \times \mathcal{T}}_{k \text{ times}}.$$

The torus algebra is generated by eight elements as an $\mathbb{F}_2$-module, which are the two idempotents ι_m and ι_l, and the six Reeb chords ρ_1, ρ_2, ρ_3, ρ_{12}, ρ_{23} and ρ_{123}. The differential is trivial, and the idempotents satisfy $\iota_m\iota_l = \iota_l\iota_m = 0$ — we also write 1 for the element $\iota_m + \iota_l$. The Reeb chords satisfy $\rho_1\rho_2 = \rho_{12}, \rho_2\rho_3 = \rho_{23}$ and $\rho_1\rho_{23} = \rho_{12}\rho_3 = \rho_{123}$. The other nonzero products are

$$\iota_m\rho_1 = \rho_1\iota_l = \rho_1, \qquad \iota_l\rho_2 = \rho_2\iota_m = \rho_2, \qquad \iota_m\rho_3 = \rho_3\iota_l = \rho_3,$$
$$\iota_m\rho_{12} = \rho_{12}\iota_m = \rho_{12}, \quad \iota_l\rho_{23} = \rho_{23}\iota_l = \rho_{23}, \quad \iota_m\rho_{123} = \rho_{123}\iota_l = \rho_{123}.$$

These algebra elements may be interpreted in terms of *Reeb chords*. Let Z denote an oriented circle with four marked points $\boldsymbol{a} = \{a_1, \dots, a_4\} \subset Z$, together with a basepoint $b \in Z \setminus \boldsymbol{a}$ — we suppose that $\boldsymbol{a}$ is numbered in such a way that a positive path from b encounters them in order. A *Reeb chord* $\rho \in (Z \setminus b, \boldsymbol{a})$ is an immersed arc in $Z \setminus b$ with endpoints in $\boldsymbol{a}$, and orientation agreeing with that on Z. To each Reeb chord ρ we associate the algebra element $a(\rho) := \rho_{[i,j]}$, where $\partial\rho = a_j - a_i$ and $[i, j]$ is the unique substring of 123 beginning with i and ending with $j - 1$. We will often conflate a Reeb chord with its corresponding algebra element.

We say that two Reeb chords ρ and ρ' are *composable* if the product $\rho\rho'$, which we call their *composition*, is nonzero. More generally, we call sequence $(\rho^1, \dots, \rho^k)$ of Reeb chords *composable* if for every $i = 1, \dots, k-1$, the chords ρ^i and ρ^{i+1} are composable. If $\vec{\rho}$ is a sequence of Reeb chords, then there is an associated sequence of Reeb chords $\vec{\rho}'$ formed by replacing any composable subsequences of Reeb chords with their composition, which we call the *reduction* of $\vec{\rho}$. There is an equivalence relation upon sequences of Reeb chords, given by saying that $\vec{\rho}$ and $\vec{\rho}'$ are *composable-equivalent* if and only if their reductions agree.

In a similar manner, we say that two chords, ρ and ρ', are 123-*composable* if they are composable and their composition is ρ_{123}, and make a similar definition for the 123-*reduction* of a sequence of Reeb chords. We shall say that a pair of sequences of Reeb chords are 123-*equivalent* if and only if their 123-reductions agree.

Remark 2.1 Being 123-composable depends only upon *pairs* of adjacent chords. In particular, the sequence (ρ_1, ρ_2, ρ_3) is 123-reduced, whereas the sequence (ρ_1, ρ_{23}) has 123-reduction ρ_{123}.

One can see from the definition of the torus algebra that some Reeb chords behave rather differently to others with respect to products by the elementary idempotents. If $\iota_\delta \cdot \rho \neq 0$ we will say that $\iota_{\text{start}}(\rho) = \iota_\delta$, and if $\rho \cdot \iota_\delta \neq 0$ we will say that $\iota_{\text{end}}(\rho) = \iota_\delta$.

It is straightforward to see that a for a sequence of Reeb chords $\vec{\rho} = (\rho^1, \dots, \rho^m)$, the tensor product $\rho^1 \otimes \cdots \otimes \rho^m$ over the ring of idempotents of $\mathcal{T}$ is nonzero if and only if $\iota_{\text{end}}(\rho^i) = \iota_{\text{start}}(\rho^{i+1})$. Accordingly, we make the convention throughout this paper that a *sequence of Reeb chords* satisfies this condition — ie that it is shorthand for an element of the tensor product $\mathcal{T}^{\otimes m}$.

Definition 2.2 Suppose that $\vec{\rho} = (\rho^1, \dots, \rho^m)$ is a sequence of elements of the torus algebra. To the sequence $\vec{\rho}$ we associate the sequence of its *jumping chords*. That is, the subsequence

$$\text{jump}(\vec{\rho}) := (\rho^i \in \vec{\rho} \mid \iota_{\text{start}}(\rho^i) \neq \iota_{\text{end}}(\rho^i))$$

of all chords in the sequence that have different starting to ending idempotent.

Example 2.3 For instance, the sequence $\vec{\rho} = (\rho_2, \rho_{12}, \dots, \rho_{12}, \rho_1)$ has $\text{jump}(\vec{\rho}) = (\rho_2, \rho_1)$.

We turn our attention to what we term a *splicing* of a series of Reeb chords. These should be thought of as something between a set of elements $\vec{b}^1, \dots, \vec{b}^k$ where $\vec{b}^i \in \mathcal{T}^{\otimes h}$, and a set of sequences of elements $\vec{a}_1, \dots, \vec{a}_h$ where each $\vec{a}_j \in \mathcal{T}^{k_j}$. The point is that, later, we will make constructions (of *diagrams*) which correspond with a given set of sequences of Reeb chords: these depend upon the corresponding set of sequences of jumping chords (and ignore any nonjumping chords), and manifestly depend upon some overall ordering of the jumping chords from all of the sequences — a splicing is a way to specify this overall ordering.

Definition 2.4 Suppose that $\vec{\rho}_1, \dots, \vec{\rho}_h$ are a series of sequences of Reeb chords $\vec{\rho}_j := (\rho_j^1, \dots, \rho_j^{i_j})$ with corresponding jumping sequences $\text{jump}(\vec{\rho}_1), \dots, \text{jump}(\vec{\rho}_j)$.

An *m-splicing* of $\vec{\rho}_1, \dots, \vec{\rho}_h$ is an ordered length m partition $P_j := (P_j^1, \dots, P_j^m)$ of each $\text{jump}(\vec{\rho}_j)$, allowing for parts P_j^i to be empty, and such that

- $|P_j^i| \leq 1$ for every i and j, and
- for each i, the union $P_1^i \cup \cdots \cup P_h^i$ is nonempty.

Every splicing $\vec{\sigma}$ has an underlying set of unordered partitions, which we shall often denote by σ. We often write $|\vec{\sigma}| = m$ to indicate that $\vec{\sigma}$ is an m-splicing. We also write $\mathrm{Col}(\vec{\sigma})$ for the number of elements in the set

$$\{i \mid P_j^i \text{ is nonempty for more than one } j\}.$$

We shall say that $\vec{\sigma}$ is an *interleaving* of $\vec{\rho}_1, \dots, \vec{\rho}_h$, or that $\vec{\sigma}$ is *interleaved* if $\mathrm{Col}(\vec{\sigma}) = 0$; that is, for every i, precisely one P_j^i is nonempty.

Example 2.5 For instance, consider $\vec{\rho}_1 = (\rho_1, \rho_{23}, \rho_2)$ and $\vec{\rho}_2 = (\rho_{12}, \rho_1)$. A noninterleaved 2-splicing of these is given by $P_1 = (\rho_1, \rho_2)$ and $P_2 = (\rho_1, \varnothing)$; an interleaved 3-splicing $\vec{\sigma}_1$ is given by $P_1 = (\rho_1, \varnothing, \rho_2)$ and $P_2 = (\varnothing, \rho_1, \varnothing)$.

If we are given two sets of sequences of Reeb chords $\vec{\lambda} = \vec{\lambda}_1, \dots, \vec{\lambda}_h$ and $\vec{\delta} = \vec{\delta}_1, \dots, \vec{\delta}_h$, where each $\vec{\lambda}_j$ is of length l_j and each $\vec{\delta}_j$ is of length k_j, there is a corresponding set of sequences of Reeb chords $\vec{\lambda} \star \vec{\delta}$ defined by $(\vec{\lambda} \star \vec{\delta})_j := (\vec{\lambda}_j, \vec{\delta}_j)$. Similarly, given splicings $\vec{\sigma}_1 = P_1, \dots, P_h$ and $\vec{\sigma}_2 = Q_1 \cdots Q_h$ for $\vec{\rho}$ and $\vec{\delta}$ respectively, then there is a corresponding splicing $\vec{\sigma}_1 \star \vec{\sigma}_2$ of $\vec{\rho} \star \vec{\delta}$, formed by $(\vec{\sigma}_1 \star \vec{\sigma}_2) = (P_1, Q_1), \dots, (P_h, Q_h)$.

Example 2.6 With $\vec{\rho} = \vec{\rho}_1, \vec{\rho}_2$ as in Example 2.5, and $\vec{\lambda} = \vec{\lambda}_1, \vec{\lambda}_2$ for $\vec{\lambda}_1 = (\rho_{23})$ and $\vec{\lambda}_2 = (\rho_2)$, with corresponding splicing $\vec{\sigma}_2$ given by $Q_1 = (\varnothing)$ and $Q_2 = (\rho_2)$, we have that $(\vec{\rho} \star \vec{\lambda})_1 = (\rho_1, \rho_{23}, \rho_2, \rho_{23})$ and $(\vec{\rho} \star \vec{\lambda})_2 = (\rho_{12}, \rho_1, \rho_2)$. The corresponding splicing is $(\vec{\sigma}_1 \star \vec{\sigma}_2)$, given by the partitions $(\rho_1, \varnothing, \rho_2, \varnothing)$ and $(\varnothing, \rho_1, \varnothing, \rho_2)$.

A splicing naturally gives rise to a sequence of sets of idempotents in the torus algebra:

Definition 2.7 Let $\vec{\rho} := \vec{\rho}_1, \dots, \vec{\rho}_h$ be a series of sequences of algebra elements $\vec{\rho}_j$, and let $\vec{\sigma} = P_1, \dots, P_h$ be an m-splicing for $\vec{\rho}$.

For each j, let $P_j^{i_1}, \dots, P_j^{i_{|\mathrm{jump}(\vec{\rho}_j)|}}$ denote the parts of P_j which are nonempty, and let $\rho_j^{i_k}$ denote the (unique) Reeb chord in each $P_j^{i_k}$. Each $P_j^{i_k}$ has an associated sequence of parts

$$\vec{P}_j^{i_k} := (P_j^{i_k}, P_j^{i_k+1}, \dots, P_j^{i_{k+1}-1}),$$

its *trailed* sequence, where only $P_j^{i_k} \in \vec{P}_j^{i_k}$ is nonempty. The sequence P_j decomposes as a union

$$P_j = \bigcup_{k=0}^{|\mathrm{jump}(\vec{\rho}_j)|} \vec{P}^{i_k},$$

where we define $\vec{P}_j^{i_0}$ to be the union of the empty partitions $(P_j^1, \dots, P_j^{i_1-1})$.

We define a set of idempotents by

$$\iota_j^i := \begin{cases} \iota_{\mathrm{start}}(\rho_j^{i_1}) & \text{if } P_j^i \in \vec{P}_j^{i_0} \text{ or } i = 0, \\ \iota_{\mathrm{end}}(\rho_j^{i_k}) & \text{if } P_j^i \in \vec{P}_j^{i_k}, k \neq 0. \end{cases}$$

We collect these into sequences of idempotents $\vec{\iota}^0, \dots, \vec{\iota}^m$ by $\vec{\iota}^i = \iota^i_1, \dots, \iota^i_h$, and denote by $\iota_{\vec{\sigma}} := (\vec{\iota}^0, \dots, \vec{\iota}^m)$.

Example 2.8 With $\vec{\rho} = \vec{\rho}_1, \vec{\rho}_2$ and $\vec{\sigma}_1$ as in Example 2.5, we have corresponding sequences of idempotents $\vec{\iota}^0 = (\iota_m, \iota_m)$, $\vec{\iota}^1 = (\iota_l, \iota_m)$, $\vec{\iota}^2 = (\iota_l, \iota_l)$ and $\vec{\iota}^3 = (\iota_m, \iota_l)$.

We can also go in the other direction, associating a splicing to a pair of a set of sequences of Reeb chords and a set of sequences of idempotents:

Definition 2.9 Suppose that we are provided with a set of sequences of Reeb chords $\vec{\rho}_1, \dots, \vec{\rho}_h$, and a sequence of sets of idempotents $\vec{\iota} = (\vec{\iota}^0, \dots, \vec{\iota}^m)$ where each $\vec{\iota}^i = \iota^i_1, \dots, \iota^i_h$ satisfies that for $j = 1, \dots, h$, the sequence of idempotents $\vec{\iota}_j := \iota^0_j, \dots, \iota^m_j$ has

$$|\{\iota^i_j \in \vec{\iota}_j \mid \iota^i_j \neq \iota^{i-1}_j\}| = |\text{jump}(\vec{\rho}_j)|,$$

and that for every $i = 1, \dots, m$ there is some j such that $\iota^i_j \neq \iota^{i-1}_j$.

There is a corresponding splicing given by setting $P^i_j = \varnothing$ when $\iota^i_j = \iota^{i-1}_j$, and for each $\iota^{i_k}_j \neq \iota^{i_k-1}_j$ we set P^i_j to be the part $\{\rho^{i_k}_j\}$, where $\rho^{i_k}_j$ is the k^{th} chord in $\text{jump}(\vec{\rho}_j)$ — we call this $\vec{\sigma}(\vec{\rho}, \vec{\iota})$.

Example 2.10 Suppose $\vec{\rho} = \vec{\rho}_1, \vec{\rho}_2$ are as in Example 2.5, and $\vec{\iota} = (\vec{\iota}^0, \dots, \vec{\iota}^m)$ as in Example 2.8. The corresponding splicing is the interleaved splicing given in Example 2.5.

We will also need to consider splitting a sequence of Reeb chords in two — roughly, in what follows, one half $\vec{\rho}_{\blacktriangle}$ corresponds with genuine asymptotic behaviour of a holomorphic curve, whereas $\vec{\rho}_{\bigstar}$ will be extracted from the topology of the homology class of the curve.

Definition 2.11 Suppose that $\vec{\rho} = \vec{\rho}_1, \dots, \vec{\rho}_h$ is a set of sequences of Reeb chords, and $\vec{\sigma}$ is an interleaved splicing of $\vec{\rho}$. We shall say that a *k-shipping* of $(\vec{\rho}, \vec{\sigma})$ is a pair $\vec{\rho}_{\blacktriangle}, \vec{\rho}_{\bigstar}$ (where $\vec{\rho}_{\blacktriangle} = \vec{\rho}_{\blacktriangle,1}, \dots, \vec{\rho}_{\blacktriangle,h}$ and $\vec{\rho}_{\bigstar} = \vec{\rho}_{\bigstar,1}, \dots, \vec{\rho}_{\bigstar,h}$) together with a pair of splicings $\vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\bigstar}$ such that

- for some j, we have that $\vec{\rho}_{\bigstar,j}$ begins with a jumping chord;
- $\vec{\sigma}_{\bigstar}$ is a k'-splicing of $\vec{\rho}_{\bigstar}$, where
$$k' = \min\{|\text{jump}(\vec{\rho}_1)| + \cdots + |\text{jump}(\vec{\rho}_h)|, k\};$$
- $(\vec{\rho}_{\blacktriangle}, \vec{\sigma}_{\blacktriangle}) \star (\vec{\rho}_{\bigstar}, \vec{\sigma}_{\bigstar}) = (\vec{\rho}, \vec{\sigma})$.

If $(\vec{\rho}_{\blacktriangle}, \vec{\sigma}_{\blacktriangle})$ and $(\vec{\rho}_{\bigstar}, \vec{\sigma}_{\bigstar})$ form a k-shipping of $(\vec{\rho}, \vec{\sigma})$, and $(\vec{\rho}'_{\blacktriangle}, \vec{\sigma}'_{\blacktriangle})$ and $(\vec{\rho}'_{\bigstar}, \vec{\rho}'_{\bigstar})$ form a $(k+1)$-shipping of $(\vec{\rho}, \vec{\sigma})$ then we shall say that the pair $(\vec{\rho}'_{\blacktriangle}, \vec{\sigma}'_{\blacktriangle})$ and $(\vec{\rho}'_{\bigstar}, \vec{\rho}'_{\bigstar})$ *splays* $(\vec{\rho}_{\blacktriangle}, \vec{\sigma}_{\blacktriangle})$ and $(\vec{\rho}_{\bigstar}, \vec{\sigma}_{\bigstar})$.

Example 2.12 Suppose $\vec{\rho} = \vec{\rho}_1, \vec{\rho}_2$ and $\vec{\sigma}_1$ are as in Example 2.5.

Let

$$\vec{\rho}_{\blacktriangle,1} = (\rho_1, \rho_{23}), \quad \vec{\rho}_{\bigstar,1} = (\rho_2),$$
$$\vec{\rho}_{\blacktriangle,2} = (\rho_{12}, \rho_1), \quad \vec{\rho}_{\bigstar,2} = \varnothing,$$

so that we have

$$\vec{\rho}_{\blacktriangle,1} \star \vec{\rho}_{\mathfrak{t},1} = (\rho_1, \rho_{23}, \rho_2) = \vec{\rho}_1, \quad \vec{\rho}_{\blacktriangle,2} \star \vec{\rho}_{\mathfrak{t},2} = \vec{\rho}_2.$$

Letting

$$\vec{\sigma}_{\blacktriangle,1} = (\rho_1, \varnothing), \quad \vec{\sigma}_{\mathfrak{t},1} = (\rho_2),$$
$$\vec{\sigma}_{\blacktriangle,2} = (\varnothing, \rho_1), \quad \vec{\sigma}_{\mathfrak{t},2} = (\varnothing),$$

we see that

$$\vec{\sigma}_{\blacktriangle,1} \star \vec{\sigma}_{\mathfrak{t},1} = (\rho_1, \varnothing, \rho_2) = P_1, \quad \vec{\sigma}_{\blacktriangle,2} \star \vec{\sigma}_{\mathfrak{t},2} = (\varnothing, \rho_1, \varnothing) = P_2.$$

As $\vec{\sigma}_{\mathfrak{t}} = (\sigma_{\mathfrak{t},1}, \sigma_{\mathfrak{t},2})$ is a 1-splicing of $(\vec{\rho}_{\mathfrak{t},1}, \vec{\rho}_{\mathfrak{t},2})$ it follows that $((\vec{\rho}_{\blacktriangle,1}, \vec{\rho}_{\blacktriangle,2}), (\vec{\sigma}_{\mathfrak{t},1}, \vec{\sigma}_{\mathfrak{t},2}))$ is a 1-shipping of $(\vec{\rho}, \vec{\sigma}_1)$.

Similarly, a 2-shipping of $(\vec{\rho}, \vec{\sigma}_1)$ is given by

$$\vec{\rho}_{\blacktriangle,1} = (\rho_1, \rho_{23}), \quad \vec{\rho}_{\mathfrak{t},1} = (\rho_2),$$
$$\vec{\rho}_{\blacktriangle,2} = (\rho_{12}), \qquad \vec{\rho}_{\mathfrak{t},2} = (\rho_1),$$

with corresponding splicings

$$\vec{\sigma}_{\blacktriangle,1} = (\rho_1), \quad \vec{\sigma}_{\mathfrak{t},1} = (\varnothing, \rho_2),$$
$$\vec{\sigma}_{\blacktriangle,2} = (\varnothing), \quad \vec{\sigma}_{\mathfrak{t},2} = (\rho_1, \varnothing).$$

Another 2-shipping is given by

$$\vec{\rho}_{\blacktriangle,1} = (\rho_1), \quad \vec{\rho}_{\mathfrak{t},1} = (\rho_{23}, \rho_2),$$
$$\vec{\rho}_{\blacktriangle,2} = (\rho_{12}), \quad \vec{\rho}_{\mathfrak{t},2} = (\rho_1),$$

with corresponding splicings

$$\vec{\sigma}_{\blacktriangle,1} = (\rho_1), \quad \vec{\sigma}_{\mathfrak{t},1} = (\varnothing, \rho_2),$$
$$\vec{\sigma}_{\blacktriangle,2} = (\varnothing), \quad \vec{\sigma}_{\mathfrak{t},2} = (\rho_1, \varnothing).$$

2.2 $\mathcal{A}_\infty$ multimodules

We begin with $\mathcal{A}_\infty$-modules as a warm-up. Let $\mathcal{A}$ be a differential graded algebra over $\mathbb{F}_2$, with differential ∂, multiplication μ and ring of idempotents $\mathcal{I} \subset \mathcal{A}$. A (right) $\mathcal{A}_\infty$ module over $\mathcal{A}$ is a module over $\mathbb{F}_2$ with right action of $\mathcal{I}$ and maps

$$m_{k+1} \colon M \otimes_{\mathcal{I}} \mathcal{A} \otimes_{\mathcal{I}} \otimes \cdots \otimes_{\mathcal{I}} \mathcal{A} \to M,$$

satisfying the $\mathcal{A}_\infty$ relation

$$0 = \sum_{i=0}^{n} m_{n-i+1}(m_{i+1}(x, a_1, \dots, a_i)a_{i+1}, \dots, a_n)$$
$$+ \sum_{i=1}^{n-1} m_n(x, a_1, \dots, \mu(a_i, a_{i+1}), a_{i+2}, \dots, a_n) + \sum_{i=1}^{n} m_{n+1}(x, a_1, \dots, \partial a_i, \dots, a_n)$$

for every $x \in M$ and $a_1, \dots, a_n \in \mathcal{A}$.

More general is the notion of an $\mathcal{A}_\infty$ *multimodule*. If $\mathcal{A}_1, \dots, \mathcal{A}_h$ are differential graded algebras over $\mathbb{F}_2$, each with corresponding ring of idempotents $\mathcal{I}_1, \dots, \mathcal{I}_h$, then a (right) $\mathcal{A}_\infty$ multimodule over $\mathcal{A}_1, \dots, \mathcal{A}_h$ is a module M over $\mathbb{F}_2$ with a right action of $\mathcal{I} := \mathcal{I}_1 \otimes \cdots \otimes \mathcal{I}_h$, and equipped with maps

$$m_{i_1,\dots,i_h} \colon M \otimes_{\mathcal{I}} \mathcal{A}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{A}_h^{\otimes i_h} \to M,$$

where $\mathcal{A}_i^{\otimes i_j}$ is the result of taking the tensor product of $\mathcal{A}_i$ with itself i_j times, over the ring $\mathcal{I}_j$.

The relations these maps must satisfy are more difficult to write down; we roughly follow Hanselman's presentation in [6, Section 2.1], with some modifications. If $\vec{a}_j := (a_j^1, \dots, a_j^{i_j}) \in \mathcal{A}_j^{\otimes i_j}$ is a sequence of elements of $\mathcal{A}_j$, then for every $0 \le k \le i_j$ we define the truncations

$$T_k(\vec{a}_j) := (a_j^1, \dots, a_j^k), \quad T^k(\vec{a}_j) := (a_j^k, \dots, a_j^{i_j}),$$

and the maps

$$\bar{\mu}^k(\vec{a}_j) := (a_j^1, \dots, a_j^{k-1}, \mu(a_j^k, a_j^{k+1}), a_j^{k+2}, \dots a_j^{i_j}),$$

$$\bar{\partial}(\vec{a}_j) := \sum_{k=1}^{i_j} (a_j^1, \dots, a_j^{k-1}, \partial a_j^k, a_j^{k+1}, \dots a_j^{i_j}).$$

This allows us to define the sequence

$$\bar{\mu}_j^k(\vec{a}_1, \dots, \vec{a}_h) := (\vec{a}_1, \dots, \bar{\mu}^k(\vec{a}_j), \dots, \vec{a}_h).$$

With this in mind, and omitting subscripts, the relations which the multiplication maps must satisfy is given by

$$\begin{aligned} 0 = \sum_{j_1,\dots,j_h} & m(m(x, T_{j_1}(\vec{a}_1), \dots, T_{j_h}(\vec{a}_h)), T^{j_1}(\vec{a}_1), \dots, T^{j_h}(\vec{a}_h)) \\ & + \sum_{j=1}^{h} \sum_{k=1}^{i_j - 1} m(x, \bar{\mu}_j^k(\vec{a}_1, \dots, \vec{a}_h)) + \sum_{k=1}^{h} m(x, \vec{a}_1, \dots, a_{\vec{k}-1}, \bar{\partial}(\vec{a}_k), a_{\vec{k}+1}, \dots, \vec{a}_h). \end{aligned}$$

When $\mathcal{A}_1, \dots, \mathcal{A}_h$ are all copies of the torus algebra, we will have cause to consider decompositions of these maps according to the extra data of a splicing of $\vec{a}_1, \dots, \vec{a}_h$ — the easiest way to define the $\mathcal{A}_\infty$ structure maps we will be interested in later in the paper is in terms of a series of auxiliary maps which sum to give the structure maps, and themselves satisfy a relation that implies the $\mathcal{A}_\infty$ relations for their sums. First, we need to discuss operations upon interleaved splicings which are compatible with the operations T_*, T^* and $\bar{\mu}_j^i$.

Fix an interleaved splicing $\vec{\sigma}$ of a set of sequences of Reeb chords $\vec{\rho} = \vec{\rho}_1, \dots, \vec{\rho}_h$. We shall say a pair of indices (i, j) is *compatible* with $(\vec{\rho}, \vec{\sigma})$ if either

(1) $\rho_j^i \rho_j^{i+1}$ is nonzero and precisely one of ρ_j^i or ρ_j^{i+1} is jumping (in part P_j^k, say), or

(2) $\rho_j^i \rho_j^{i+1}$ is nonzero, both of $\rho_j^i \in P_j^{k_i}$ and $\rho_j^{i+1} \in P_j^{k_{i+1}}$ are jumping, and we have that they are in consecutive parts — that is, $\rho_j^i \in P_j^{k_i}$ and $\rho_j^{i+1} \in P_j^{k_i+1}$ for some k_i.

If (i, j) are compatible with $\vec{\rho}, \vec{\sigma}$, then there is an associated splicing of $\bar{\mu}^i_j(\vec{\rho})$ which we call $\vec{\sigma}^i_j$. In case (1), we set $\vec{\sigma}^i_j$ to be the splicing of $\bar{\mu}^i_j(\vec{\rho})$ obtained from $\vec{\sigma}$ by replacing P^k_j with the part $\{\rho^i_j \rho^{i+1}_j\}$. In case (2), we set $\vec{\sigma}^i_j$ to be the splicing given by omitting, for each l, the parts $P^{k_i}_l$ and $P^{k_i+1}_l$ from P_l.

If $\vec{\sigma}$ is interleaved, we shall say that an index $0 < k < m$ is *collidable* for $(\vec{\rho}, \vec{\sigma})$ if for every j we have that only one of P^k_j and P^{k+1}_j is nonempty. In this circumstance, we can define an associated $(m-1)$-splicing of $\vec{\rho}$ associated to $\vec{\sigma}$ and k which we call $\vec{\sigma}(k)$, the *collision of* $\vec{\sigma}$ *at* k. $\vec{\sigma}(k)$ consists of the partitions Q_j, where each Q_j is formed by

$$Q^i_j := \begin{cases} P^i_j & \text{if } i < k, \\ P^i_j \cup P^{i+1}_j & \text{if } i = k, \\ P^{i+1}_j & \text{if } i > k+1. \end{cases}$$

In words, we amalgamate the k and $(k+1)^{\text{st}}$ parts of each partition. Colliding does not preserve being interleaved.

Example 2.13 Suppose that $\vec{\rho} = \vec{\rho}_1 = (\rho_2, \rho_{12}, \rho_3, \rho_2, \rho_3)$, with splicing $\vec{\sigma} = (\rho_2, \rho_3, \rho_2, \rho_3)$. The indices $(2, 1)$ are compatible, in case (1), as $\rho_{12} \cdot \rho_3 = \rho_{123} \neq 0$ and ρ_3 is the only jumping chord of the two — here, the associated splicing of $\bar{\mu}^2_1(\vec{\rho})$ is given by $(\rho_2, \rho_{123}, \rho_2, \rho_3)$. The indices $(4, 1)$ are compatible, in case (2), as (trivially) the final ρ_2 and ρ_3 belong to consecutive parts of $\vec{\sigma}$ — the associated splicing of $\bar{\mu}^4_1(\vec{\rho})$ is given by (ρ_2, ρ_3).

Suppose that $\vec{\rho}_2 = (\rho_{12}, \rho_1, \rho_{23}, \rho_2)$, and let

$$\vec{\sigma}_2 := (\varnothing, \varnothing, \rho_1, \varnothing, \rho_2, \varnothing), \quad \vec{\sigma}_1 := (\rho_2, \rho_3, \varnothing, \rho_2, \varnothing, \rho_3).$$

Together, $(\vec{\sigma}_1, \vec{\sigma}_2)$ form an interleaved splicing of $(\vec{\rho}_1, \vec{\rho}_2)$ and we have that the index 1 is *not* collidable as P^1_1 and P^2_1 are both nonempty. The index 3, on the other hand, is collidable: the collision of $(\vec{\sigma}_1, \vec{\sigma}_2)$ at 3 is given by $((\rho_2, \rho_3, \rho_2, \varnothing, \rho_3), (\varnothing, \varnothing, \rho_1, \rho_2, \varnothing))$ which is no longer interleaved.

Definition 2.14 Let $\vec{\sigma}(i_1, \dots, i_h)$ denote the set of all splicings of all sets of sequences of Reeb chords in $\mathcal{A}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{A}_h^{\otimes i_h}$. We shall say a sequence of maps

$$n_{i_1,\dots,i_h} \colon M \otimes_{\mathcal{I}} \mathcal{A}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{A}_h^{\otimes i_h} \times \vec{\sigma}(i_1, \dots, i_h) \to M$$

satisfies a *partial* $\mathcal{A}_\infty$ *relation* if, for every $x \in M$, set of sequences of Reeb chords $\vec{\rho}$, and interleaving $\vec{\sigma}$ of $\vec{\rho}$, we have that

$$0 = \sum_{\substack{\vec{\lambda} \star \vec{\delta} = \vec{\rho} \\ \vec{\sigma}_1 \star \vec{\sigma}_2 = \vec{\sigma}}} n(n(x, \vec{\lambda}, \vec{\sigma}_1), \vec{\delta}, \vec{\sigma}_2) + \sum_{i,j \text{ compatible}} n(x, \bar{\mu}^i_j(\vec{\rho}), \vec{\sigma}^i_j) + \sum_{k \text{ collidable}}^{|\vec{\sigma}|-1} n(x, \vec{\rho}, \vec{\sigma}(k)),$$

Proposition 2.15 *Let $n_{i_1,\dots,i_h}$ be a sequence of maps satisfying the partial $\mathcal{A}_\infty$ relation. Then the maps*

$$m_{i_1,\dots,i_h} \colon M \otimes_{\mathcal{I}} \mathcal{A}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{A}_h^{\otimes i_h} \to M$$

defined by

$$m_{i_1,\dots,i_h}(x,\vec{\rho}) := \sum_{\vec{\sigma} \text{ an interleaving for } \vec{\rho}} n_{i_1,\dots,i_h}(x,\vec{\rho},\vec{\sigma})$$

satisfy the $\mathcal{A}_\infty$ relations.

Proof This follows readily by fixing x and $\vec{\rho}$, and summing the partial $\mathcal{A}_\infty$ relations over all interleavings for $\vec{\rho}$. The first two terms in the partial relation sums to the first two terms in the $\mathcal{A}_\infty$ relations, and the final terms cancel in pairs where $\vec{\sigma}$ and $\vec{\sigma}'$ are the two unique interleavings which satisfy $\vec{\sigma}(k) = \vec{\sigma}'(k)$. □

3 Diagrams

In this section, we will be concerned with *diagrams* of various sorts. The reader is encouraged to read Sections 3.0.1 and 3.0.3 before this section, for some precise cases of the construction and for some examples. We make a broad but vague definition here to try and minimise repetition of associated definitions later:

Definition 3.1 A *diagram* is a tuple $\mathcal{D} = (\Sigma, \boldsymbol{\gamma}; \boldsymbol{b}, \boldsymbol{p})$ consisting of the following data:

(1) a connected, oriented surface Σ of genus g, possibly with boundary;

(2) an ordered tuple of *attaching curves* $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m\}$, where each $\boldsymbol{\gamma}^i$ is a set of embedded curves and arcs in Σ;

(3) a finite set of marked points $\boldsymbol{b} \subset \partial\Sigma$;

(4) a finite set of distinguished points $\boldsymbol{p} \subset \mathrm{Int}(\Sigma)$.

We make the following broad definition:

Definition 3.2 Let $\mathcal{D}$ be a diagram with two sets of attaching curves and arcs, $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$. A *generator* for $(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ is a set of intersection points $\boldsymbol{x} = \{x_1, \dots, x_\kappa\}$ where the x_i are members of $\boldsymbol{\gamma} \cap \boldsymbol{\gamma}'$, such that

- every closed curve in $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ contains precisely one x_i, and
- exactly half of the arcs in each of $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ contain precisely one x_i.

We will denote the set of generators for $(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ by $\mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$. With the genericity of the above definition of a diagram, of course, it is possible that $\mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ is empty.

In more generality, we will say a *sequence of generators* for a diagram with more than two sets of attaching curves $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$ is a sequence $\boldsymbol{x}^{1,2}, \dots, \boldsymbol{x}^{m,1}$ where $\boldsymbol{x}^{i,i+1} \in \mathcal{G}(\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{i+1})$ (here, the superscripts are taken modulo m).

If $\mathcal{D}$ is a diagram with m sets of attaching curves, then for any subset $U \subset \{1, \dots, m\}$ of size l, there is an associated set of attaching curves $\boldsymbol{\gamma}(U) := \{\boldsymbol{\gamma}^i \mid i \in S\}$, and the data $\mathcal{D}|_U := (\Sigma; \boldsymbol{\gamma}(U); \boldsymbol{p}; \boldsymbol{b})$ also constitutes a diagram, which we call an *l-subdiagram of $\mathcal{D}$*.

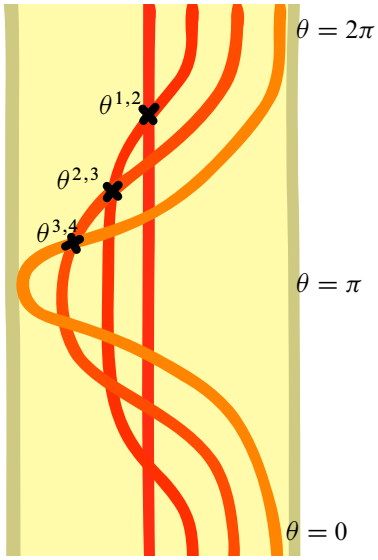


Figure 1: A neighbourhood of an attaching curve γ, shown in red, and the images of the Hamiltonian isotopy $H(\gamma, t)$ for various $t \in [0, \varepsilon]$, where the curves get progressively more orange as t nears ε.

We will later be interested in perturbing some of the curves in a diagram. A way to do this is given in [10, Lemma 11.8]. Namely, Lipshitz describes a Hamiltonian $H\colon \Sigma \to \mathbb{R}$ which is supported within a tubular neighbourhood of a set of attaching curves, such that the corresponding Hamiltonian isotopy $H(x, t)$ behaves as in Figure 1 (θ refers to the S^1 coordinate of the tubular neighbourhood, identified with $S^1 \times [0, 1]$).

Definition 3.3 Suppose $\boldsymbol{\gamma}$ is a collection of curves in some diagram. Then we define the ε-small m-fold approximation of $\boldsymbol{\gamma}$ to be the set of sets of arcs and curves $\boldsymbol{\gamma}(\varepsilon, m) = \{\boldsymbol{\gamma}^1(\varepsilon, m), \dots, \boldsymbol{\gamma}^j(\varepsilon, m)\}$, where $\boldsymbol{\gamma}^i(\varepsilon, m) = H(\boldsymbol{\gamma}, i\varepsilon/m)$. We shall sometimes write $\boldsymbol{\gamma}^{\varepsilon/m;i}$ for $\boldsymbol{\gamma}^i(\varepsilon, m)$ when it is more convenient.

If $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$ is a sequence of curves that may be written as $\boldsymbol{\gamma}(\varepsilon, m)$ for some $\boldsymbol{\gamma}$, ε and m, we will often say that $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$ are *small approximations of one another*, and, when $m = 2$, we shall write $\boldsymbol{\gamma}^1 \approx \boldsymbol{\gamma}^2$. If $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \subset \Sigma$ are small approximations of one another for some ε, we will say that the subset of Σ given by $\bigcup_{t \in [0,\varepsilon]} H(\boldsymbol{\gamma}^1, t)$ is the *approximation region* for $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2$. More generally, if $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$ are all small approximations of one another, we say the approximation region for $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$ is the union of the approximation regions for each $\boldsymbol{\gamma}^i \approx \boldsymbol{\gamma}^{i+1}$.

Note that in an ε-small m-fold approximation, any consecutive set of arcs and curves $\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{i+1}$ constitute the ε/m-small 2-fold approximation of $\boldsymbol{\gamma}$. For any 2-fold approximation $\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{i+1}$, we can choose a *distinguished generator* $\boldsymbol{\theta}^{i,i+1}$: a choice of, for every $i = 1, \dots, g + k$, one of the two elements of $\boldsymbol{\gamma} \cap \boldsymbol{\gamma}'$. Hence, given an ε-small m-fold approximation, we may choose a distinguished generator for each consecutive pair $\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{i+1}$ to give a set of distinguished generators $\vec{\boldsymbol{\theta}} = \{\boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,m}\}$.

We will be largely concerned with a canonically defined set of distinguished generators. Namely:

Example 3.4 Consider the intersection points $\theta^{i,i+1}_{+,j} \in \gamma^i_j \cap \gamma^{i+1}_j$ as shown in Figure 1. The collection of $\boldsymbol{\theta}^{i,i+1}_{+} = \bigcup_j \theta^{i,i+1}_{+,j}$ for every i gives a set of distinguished generators $\vec{\boldsymbol{\theta}}_+$.

Sometimes, we refer to the other choice of intersection point in $\gamma^i_j \cap \gamma^{i+1}_j$ as $\theta^{i,i+1}_{-,j}$.

Definition 3.5 Suppose $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ are two collections of arcs and curves in some diagram, and let $\boldsymbol{x} \in \mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ be a generator.

If $\boldsymbol{\gamma}^{\varepsilon/m;1}, \dots, \boldsymbol{\gamma}^{\varepsilon/m;m}$ is the ε-small m-fold approximation to $\boldsymbol{\gamma}$, then, provided ε is sufficiently small, there are a series of canonical *nearest point maps*, which are given by, for each i, the map

$$\varphi_i : \mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}') \to \mathcal{G}(\boldsymbol{\gamma}^{\varepsilon/m;i}, \boldsymbol{\gamma}')$$

which sends $x_j \in \gamma_j \cap \gamma'_{\sigma(j)}$ to the canonical nearest point $x^{\varepsilon/m;i}_j \in \gamma^{\varepsilon/m;i}_j \cap \gamma'_{\sigma(j)}$.

If we are given a set of distinguished generators $\vec{\boldsymbol{\theta}}$ for the approximation, then this allows us to associate to any pair of generators $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{G}(\boldsymbol{\gamma}, \boldsymbol{\gamma}')$ a sequence of generators

$$(\boldsymbol{x}', \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,m}, \boldsymbol{y}')$$

in the set

$$\mathcal{G}(\boldsymbol{\gamma}', \boldsymbol{\gamma}^{\varepsilon/m;1}) \times \mathcal{G}(\boldsymbol{\gamma}^{\varepsilon/m;1}, \boldsymbol{\gamma}^{\varepsilon/m;2}) \times \cdots \times \mathcal{G}(\boldsymbol{\gamma}^{\varepsilon/m;m-1}, \boldsymbol{\gamma}^{\varepsilon/m;m}) \times \mathcal{G}(\boldsymbol{\gamma}^{\varepsilon/m;m}, \boldsymbol{\gamma}'),$$

where $\boldsymbol{x}' = \boldsymbol{x}$, and $\boldsymbol{y}' = \varphi_m(\boldsymbol{y})$. We call this sequence the *ε-small m-fold splaying* of $(\boldsymbol{x}, \boldsymbol{y})$, and shall often abuse notation and write $(\boldsymbol{x}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1m}, \boldsymbol{y})$ for the relevant generator.

3.0.1 Toroidal bordered Heegaard diagrams

One of the main instances of a diagram we shall consider is when $\mathcal{D}$ specifies a three-manifold with boundary that is a disjoint union of tori. The definition we shall use is:

Definition 3.6 Suppose that $\mathcal{D} = (\Sigma; \boldsymbol{\gamma}; \boldsymbol{b}; \boldsymbol{p})$ is a diagram. Then we shall say that $\mathcal{D}$ is a *toroidal bordered Heegaard diagram* if:

(1) Σ is a genus g surface with h boundary components.

(2) $\boldsymbol{\gamma}$ consists of only two sets of pairwise disjoint curves $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, which intersect transversely, such that $\boldsymbol{\beta}$ consists of $g - 1 + h$ simple closed curves, and $\boldsymbol{\alpha}$ can be partitioned into two sets, $\boldsymbol{\alpha} = \boldsymbol{\alpha}^a \cup \boldsymbol{\alpha}^c$, where $\boldsymbol{\alpha}^a$ consists of $2h$ arcs and $\boldsymbol{\alpha}^c$ consists of $g - 1$ closed curves.

(3) The set $\boldsymbol{\alpha}^a$ can be written $\boldsymbol{\alpha}^a = \{\alpha^m_1, \alpha^l_1, \dots, \alpha^m_h, \alpha^l_h\}$, where α^m_i and α^l_i intersect $\partial\Sigma_i$ relative to the point $b_i \in \partial\Sigma_i$ as shown in Figure 2, and is disjoint from $\partial\Sigma_j$ for $i \neq j$.

(4) Every region of the surfaces $\Sigma - \boldsymbol{\alpha}$ and $\Sigma - \boldsymbol{\beta}$ meets some basepoint $b_i \in \boldsymbol{b}$ (we sometimes call this *homological linear independence*).

(5) The set of points $\boldsymbol{p}$ is empty (we therefore usually omit this from the notation).

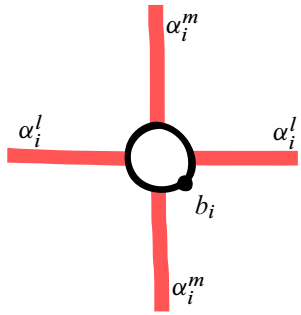


Figure 2: Labelling conventions for α_i^m and α_i^l.

An example of a toroidal bordered Heegaard diagram with 2 boundary components can be found on the left of Figure 8.

In the case where $\mathcal{D}$ is a toroidal bordered Heegaard diagram with $|\partial\Sigma| = 1$, this constitutes a bordered Heegaard diagram compatible with a manifold with torus boundary in the sense of [14]. When $|\partial\Sigma| > 1$, the data corresponds to a *bordered sutured Heegaard diagram* in the sense of [26]. The data of $\boldsymbol{b} = \{b_1, \dots, b_h\}$ corresponds to a series of arcs $\Gamma_{\boldsymbol{b}} \subset \partial\Sigma$, so that $\boldsymbol{\alpha}^a$ and $\Gamma_{\boldsymbol{b}}$ specify an embedding of the disjoint union of h arc diagrams for the torus into Σ; that is, $\mathcal{D}$ is equivalent to a bordered sutured Heegaard diagram for a three-manifold with h torus boundary components. We will tend to ignore the distinction between $\mathcal{D}$ and the corresponding bordered sutured diagram.

Convention 3.7 To avoid proliferation of adjectives, we will often refer to a toroidal bordered Heegaard diagram as simply a *bordered diagram*.

Note that when $\mathcal{D}$ is a bordered diagram, a generator for $\mathcal{D}$ is a set of intersection points $x_1, \dots, x_{g+h-1}$, where $x_1, \dots, x_h \in \boldsymbol{\alpha}^a \cap \boldsymbol{\beta}$ and $x_{h+1}, \dots, x_{g+h-1} \in \boldsymbol{\alpha}^c \cap \boldsymbol{\beta}$. Without loss of generality, for $i = 1, \dots, h$, we have that $x_i \in \alpha_i^{\delta_i} \cap \beta_{\sigma(i)}$, where $\delta_i \in \{m, l\}$ and σ is some permutation of $\{1, \dots, g\}$. If $\delta = (\delta_1, \dots, \delta_h)$ is a sequence of elements of $\{m, l\}$, then we can associate the idempotent $\iota(\delta) = (\iota_{\delta_1}, \dots, \iota_{\delta_h})$ in $\mathcal{T}^h$. Accordingly, we may associate to any generator $\boldsymbol{x}$ the corresponding idempotent $\iota(\boldsymbol{x}) := \iota(\delta)$ — this is called $I_A(\boldsymbol{x})$ in [14].

It will be of use to us later to note that although the set $\boldsymbol{p}$ is empty if $\mathcal{D}$ is a bordered diagram, there is a set of points $\boldsymbol{p}(\boldsymbol{b}) \subset \mathrm{Int}(\Sigma)$ given by pushing each $b_i \in \boldsymbol{b}$ slightly off of the boundary $\partial\Sigma_i$ into the region of $\Sigma - \boldsymbol{\alpha} - \boldsymbol{\beta}$ adjacent to b_i.

3.0.2 Link diagrams Another important instance of a diagram is what we term a *$2h$-pointed link diagram*; cf [20, Definition 3.1] where these are called *$2l$-pointed Heegaard diagrams*; these are

Definition 3.8 Let $\mathcal{D} = (\Sigma; \boldsymbol{\gamma}; \boldsymbol{b}; \boldsymbol{p})$ be a diagram. Then we say $\mathcal{D}$ is a *$2h$-pointed link diagram* if:

(1) Σ is a closed surface of genus g.

(2) $\boldsymbol{\gamma}$ consists of two sets of pairwise disjoint curves $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$, which intersect transversely, such that each is a set of $g+h-1$ simple closed curves (ie none of them are arcs).

(3) Each set $\boldsymbol{\gamma}^i$ spans a g-dimensional sublattice of $H_1(\Sigma)$. Equivalently, for each i, the surface $\Sigma-\boldsymbol{\gamma}^i$ has precisely h components $A^i_1,\dots,A^i_l$.

(4) The set $\boldsymbol{b}$ is empty (this is implied by Σ being closed, and we will often omit it from the notation).

(5) The set $\boldsymbol{p}$ can be ordered as $\boldsymbol{p}=(w_1,z_1,\dots,w_h,z_h)$, such that $w_j,z_j\in A^1_j\cap A^2_j$. We will often write $\boldsymbol{p}=\boldsymbol{w}\cup\boldsymbol{z}$ for the partition of $\boldsymbol{p}$ into $(w_1,\dots,w_h)$ and $(z_1,\dots,z_h)$ respectively.

Often, we write $\boldsymbol{\gamma}^1:=\boldsymbol{\beta}$ and $\boldsymbol{\gamma}^2:=\boldsymbol{\alpha}$. For an example of a 4-pointed link diagram with $g=0$ and $h=2$, please see Figure 3 — at this stage, the purple and orange curves are to be ignored. Another example of a link diagram with basepoints omitted can be found on the right of Figure 8.

3.0.3 Heegaard multidiagrams Another important instance of a diagram is a *Heegaard multidiagram*. We follow approximately the treatment given in [13], though the notion goes back as far as [18, Section 8.1.5].

Definition 3.9 Let $\mathcal{D}=(\Sigma;\boldsymbol{\gamma};\boldsymbol{b};\boldsymbol{p})$ be a diagram. Then we say $\mathcal{D}$ is a *Heegaard m-diagram* if:

(1) Σ is a closed surface of genus g.

(2) $\boldsymbol{\gamma}$ consists of m sets of $g+h-1$ pairwise disjoint simple closed curves $\boldsymbol{\gamma}^i$, such that each $\boldsymbol{\gamma}^i$ intersects $\boldsymbol{\gamma}^j$ transversally for $i\neq j$.

(3) For each i, the surface $\Sigma-\boldsymbol{\gamma}^i$ has precisely h components $A^i_1,\dots,A^i_h$.

(4) The set $\boldsymbol{b}$ is empty (this is implied by Σ being closed, and we will often omit it from the notation).

(5) The set $\boldsymbol{p}$ consists of h points, $\boldsymbol{p}=(z_1,\dots,z_h)$, such that (without loss of generality) $z_j\in\bigcap_{i=1}^m A^i_j$.

Examples of 3- and 4-diagrams with basepoints omitted can be found on the left and right (respectively) of Figure 10.

3.0.4 Bordered multidiagrams We roughly generalise both a toroidal bordered diagram and a Heegaard m-diagram:

Definition 3.10 Let $\mathcal{D}=\mathcal{D}=(\Sigma;\boldsymbol{\gamma}^0,\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^{m-1};\boldsymbol{b};\boldsymbol{p})$ be a diagram. Then we say $\mathcal{D}$ is a *toroidal bordered multidiagram* if:

(1) Σ is a surface with h boundary components $\partial\Sigma_1,\dots,\partial\Sigma_h$.

(2) $(\Sigma,\boldsymbol{\gamma}^0,\boldsymbol{\gamma}^1,\boldsymbol{b})$ is a toroidal bordered Heegaard diagram (such that $\boldsymbol{\gamma}^0$ are closed curves, and $\boldsymbol{\gamma}^1=\boldsymbol{\gamma}^{1,a}\cup\boldsymbol{\gamma}^{1,c}$ splits as a series of $2h$ arcs and $g-1$ closed curves).

(3) For $i\geq 2$ each of the sets of curves $\boldsymbol{\gamma}^i$ is a set of $g+h-1$ pairwise disjoint simple closed curves such that each $\boldsymbol{\gamma}^i$ intersects $\boldsymbol{\gamma}^j$ transversally for $i\neq j$.

(4) For each $i > 1$, the surface $\Sigma - \boldsymbol{\gamma}^i$ has precisely h components $A_1^i, \dots, A_h^i$.

(5) The set $\boldsymbol{p} = (z_1, \dots, z_h)$ is given by the push-off $\boldsymbol{p}(\boldsymbol{b})$, and (without loss of generality) $z_j \in \bigcap_{i=1}^m A_j^i$.

Given a toroidal bordered multidiagram, there are two associated surfaces which we will have cause to mention later. The first is the associated compact surface $\overline{\Sigma}$, which comes with associated arcs $\bar{\boldsymbol{\gamma}}^1$ — we may view Σ as $\overline{\Sigma} - \partial\overline{\Sigma}$. The second is the surface $\Sigma_{\bar{e}}$ which is the result of collapsing each boundary circle $\partial\overline{\Sigma}_i$ of $\overline{\Sigma}$ down to a point e_j^∞ — thus we can also view Σ as a punctured surface $\Sigma_{\bar{e}} \setminus \{e_1^\infty, \dots, e_h^\infty\}$. This comes endowed with natural closed curves $\boldsymbol{\gamma}_{\bar{e}}$ which are the closures of $\boldsymbol{\gamma}$ under this identification, and basepoints $\boldsymbol{p}_{\bar{e}}$ such that $\mathcal{D}_{\bar{e}} = (\Sigma_{\bar{e}}; \boldsymbol{\gamma}_{\bar{e}}; \boldsymbol{p}_{\bar{e}})$ is a diagram, which we sometimes call the *east compactification* of $\mathcal{D}$.

3.0.5 Converting between diagrams Most of the content of this paper is in relating counts of certain moduli spaces related to bordered diagrams with those related to multipointed Heegaard multidiagrams. As such, we will need to have a dictionary between concepts for the above two diagrams.

We will work with a special class of diagrams in both cases. We first need to introduce the notion of a *handle*. Please see Figure 6 for a picture of the closed -6-twisted handle with the curves $\beta_{\mathfrak{H}}$ and $\alpha_{\mathfrak{H}}(l)$ pictured; and Figure 5 for a picture of a closed handle with the curves $\beta_{\mathfrak{H}}$ and $\alpha_{\mathfrak{H}}(m)$ shown. A bordered -6-twisted handle is also shown in Figure 4.

Definition 3.11 For $n \in \mathbb{Z}$, the *closed n-twisted handle* is a tuple

$$\mathfrak{H}(n) = (H; \alpha_{\mathfrak{H}}(m), \beta_{\mathfrak{H}}, \alpha_{\mathfrak{H}}(l); \boldsymbol{p}_{\mathfrak{H}}),$$

where

- H is the tube $S^1 \times [-1, 1]$, with left and right ends $\mathfrak{H}_L := S^1 \times \{-1\}$ and $\mathfrak{H}_R := S^1 \times \{1\}$ respectively;
- $\alpha_{\mathfrak{H}}(m)$ is the curve $S^1 \times \{0\}$;
- $\beta_{\mathfrak{H}}$ is the arc $\{0\} \times [-1, 1]$;
- $\alpha_{\mathfrak{H}}(l)$ is the arc $\{(e^{ni\pi t}, t) \mid t \in [-1, 1]\}$.
- $\boldsymbol{p}_{\mathfrak{H}}$ is the pair of basepoints $(w_{\mathfrak{H}}, z_{\mathfrak{H}})$ as defined below.

For the definition of $\boldsymbol{p}_{\mathfrak{H}}$, we choose a disc neighbourhood of the point $\alpha_{\mathfrak{H}}(m) \cap \alpha_{\mathfrak{H}}(l)$ which is sufficiently small that the triple $(D, \alpha_{\mathfrak{H}}(m) \cap D, \alpha_{\mathfrak{H}}(l) \cap D)$ is modelled on the unit disc $\mathbb{D} \subset \mathbb{C}$, together with the imaginary and real axes respectively. Under this identification, we put $w_{\mathfrak{H}} = \frac{1}{2}e^{3\pi i/4}$ and $z_{\mathfrak{H}} = \frac{1}{2}e^{-3\pi i/4}$.

The *bordered n-twisted handle* is a tuple $\boldsymbol{\mathfrak{H}}(n) = (\boldsymbol{H}; \alpha_{\mathfrak{H}}(m), \beta_{\mathfrak{H}}, \alpha_{\mathfrak{H}}(l); b)$, where $\boldsymbol{H}$, $\alpha_{\mathfrak{H}}(m)$, $\beta_{\mathfrak{H}}$ and $\alpha_{\mathfrak{H}}(l)$ are the punctured tube and arcs induced from H, $\alpha_{\mathfrak{H}}(m)$, $\beta_{\mathfrak{H}}$ and $\alpha_{\mathfrak{H}}(l)$ by removing D from H. Under the identification of $\mathbb{D}$ with D, the boundary of $\boldsymbol{H}$ consists of three components: $\boldsymbol{\mathfrak{H}}_L, \boldsymbol{\mathfrak{H}}_R$ and a new circle $\boldsymbol{\mathfrak{H}}_\partial$ modelled upon $S^1 = \partial\mathbb{D}$. We define b to be the point in η_∂ that corresponds to $e^{-\pi i/4}$.

It will be useful for us to note here that there is a clear bijection between the sets of curves $\alpha_{\mathfrak{H}}(m), \beta_{\mathfrak{H}}, \alpha_{\mathfrak{H}}(l)$ and $\alpha_{\mathfrak{H}}(m), \beta_{\mathfrak{H}}, \alpha_{\mathfrak{H}}(l)$ respectively, given by the map $\cap_D(\gamma) = \gamma \cap (\Sigma \setminus D)$.

We will shortly construct some diagrams by gluing handles to a multipointed Heegaard multidiagram in a way specified by some fixed data. First, we consider sequences of idempotents:

Definition 3.12 An *h-splaying sequence of idempotents* is a sequence $\vec{\iota}$ of $m+1$ sequences of elements in the subring of idempotents of the h-torus algebra; $\vec{\iota} = \vec{\iota}^0, \dots, \vec{\iota}^m$, where $\vec{\iota}^i = (\iota^i_1, \dots, \iota^i_h)$ and each ι^i_j is one of ι_m or ι_l.

Note that if $\vec{\rho} = \vec{\rho}_1, \dots, \vec{\rho}_h$ is a set of sequences of Reeb chords and $\vec{\sigma}$ is a splicing for $\vec{\rho}$, the associated sequence $\vec{\iota}_\sigma$ is an h-splaying sequence of idempotents. In general, the data we shall require is as follows.

Definition 3.13 Suppose $\mathcal{D}$ is a link diagram $(\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{p})$, where $|\boldsymbol{p}| = 2h$. By definition, we can number the regions A of $\Sigma - \boldsymbol{\alpha}$ and B of $\Sigma - \boldsymbol{\beta}$ by the numbers $1, \dots, h$ so that the points $\boldsymbol{p} = \{w_1, z_1, \dots, w_h, z_h\}$ satisfy $(w_i, z_i) \in A_i \cap B_i$.

A set of *stabilisation data for* $\mathcal{D}$ is a tuple $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$, where:

(1) $\boldsymbol{\lambda}_{w,z} = \{\lambda_{w_1,z_1}, \dots, \lambda_{w_h,z_h}\}$ and $\boldsymbol{\lambda}_{z,w} = \{\lambda_{z_1,w_1}, \dots, \lambda_{z_h,w_h}\}$ are oriented arcs in Σ chosen such that
   - $\partial \lambda_{w_i,z_i} = z_i - w_i$,
   - $\partial \lambda_{z_i,w_i} = w_i - z_i$,
   - $\boldsymbol{\lambda}_{w,z} \cap \boldsymbol{\alpha} = \varnothing$, and
   - $\boldsymbol{\lambda}_{z,w} \cap \boldsymbol{\beta} = \varnothing$.

(2) $\boldsymbol{n} = (n_1, \dots, n_h)$ is a sequence of integers.

(3) ε is a small approximation parameter.

(4) $\vec{\iota}$ is an h-splaying sequence of idempotents.

This is demonstrated in Figure 3 for a specific choice of link diagram. (This is a 4-pointed Heegaard diagram for the Hopf link in S^3.)

We now describe the process of gluing handles to a link diagram. The reader is strongly encouraged to refer to Figure 4 either whilst or instead of reading the following definition.

Definition 3.14 Let $\mathcal{D} = (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{p})$ be a link diagram, where $|\boldsymbol{p}| = 2h$, and let $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$ be a set of stabilisation data for $\mathcal{D}$. Suppose that $(\mathfrak{H}_1, \dots, \mathfrak{H}_h)$ is a set of bordered handles.

We label the elements of $\boldsymbol{p}$ by $(w_1, z_1, \dots, w_h, z_h)$, where w_i and z_i both belong to the same component of $\Sigma - \boldsymbol{\alpha}$ and $\Sigma - \boldsymbol{\beta}$. Choose a pair of small neighbourhoods N^w_i and N^z_i around w_i and z_i which are holomorphically identified with the unit disc $\mathbb{D} \subset \mathbb{C}$ in such a way that the basepoint z_i is identified with the origin and the arcs $\lambda_{z_i,w_i} \cap N^z_i$ and $\lambda_{w_i,z_i} \cap N^z_i$ are identified with the positive and negative sides of the real axis respectively; and that $\lambda_{w_i,z_i} \cap N^w_i$ and $\lambda_{w_i,z_i} \cap N^w_i$ are identified with the positive and negative sides of the real axis respectively.

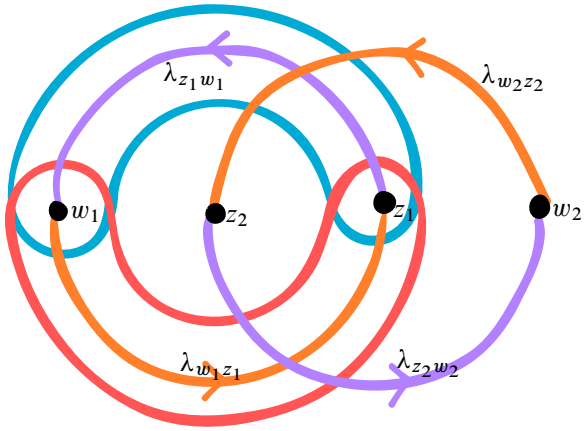


Figure 3: Stabilisation data for a 4-pointed link diagram. Here, the $\boldsymbol{\alpha}$ curves are shown in red and the $\boldsymbol{\beta}$ in blue.

Now choose homeomorphisms h_i^L (resp. h_i^R) from $\mathfrak{H}_i^L$ (resp. $\mathfrak{H}_i^R$) to ∂N_i^w (resp. ∂N_i^z), in such a way that $h_i^L(\alpha_{\mathfrak{H}_i}(l)\cap\mathfrak{H}_i^L)=\lambda_{z_i,w_i}\cap N_i^w$ and $h_i^L(\beta_{\mathfrak{H}_i}\cap\mathfrak{H}_i^L)=\lambda_{w_i,z_i}\cap N_i^w$ (resp. $h_i^R(\alpha_{\mathfrak{H}_i}(l)\cap\mathfrak{H}_i^L)=\lambda_{z_i,w_i}\cap N_i^w$ and $h_i^R(\beta_{\mathfrak{H}_i}\cap\mathfrak{H}_i^L)=\lambda_{w_i,z_i}\cap N_i^w$).

Form the surface

$$\boldsymbol{\Sigma}(d) := (\Sigma - N_1^w - N_2^z - \cdots - N_h^w - N_h^z) \cup_{h^L,h^R} (H_1, \ldots, H_h)$$

by gluing the handles to Σ via the homeomorphisms h_i^L and h_i^R. We denote by **glue** the composition of the projection map from Σ to $(\Sigma - N_1^w - N_2^z - \cdots - N_h^w - N_h^z)$ with the gluing map; that is, it is a map

$$\mathbf{glue}\colon \Sigma \sqcup H_1 \sqcup \cdots \sqcup H_h \to \boldsymbol{\Sigma}(d).$$

There is a set of h curves $\boldsymbol{\beta}_{\mathfrak{H}} := (\lambda_{w_1,z_1} \cup_{h_i^L,h_i^R} \beta_{\mathfrak{H}_1}, \ldots, \lambda_{w_h,z_h} \cup_{h_h^L,h_h^R} \beta_{\mathfrak{H}_h})$ contained in $\Sigma(d)$, and also a sequence of arcs $\boldsymbol{\alpha}^a_{\mathfrak{H}}$ given by $\alpha^a(m_i) := \mathbf{glue}(\alpha_{\mathfrak{H}_i}(m))$ and $\alpha^a(l_i) := \alpha_{\mathfrak{H}_i}(l) \cup_{h_h^L,h_h^R} \lambda_{z_1,w_1}$, for $i = 1, \ldots, h$.

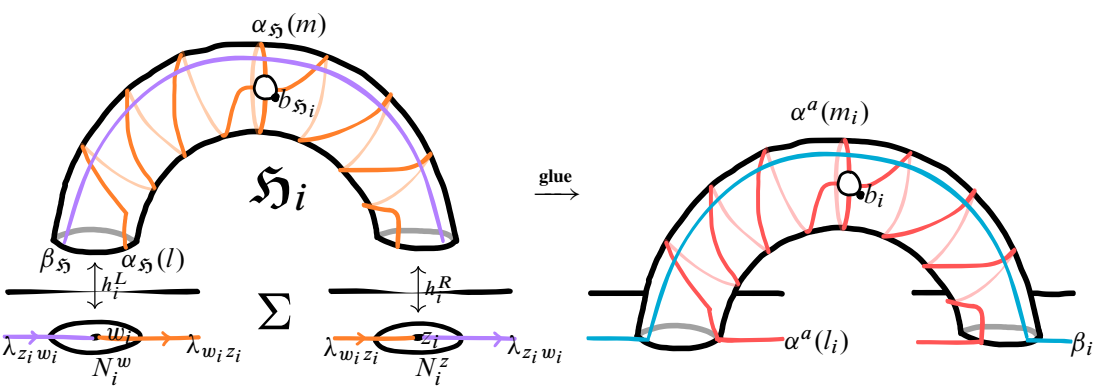


Figure 4: The bordered stabilisation of a link diagram at one pair of basepoints.

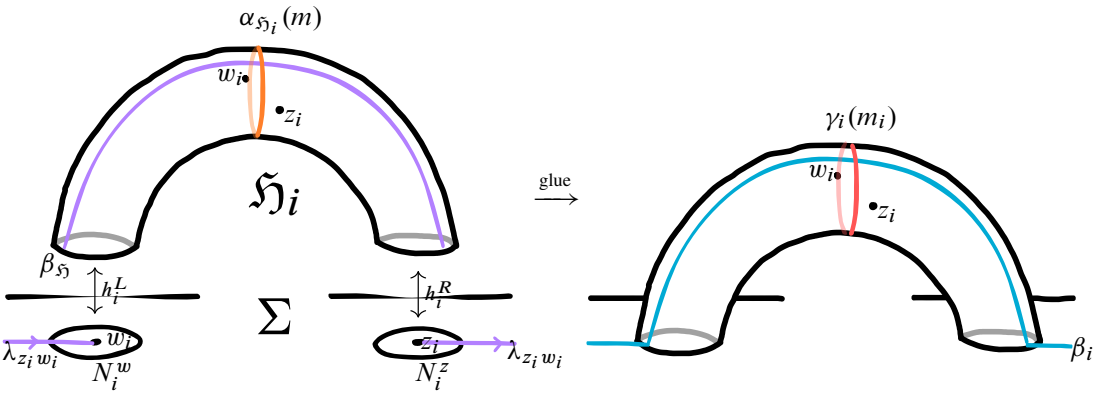


Figure 5: The closed stabilisation of a link diagram at one pair of basepoints w_i, z_i, with meridional corresponding idempotent.

These allow us to form the bordered diagram $\mathcal{D}(d) = (\boldsymbol{\Sigma}(d), \boldsymbol{\alpha}(d)^a, \boldsymbol{\alpha}(d)^c, \boldsymbol{\beta}(d), \boldsymbol{b}(d))$, where

- $\Sigma(d)$ is as above,
- $\boldsymbol{\alpha}(d)^a$ are the arcs $\boldsymbol{\alpha}^a_{\mathfrak{H}}$,
- $\boldsymbol{\alpha}(d)^c$ are the images $\mathbf{glue}(\boldsymbol{\alpha})$,
- $\boldsymbol{\beta}(d)$ is the union of $\boldsymbol{\beta}_{\mathfrak{H}}$ and the images $\mathbf{glue}(\boldsymbol{\beta})$, and
- $\boldsymbol{b}(d)$ is the image $\mathbf{glue}(\boldsymbol{b}(\mathfrak{H}_1) \cup \cdots \cup \boldsymbol{b}(\mathfrak{H}_h))$.

We call this diagram the *bordering* of $\mathcal{D}$; any bordered diagram that arises as $\mathcal{D}(d)$ for some $(\mathcal{D}, d)$ we will call *stable*.

Note that in this construction, the elements of the approximation data ε and $\vec{\iota}$ were redundant. A small adaptation of the above also gives us a multipointed Heegaard multidiagram (illustrated in Figures 5 and 6):

Definition 3.15 Let $\mathcal{D} = (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{p})$ be a link diagram, where $|\boldsymbol{p}| = 2h$, and let $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$ be a set of stabilisation data, such that $\vec{\iota}$ consists of a single idempotent of $\mathcal{T}^h$, $\iota = (\iota_{\delta_1}, \dots, \iota_{\delta_h})$ where each $\delta_i \in \{m, l\}$. Suppose further that $(\mathfrak{H}_1, \dots, \mathfrak{H}_h)$ is a set of closed handles.

One can imitate the construction of the surface $\boldsymbol{\Sigma}(d)$ from the previous definition, but with $(\mathfrak{H}_1, \dots, \mathfrak{H}_h)$ in place of $(\mathfrak{H}_1, \dots, \mathfrak{H}_h)$ to yield a surface $\Sigma(d, \iota) := \Sigma \cup_{h^L, h^R} (H_1, \dots, H_h)$ by gluing the handles to Σ via homeomorphisms h_i^L and h_i^R. We denote by glue the corresponding gluing map

$$\mathrm{glue}\colon \Sigma \sqcup H_1 \sqcup \cdots \sqcup H_h \to \Sigma(d, \iota).$$

Again, there is a set of h curves $\boldsymbol{\beta}_{\mathfrak{H}} := (\lambda_{w_1,z_1} \cup_{h_i^L, h_i^R} \beta_{\mathfrak{H}_1}, \dots, \lambda_{w_h,z_h} \cup_{h_h^L, h_h^R} \beta_{\mathfrak{H}_h})$ contained in $\Sigma(d, \iota)$. There is also a sequence of curves contingent on ι, given by the set $\boldsymbol{\alpha}_{\mathfrak{H}}(d, \iota) = \{\gamma_i(\delta_i)\}$, where

$$\gamma_i(\delta_i) = \begin{cases} \mathrm{glue}(\alpha_{\mathfrak{H}_i}(m)) & \text{if } \delta_i = m, \\ \alpha_{\mathfrak{H}_i}(l) \cup_{h_i^L, h_i^R} \lambda_{z_i, w_i} & \text{if } \delta_i = l. \end{cases}$$

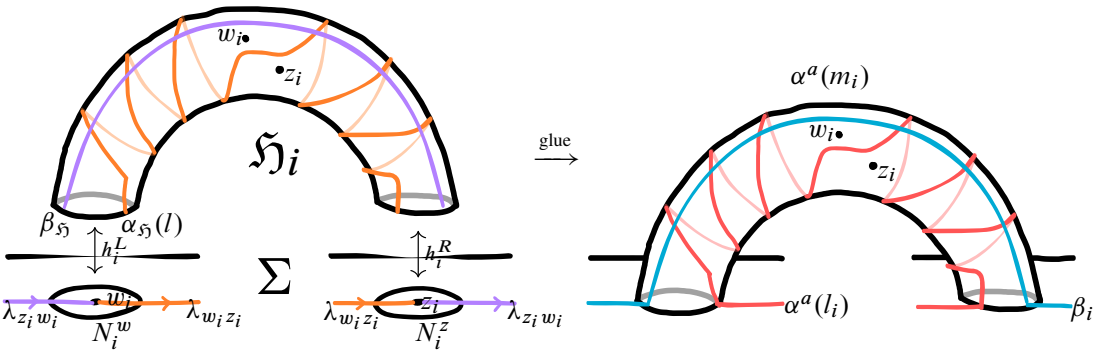


Figure 6: The closed stabilisation of a link diagram at one pair of basepoints w_i, z_i, with longitudinal corresponding idempotent.

These allow us to form the diagram $\mathcal{D}(d,\iota) = (\Sigma(d,\iota), \boldsymbol{\alpha}(d,\iota), \boldsymbol{\beta}(d,\iota), \boldsymbol{p}(d,\iota))$, where

- $\Sigma(d,\iota)$ is as above,
- $\boldsymbol{\alpha}(d,\iota)$ is the union of $\boldsymbol{\alpha}_{\mathfrak{H}}(d,\iota)$ and the images glue$(\boldsymbol{\alpha})$,
- $\boldsymbol{\beta}(d,\iota)$ is the union of $\boldsymbol{\beta}_{\mathfrak{H}}$ and the images glue$(\boldsymbol{\beta})$, and
- $\boldsymbol{p}(d,\iota)$ is the image glue$(\boldsymbol{p}_{\mathfrak{H}_1} \cup \cdots \cup \boldsymbol{p}_{\mathfrak{H}_h})$.

Note that $\cap_D : \mathfrak{H} \to \boldsymbol{\mathfrak{H}}$ extends to a map $\cap_D : \mathcal{D} \to \boldsymbol{\mathcal{D}}(d,\iota)$, which yields a bijection $\cap_D : \boldsymbol{\beta}(d,\iota) \to \boldsymbol{\beta}(d)$. It no longer yields a bijection of the sets $\boldsymbol{\alpha}(d,\iota)$; rather, it gives an injective map sending glue$(\boldsymbol{\alpha}) \mapsto \mathbf{glue}(\boldsymbol{\alpha})$ bijectively, and each element $\gamma_i(\delta_i) \in \boldsymbol{\alpha}_{\mathfrak{H}}(d,\iota)$ to the corresponding arc $\cap_D(\gamma_i(\delta_i)) = \alpha^a(\delta_i)$. It follows that to every generator $\boldsymbol{x} \in \mathcal{G}(\boldsymbol{\mathcal{D}})$ we may associate a generator $\cap_D^{-1}(\boldsymbol{x}) \in \mathcal{D}(d, \iota(\boldsymbol{x}))$ (we often abuse notation for this generator and also denote it by $\boldsymbol{x}$).

The diagram $\mathcal{D}(d,\iota)$ as defined above constitutes a multipointed Heegaard multidiagram (in fact a simple Heegaard diagram), but does not use the information ε from the approximation data d, and is restricted to the case where $\vec{\iota}$ is a single set of idempotents. In more generality, we make a similar construction:

Definition 3.16 Let $\mathcal{D} = (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{p})$ be a link diagram, where $|\boldsymbol{p}| = 2h$, and let $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$ be a set of stabilisation data for $\mathcal{D}$ such that $\vec{\iota} = \iota^1, \dots, \iota^m$.

The *splayed diagram* associated to the pair $(\mathcal{D}, d)$ is the tuple

$$\mathcal{D}(d) := (\Sigma(d), \boldsymbol{\alpha}^1(d), \dots, \boldsymbol{\alpha}^m(d), \boldsymbol{\beta}(d), \boldsymbol{p}(d))$$

given by:

- $\Sigma(d) = \Sigma(d, \iota^1)$ (this is the same as $\Sigma(d, \iota^i)$ for any i).
- $\boldsymbol{\alpha}^i(d)$ is given by the ε-small approximation $H(\boldsymbol{\alpha}(d, \iota(\iota_i)), i\varepsilon/m)$ to the set of curves $\boldsymbol{\alpha}(d, \iota(\iota_i))$ for each $i = 1, \dots, m$, as described in Definition 3.15.

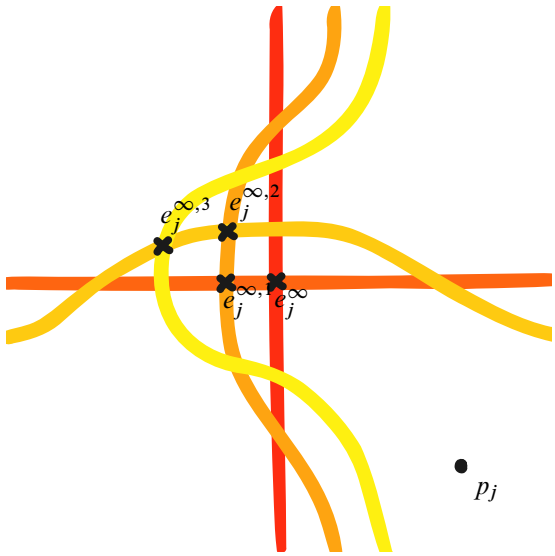


Figure 7: A splayed diagram near e_j^{∞}.

Here, for each $i = 1, \dots, h$, if we denote by $\alpha(d, \iota, \mathfrak{H}_i)$ the curve of $\boldsymbol{\alpha}(d, \iota)$ which intersects $\mathfrak{H}_i$, then we chose the identification of the tubular neighbourhood $N(\alpha, i)$ of $\alpha(d, \iota, \mathfrak{H}_i)$ with $S^1 \times [0, 1]$ so that the intersection of the complementary curve $\alpha(d, \iota^c, \mathfrak{H}_i)$ with $N(\alpha, i)$ coincides with the arc $\pi \times [0, 1]$. (See Figure 7.)

- $\boldsymbol{\beta}(d)$ is given by the set $\boldsymbol{\beta}(d, \iota_1)$ (this, again, is really independent of ι_1).
- $\boldsymbol{p}(d)$ is given by the set $\bigcup_{i=1}^{m} \boldsymbol{p}_{\mathfrak{H}_i}$.

If $\boldsymbol{\mathcal{D}}$ is a stable bordered diagram $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}(d)$ for some d and $\mathcal{D}$, then we will refer to the corresponding splayed diagram $\mathcal{D}(d)$ as the *splaying* of $\boldsymbol{\mathcal{D}}$. We will often denote this simply by $\mathcal{D}(\varepsilon, \vec{\iota})$, as these are the only parts of the splaying data d not determined by the bordered diagram $\boldsymbol{\mathcal{D}}(d)$.

There is one further variant which we shall need to consider:

Definition 3.17 Suppose that $\mathcal{D}$ is a link diagram, and that d is a set of stabilisation data for $\mathcal{D}$. Consider the corresponding splayed diagram

$$\mathcal{D}(d, \vec{\iota}) := (\Sigma(d), \boldsymbol{\alpha}^1(d), \dots, \boldsymbol{\alpha}^m(d), \boldsymbol{\beta}(d), \boldsymbol{p}(d)).$$

We note that each $\boldsymbol{\alpha}^i(d)$ is disjoint from the points e^{∞}. It follows that there is an induced set of sets of curves in the surface $\boldsymbol{\Sigma}(d)$, given by the images of the $\boldsymbol{\alpha}^i(d)$ under the map $\cap_{\boldsymbol{D}}$. As such, there is a corresponding bordered multidiagram $\mathcal{D}^{\partial}(d, \vec{\iota})$ given by

$$\mathcal{D}^{\partial}(d, \vec{\iota}) := (\Sigma(d, \vec{\iota}), \boldsymbol{\alpha}^a(d, \vec{\iota}) \cup \boldsymbol{\alpha}^c(d, \vec{\iota}), \boldsymbol{\alpha}^1(d, \vec{\iota}), \dots, \boldsymbol{\alpha}^m(d, \vec{\iota}), \boldsymbol{\beta}(d, \vec{\iota}), \boldsymbol{p}(d, \vec{\iota})).$$

(We often relabel so that $\Sigma(d, \vec{\iota}) = \Sigma(d)$, etc.)

We call this the *partial splaying of* $\mathcal{D}$ associated to the data $d, \vec{\iota}$.

Later, we will need to perturb subdiagrams into a more standard form:

Definition 3.18 Suppose that $\mathcal{D}$ is a splayed or partially splayed diagram, and consider a subdiagram $\mathcal{D}'$ with attaching curves $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{i_1}, \dots, \boldsymbol{\gamma}^{i_k}$. Let $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^{i'_1}, \dots, \boldsymbol{\gamma}^{i'_{k'}}$ denote the subsequence of these attaching curves obtained by omitting any $\boldsymbol{\gamma}^{i_j}$ such that $\boldsymbol{\gamma}^{i_{j-1}} \equiv \boldsymbol{\gamma}^{i_j}$.

Each set of curves $\boldsymbol{\gamma}^{i'_j}$ is given either by $H(\boldsymbol{\alpha}(d, \vec{\iota}^{\,i'_j}, i'_j\varepsilon/m)$ for some sequence of idempotents $\vec{\iota}^{\,i'_j}$ or is equal to $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$ — say there are m' curves of the first sort. We let $\boldsymbol{\gamma}_{\mathrm{reg}}$ denote the sequence of attaching curves obtained from $\boldsymbol{\gamma}'$ by replacing each $\boldsymbol{\gamma}^{i'_j}$ of the first sort with the perturbation

$$\boldsymbol{\gamma}^{i'_j}_{\mathrm{reg}} := H(\boldsymbol{\alpha}(d, \vec{\iota}^{\,i'_j}, j\varepsilon/m')),$$

and $\mathcal{D}'_{\mathrm{reg}}$ be the diagram obtained from $\mathcal{D}'$ by replacing the attaching curves used with $\boldsymbol{\gamma}_{\mathrm{reg}}$.

We shall call this diagram the *regularisation* of $\mathcal{D}'$.

The bordered stabilisation and splaying of a Heegaard diagram exhibit a number of similarities.

Let $\mathcal{D} := (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{p})$ be a link diagram, and $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$ some stabilisation data for $\mathcal{D}$. Recall that associated to the curves $\boldsymbol{\alpha}$ are the m-fold ε approximations $\boldsymbol{\alpha}^1, \dots, \boldsymbol{\alpha}^m$.

The image of $\boldsymbol{\alpha}^i$ under the map glue coincides with a set of $g-1$ of the closed stabilisation curves $\boldsymbol{\alpha}^i(d) \subset \mathcal{D}(d) = (\Sigma(d); \boldsymbol{\alpha}^1(d), \dots, \boldsymbol{\alpha}^m(d); \boldsymbol{\beta}(d); \boldsymbol{p}(d))$, or similarly in the partial splaying of $\mathcal{D}$. We number the $\boldsymbol{\alpha}^i(d)$ so that these $g-h$ curves are given by the set $\{\alpha^i_{h+1}(d), \dots, \alpha^i_{g+h-1}(d)\} \subset \boldsymbol{\alpha}^i(d)$. It follows that to any distinguished generators $\vec{\boldsymbol{\theta}}_{\mathrm{approx}} = \boldsymbol{\theta}^{1,2}_{\mathrm{approx}}, \dots, \boldsymbol{\theta}^{m-1,m}_{\mathrm{approx}}$ for the m-fold ε-approximation of $\boldsymbol{\alpha}$ we may associate a corresponding set of partial generators $\mathrm{glue}(\vec{\boldsymbol{\theta}}_{\mathrm{approx}})$ for the splayed or partially splayed diagram $\mathcal{D}(d)$, which we also denote by $\vec{\boldsymbol{\theta}}_{\mathrm{approx}}$.

The curves in each $\boldsymbol{\alpha}^i(d)$ not in the image of $\mathrm{glue}(\boldsymbol{\alpha}^i)$, which we assume are labelled by $\alpha^i_1, \dots, \alpha^i_h$ such that, for each i either:

- $\alpha^i_j \cap \alpha^{i+1}_j$ is a single point $e^{\infty,i}_j$ (this corresponds with the idempotents ι^i_j and ι^{i+1}_j being different).
- α^i_j and α^{i+1}_j are Hamiltonian isotopic (in fact, they are ε/m-small approximations of one another).

In the second case, there are two points in $\alpha^i_j \cap \alpha^{i+1}_j$. For every such i, j, we choose an element $\theta^{i,i+1}_j \in \alpha^i_j \cap \alpha^{i+1}_j$. We term such a choice a *splaying extension* of the distinguished generators $\vec{\boldsymbol{\theta}}_{\mathrm{approx}}$. For every fixed i, we collect the intersection points $e^{\infty,i,i+1}_j$ and $\theta^{i,i+1}_j$ into a set $\boldsymbol{\theta}^{i,i+1}_{\mathrm{ext}}$, and set $\boldsymbol{\theta}^{i,i+1} = \boldsymbol{\theta}^{i,i+1}_{\mathrm{ext}} \cup \boldsymbol{\theta}^{i,i+1}_{\mathrm{approx}}$ — this is a bona fide set of generators dependent upon both the choice of distinguished generators and the choice of splaying extension.

Example 3.19 Suppose that $\vec{\boldsymbol{\theta}}_{\mathrm{approx}}$ is the distinguished set of generators $\vec{\boldsymbol{\theta}}_+$ described in Example 3.4. Then we shall call the corresponding set of $\boldsymbol{\theta}^{i,i+1}$ by the same name $\boldsymbol{\theta}^{i,i+1}_+$ — and hope that this does not cause confusion.

If we wish to draw attention to the number of generators in $\vec{\boldsymbol{\theta}}_+$, we will write $\vec{\boldsymbol{\theta}}_+(k)$, where $k = |\vec{\boldsymbol{\theta}}_+(k)|$.

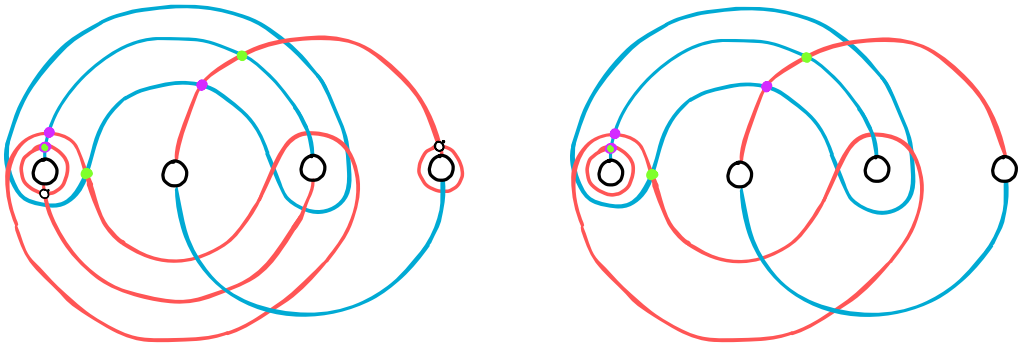


Figure 8: The splayings of the diagram in Figure 3 with splaying data d as stated in Example 3.21. On the left is shown the bordered stabilisation with a pair of generators $\boldsymbol{x}$ (violet) and $\boldsymbol{y}$ (lime green); on the right is shown the splaying by the length one sequence of pairs of idempotents $((\iota_m, \iota_l))$ together with the corresponding splayed generator.

Recall there are corresponding generators $(\boldsymbol{x}', \boldsymbol{y}')$ in the splayings $\mathcal{D}(0, \iota(\boldsymbol{x}))$ and $\mathcal{D}(0, \iota(\boldsymbol{y}))$, given by the map $\cap_D^{-1}$ (see the discussion following Definition 3.15). The nearest point map gives a generator $\bar{\boldsymbol{y}}$ corresponding with $\boldsymbol{y}'$ in the set of generators $\mathcal{G}(\boldsymbol{\alpha}^m, \boldsymbol{\beta})$; it follows that the tuple $(\boldsymbol{x}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,m}, \bar{\boldsymbol{y}})$ constitutes a sequence of generators for the splayed or partially splayed diagram $\mathcal{D}(\varepsilon, \vec{\iota})$. We call this sequence the *splaying* of $(\boldsymbol{x}, \boldsymbol{y})$.

Remark 3.20 For sufficiently small ε, for each j all of the points $e_j^{\infty,i}$ may all be assumed to be contained in a small disc neighbourhood N_j of e_j^{∞} which is disjoint from the curves $\boldsymbol{\beta}(d)$.

Example 3.21 Figures 8, 9 and 10 show the bordered stabilisation of the diagram in Figure 3 by the data $d = (\boldsymbol{\lambda}_{w,z}, \boldsymbol{\lambda}_{z,w}, \boldsymbol{n}, \varepsilon, \vec{\iota})$, where $\boldsymbol{\lambda}_{w,z}$ and $\boldsymbol{\lambda}_{z,w}$ are as given in Figure 3, $\boldsymbol{n} = (0, 0)$, ε isn't pictured, and $\vec{\iota}$ is stated below each figure. Also demonstrated are a series of generators, and their respective splayings.

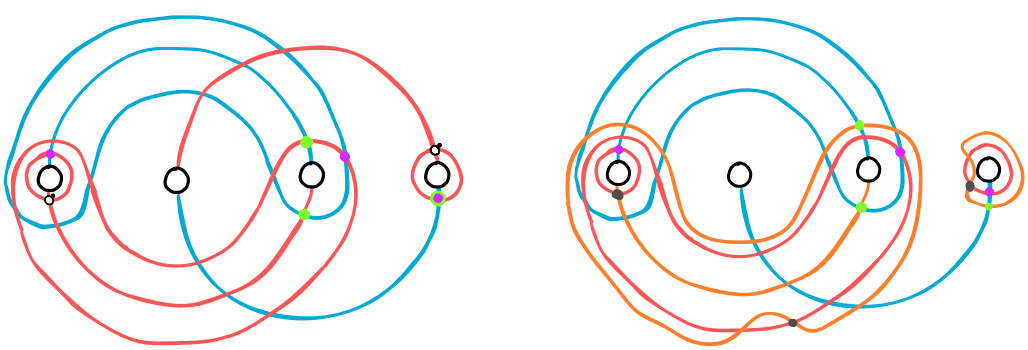


Figure 9: The splayings of the diagram in Figure 3 with splaying data d as stated in Example 3.21. On the left is shown the bordered stabilisation with a pair of generators $\boldsymbol{x}$ (violet) and $\boldsymbol{y}$ (lime green); on the right is shown the splaying by the length two sequence of pairs of idempotents $((\iota_m, \iota_m), (\iota_l, \iota_m))$, together with the corresponding splayed generators. Here, the distinguished generator is shown in dark grey.

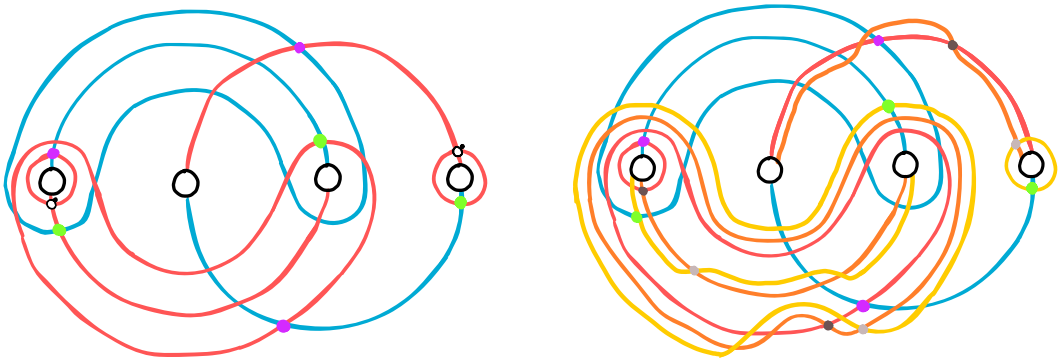


Figure 10: The splayings of the diagram in Figure 3 with splaying data d as stated in Example 3.21. On the left is shown the bordered stabilisation with $\boldsymbol{\beta}$ in blue and $\boldsymbol{\alpha}$ in red, with a pair of generators $\boldsymbol{x}$ (violet) and $\boldsymbol{y}$ (lime green). On the right is shown the splaying by the length three sequence of pairs of idempotents $((\iota_m,\iota_l),(\iota_l,\iota_l),(\iota_l,\iota_m))$, where $\boldsymbol{\beta}$ is blue, $\boldsymbol{\alpha}_1$ is red, $\boldsymbol{\alpha}_2$ is orange and $\boldsymbol{\alpha}_3$ is yellow, together with the corresponding splayed generators. Here, the distinguished generators are shown in dark grey and light grey respectively.

Convention 3.22 From here on, we will label the attaching curves of a splayed or partially splayed diagram $\mathcal{D}(\varepsilon,\vec{\iota})$ or $\mathcal{D}^{\partial}(\varepsilon,\vec{\iota})$ by $\boldsymbol{\gamma}^0,\dots,\boldsymbol{\gamma}^m$, where $\boldsymbol{\gamma}^0=\boldsymbol{\beta}$, and the curves $\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^m$ are given by the splayed curves $\boldsymbol{\gamma}^i=\boldsymbol{\alpha}^i(d)$, or, for a partially splayed diagram, $\boldsymbol{\gamma}^1=\boldsymbol{\alpha}=\boldsymbol{\alpha}^a\cup\boldsymbol{\alpha}^c$ and the subsequent curves are the closed curves $\boldsymbol{\gamma}^i=\boldsymbol{\alpha}^i(d)$.

3.1 Domains

Suppose that $\mathcal{D}=(\Sigma;\boldsymbol{\gamma}^0,\dots,\boldsymbol{\gamma}^m;\boldsymbol{b};\boldsymbol{p})$ is a diagram. In the next chapter, we will be interested in maps

$$u\colon (S,\partial S)\to(\Sigma\times D_{m+1},\boldsymbol{\gamma}^0\times e^0\cup\cdots\cup\boldsymbol{\gamma}^m\times e^m)$$

where S is a Riemann surface with boundary and punctures upon its boundary, and D_{m+1} is an $(m+1)$-gon — that is, a disc with $m+1$ boundary punctures $v^{0,1},\dots,v^{m,0}$ (see Section 4.1) — with sides $e^0,\dots,e^m$. Given any such u, one can postcompose with the projection maps $\pi_\Sigma\colon\Sigma\times D_{m+1}\to\Sigma$ and $\pi_D\colon\Sigma\times D_{m+1}\to D_{m+1}$ to yield corresponding maps u_Σ and u_D.

Near any boundary puncture q of S that is mapped to $v^{i,i+1}$ under u_D, the maps u which we study will be asymptotic to a point $I^{i,i+1}(q)$ of the form

$$I^{i,i+1}(q):=x^{i,i+1}(q)\times v^{i,i+1},$$

for some point $x^{i,i+1}(q)\in\boldsymbol{\gamma}^i\cap\boldsymbol{\gamma}^{i+1}$, such that the set

$$\boldsymbol{x}^{i,i+1}:=\bigcup_{q\in u_D^{-1}(v^{i,i+1})}x^{i,i+1}(q)$$

constitutes a generator for $(\Sigma,\boldsymbol{\gamma}^i,\boldsymbol{\gamma}^{i+1})$ — that is, the sequence $\mathcal{G}(u):=(\boldsymbol{x}^{0,1},\dots,\boldsymbol{x}^{m,0})$ is a sequence of generators for $\mathcal{D}$.

We label the following objects:

$$\boldsymbol{C}^i := \boldsymbol{\gamma}^i \times e^i, \quad \boldsymbol{I}^{i,i+1} := \bigcup_{q \in u_D^{-1}(v^{i,i+1})} I^{i,i+1}(q), \quad S^\partial := (\partial\Sigma \setminus \boldsymbol{b}) \times e^0.$$

It will follow from our assumptions on u that there is an associated map $\bar{u}$ which represents some class in the relative homology group

$$H_2\big(\overline{\Sigma} \times I_s \times I_t, (\boldsymbol{C}^0 \cup \cdots \cup \boldsymbol{C}^m) \cup (\boldsymbol{I}^{0,1}(\boldsymbol{x}^{0,1} \cup \cdots \cup \boldsymbol{I}^{m,0}(\boldsymbol{x}^{m,0})) \cup S^\partial)\big).$$

We denote by $\pi_2(\boldsymbol{x}^{0,1}, \ldots, \boldsymbol{x}^{m,0})$ the set of all elements of the above group which map to the relative fundamental class of $\boldsymbol{I}^{0,1}(\boldsymbol{x}^{0,1}) \cup \cdots \cup \boldsymbol{I}^{m,0}(\boldsymbol{x}^{m,0})$ under the composition of the boundary homomorphism and collapsing the remainder of the boundary.

A related notion is that of a *domain*. Namely, if $\mathcal{D}$ is a diagram, then a *domain* is an element of the homology group

$$\mathrm{Dom}(\mathcal{D}) := H_2(\overline{\Sigma}, \boldsymbol{\gamma}^1 \cup \cdots \cup \boldsymbol{\gamma}^m \cup \partial\overline{\Sigma}).$$

This homology group is generated by the regions R_i of $\Sigma - \boldsymbol{\gamma}$, and thus any domain is given by a formal linear combination $\sum r_i R_i$.

A *periodic* domain is a domain with zero multiplicity (coefficient) at all regions R_i where $R_i \cap (\boldsymbol{z} \cup \boldsymbol{b}) \neq \varnothing$ and with boundary equal to a linear combination of (full) curves and arcs from $\boldsymbol{\gamma}$. A *provincial periodic domain* is a domain which further has zero multiplicity at all regions R_i such that $R_i \cap \partial\Sigma \neq \varnothing$.

We say a diagram $\mathcal{D}$ is *admissible* if every periodic domain in $\mathcal{D}$ has both positive and negative coefficients; we say it is *provincially admissible* if every provincial periodic domain in $\mathcal{D}$ has both positive and negative coefficients.

Any homology class B in the group $\pi_2(\boldsymbol{x}^{0,1}, \ldots, \boldsymbol{x}^{m,0})$ can be projected under π_Σ to an element of $\mathrm{Dom}(\mathcal{D})$, which we call its *domain*. One can see that a homology class is uniquely determined by the corresponding domain; thus we often conflate the two notions.

Given a sequence of Reeb chords $(\rho^1, \ldots, \rho^k)$ of the torus algebra, there is an associated homology class $[\vec{\rho}] = [\rho^1] + [\rho^2] + \cdots + [\rho^k]$, where $[\rho^i]$ is the image of the Reeb chord $\vec{\rho}^i$ in the relative homology group $H_1(S^1, \boldsymbol{a})$. If $\mathcal{D}$ is a toroidal Heegaard multidiagram with h boundary components, each of the relative homology groups $H_1(\partial\Sigma_j, \boldsymbol{a}_j)$ are identified with $H_1(S^1, \boldsymbol{a})$.

Definition 3.23 Suppose that $B \in \pi_2(\boldsymbol{x}^{0,1}, \ldots, \boldsymbol{x}^{m,0})$ is a domain. We shall say that B is *$\vec{\rho}$-compatible* for $\vec{\rho} = (\vec{\rho}_1, \ldots, \vec{\rho}_h)$ if, for each j, we have that the image of B in $H_1(\partial\Sigma_j, \boldsymbol{a}_j)$ is equal to $[\vec{\rho}_j]$.

Convention 3.24 When $\mathcal{D}$ is the splaying of a diagram $\boldsymbol{\mathcal{D}}$, if $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{G}(\boldsymbol{\mathcal{D}})$ are generators, we shall often abbreviate the corresponding group

$$\pi_2(\boldsymbol{x}', \boldsymbol{\theta}_+^{1,2}, \ldots, \boldsymbol{\theta}_+^{m-1,m}, \boldsymbol{y}')$$

as $\pi_2^{\boldsymbol{\theta}^+}(\boldsymbol{x}, \boldsymbol{y}) \subset \mathrm{Dom}(\mathcal{D}(d))$.

We end this subsection with a brief discussion of generators and domains in east-compactified diagrams, which will be useful in Section 4.6.1. Recall that to any diagram $\mathcal{D}$ is associated a corresponding east-compactified diagram $\mathcal{D}_{\bar{e}}$. Generators for $\mathcal{D}$ have obvious counterparts in $\mathcal{D}_{\bar{e}}$, and we generally abuse notation and call these by the same label; any homology class B for $\mathcal{D}$ has a corresponding east-compactified class $B_{\bar{e}}$ in the relative homology group

$$H_2\big(\Sigma_{\bar{e}} \times D_n, ((\boldsymbol{\gamma}^1_{\bar{e}} \times e^1) \cup \cdots \cup (\boldsymbol{\gamma}^m \times e^m)) \cup (\boldsymbol{I}^{0,1} \times \cdots \times \boldsymbol{I}^{m-1,0})\big).$$

3.1.1 Splaying domains We would like to understand the relationship between domains in a bordered diagram $\mathcal{D} = (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{b})$ and a corresponding (partial) splaying $\mathcal{D}^\partial(\varepsilon, \vec{\iota})$. First, we set up some notation for describing different qualities of sequences of Reeb chords.

Fix a stable bordered Heegaard diagram $\boldsymbol{\mathcal{D}}$, generators $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{G}(\mathcal{D})$, a set of sequences of Reeb chords $\vec{\rho}_{\clubsuit} = \vec{\rho}_{\clubsuit,1}, \ldots, \vec{\rho}_{\clubsuit,h}$, and a sequence of idempotents $\vec{\iota}$. Consider the partially splayed diagram $\mathcal{D}^\partial(\varepsilon, \vec{\iota})$ and the splayed diagram $\mathcal{D}(\varepsilon, \vec{\iota})$. We shall associate to any $\vec{\rho}_{\clubsuit}$-compatible domain $B \in \pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x}, \boldsymbol{y}) \subset \mathrm{Dom}(\mathcal{D}^\partial(\varepsilon, \vec{\iota}))$, or any domain $B \in \pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x}, \boldsymbol{y}) \subset \mathrm{Dom}(\mathcal{D}(\varepsilon, \vec{\iota}))$ a sequence of Reeb chords $\vec{\rho}(B)$.

Convention 3.25 Recall from the beginning of the section that any homology class in $\mathrm{Dom}(\mathcal{D})$ corresponds with a homology class in $\mathrm{Dom}(\mathcal{D}_{\bar{e}})$ — we make this identification for the rest of this section, and consider domains in diagrams where the underlying surface is closed.

Consider a small neighbourhood N_j of $e_j^\infty \in \Sigma_{\bar{e}}$, which is disjoint from $\boldsymbol{\beta}$ and such that for every i, $e_j^{\infty,i}$ is contained in N_j, as in Remark 3.20. Suppose that $B \in \pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x}, \boldsymbol{y})$ is a domain, and that u is a map from a surface with boundary and punctures upon the boundary to $\Sigma \times D_m$ or $\Sigma_{\bar{e}} \times D_m$, with homology class B. If S_j denotes the restricted source $(\pi_\Sigma \circ u)^{-1}(N_j)$, then the map $u'_j = u|_{S_j}$ takes the form of a map

$$u'_j : (S_j, \partial S_j) \to (N_j \times D_{m+1}, \partial N_j \cup \gamma_j^1 \times e^1 \cup \cdots \cup \gamma_j^m \times e^m).$$

Maps of this form are in bijection with maps

$$u_j : (S_j, \partial S_j) \to (N_j \times D_{m+1}, \partial N_j \cup \delta_j^1 \times e^1 \cup \cdots \cup \delta_j^{m'} \times e^{m'}),$$

where $\delta_j^1, \ldots, \delta_j^{m'}$ is the subsequence of $\gamma_j^1, \ldots, \gamma_j^m$ formed by deleting every arc γ_j^i such that γ_j^i approximates γ_j^{i-1}. On the level of domains, the map u_j has an associated domain

$$B_j \in H_2(N_j, \partial N_j \cup (\delta_j^1 \cup \cdots \cup \delta_j^{m'}))$$

which can be described by choosing a point $x_{\boldsymbol{R}}$ for every region of $N_j - (\gamma_j^1 \cup \cdots \cup \gamma_j^m)$ not contained in the approximation regions between any two adjacent curves $\gamma_j^i, \gamma_j^{i+1}$ which approximate one another: B_j is the domain given by requiring that the multiplicity of B_j at each $x_{\boldsymbol{R}}$ is the same as the multiplicity of $B|_{N_j}$ at each $x_{\boldsymbol{R}}$.

In what follows, it will help us to have an understanding of the homology groups $H_2(N_j, \partial N_j \cup \delta_j^i)$ and $H_2(N_j, \partial N_j \cup \delta_j^i \cup \delta_j^{i+1})$.

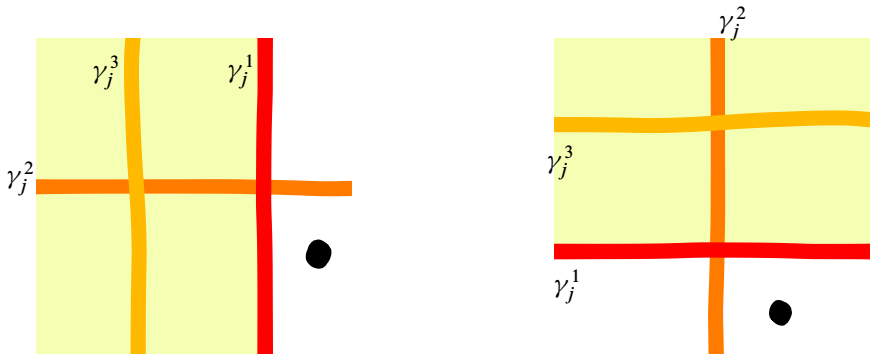


Figure 11: A series of examples of various domains (shaded green) in the neighbourhood N_j of e_j^∞. On the left is an example of $D^1(\rho_{12})$, and on the right is $D^1(\rho_{23})$.

The pair (N_j, δ_j^i) is modelled upon either the pair $(\mathbb{D}, \Re)$ or $(\mathbb{D}, \Im)$ depending upon whether δ_j^i is meridional or longitudinal. In these cases, the groups $H_2(\mathbb{D}, \Re)$ and $H_2(\mathbb{D}, \Im)$ are generated by the half-discs $\mathbb{D}-\Re$ and $\mathbb{D}-\Im$ respectively, only one of which has multiplicity zero at the basepoint p_j in each case. We label these homology classes $D(\rho_{12})$ and $D(\rho_{23})$ respectively, and the corresponding domain under the identification with (N_j, δ_j^i) by $D_j^i(\rho_{12})$ and $D_j^i(\rho_{23})$ respectively (see Figure 11). We generalise this to length n_i repeated sequences of these chords, by setting $D_j^i(\rho_{12}, \dots, \rho_{12}) := n_i \cdot D_j^i(\rho_{12})$, and similar for ρ_{23} — in both instances, we allow the degenerate length 0 sequence with associated domain equal to zero. If δ_j^i and δ_j^{i+1} are not small approximations of one another, there are two models for the ordered triple $(N_j, \delta_j^i, \delta_j^{i+1})$ depending upon the configuration of the curves:

- $(\mathbb{D}, \Im, \Re)$ if δ_j^i is meridional and δ_j^{i+1} longitudinal;
- $(\mathbb{D}, \Re, \Im)$ if δ_j^i is longitudinal and δ_j^{i+1} is meridional.

$H_2(\mathbb{D}, \Re \cup \Im)$ is generated by four quadrants $B(\rho_0)$, $B(\rho_1)$, $B(\rho_2)$ and $B(\rho_3)$, where $B(\rho_0)$ is the quadrant of $(\mathbb{D}, \Re \cup \Im)$ containing the point p_j, and the others are given in the order they are encountered clockwise from $B(\rho_0)$ (see Figure 12). There is a natural identification $\varphi\colon H_1(S^1, \boldsymbol{a}_j) \to H_2(\mathbb{D}, \Re \cup \Im)$, and we denote $B(\rho) := \varphi(\rho)$. We denote by $B_j^i(\rho)$ the corresponding domain in $H_2(N_j, \partial N_j \cup (\delta_j^i \cup \delta_j^{i+1}))$. (Please see Figure 13 for some examples.)

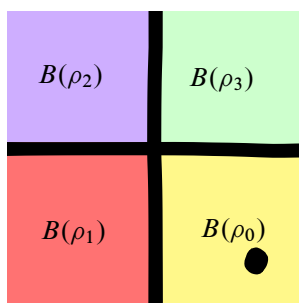


Figure 12: The four quadrants of $\mathbb{D}$.

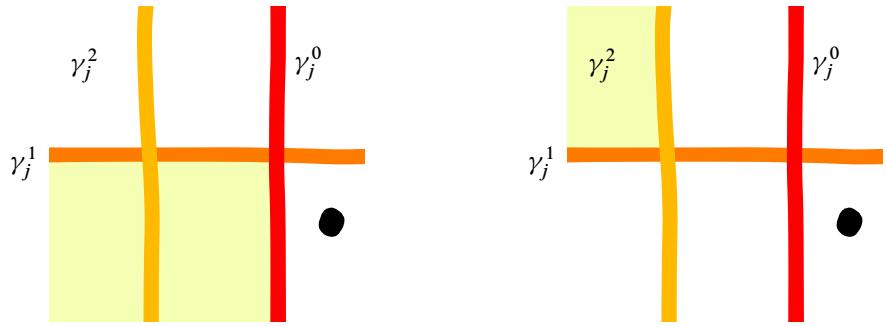


Figure 13: A series of examples of various domains (shaded green) in the neighbourhood N_j of e_j^∞. On the left is an example of $B^1(\rho_1)$, and on the right is $B^2(\rho_2)$. We note how the sum of these domains has an obtuse corner at the intersection of γ_j^1 and γ_j^2.

We also extend this assignment to the case where $\mathcal{D}$ is partially splayed and $\boldsymbol{\gamma}^1 \cap N_j = (\gamma_j^{1,m} \cup \gamma_j^{1,l}) \cap N_j$. Noting that $(N_j, \boldsymbol{\gamma}^1 \cap N_j)$ may be identified with $(\mathbb{D}, \Im \cup \Re)$, we may associate to a sequence of Reeb chords $\vec{\rho}$ the corresponding domain $B(\vec{\rho})$ identified with $\varphi(\vec{\rho})$, in the homology group

$$H_2(N_j, \partial N_j \cup (\boldsymbol{\gamma}^1 \cap N_j)).$$

Recall that two sequences of Reeb multichords, $\vec{\rho}$ and $\vec{\rho}'$, are 123-equivalent if their corresponding 123-composed sequences agree. We introduce a related notion for domains:

Definition 3.26 We shall say that a sequence of domains

$$S = (D_j^1, B_j^1, D_j^2, B_j^2, \dots, D_j^k, B_j^k, D_j^{k+1}),$$

is a *sequence of Reeb domains* if each D_j^i is equal to $D_j^i(\vec{\rho}^i)$ for some repeated sequence $\vec{\rho}^i := (\rho^i, \rho^i, \dots, \rho^i)$ of a nonjumping chord $\rho^i \in \mathcal{T}$, and each B_j^i is equal to $B_j^i(\underline{\rho}^i)$ for some jumping chord $\underline{\rho}^i \in \mathcal{T}$. For such a sequence, denote by $\vec{\rho}(S)$ the sequence of Reeb chords

$$(\vec{\rho}^1 \cup \underline{\rho}^1 \cup \vec{\rho}^2 \cup \cdots \cup \underline{\rho}^k \cup \vec{\rho}^k).$$

If S' is another sequence of Reeb domains, we shall say that S and S' are *composable-equivalent* if the homology classes in $H_2(N_j, \partial N_j \cup \delta_j^1 \cup \cdots \cup \delta_j^{m'})$ formed by summing S and S' are the same.

As the name suggests, this relation is closely related to the composable-equivalence of corresponding Reeb chords:

Proposition 3.27 *Suppose S and T are sequences of Reeb domains. Then they are composable-equivalent if and only if the sequences $\vec{\rho}(S)$ and $\vec{\rho}(T)$ are 123-equivalent.*

Proof Suppose that S is a sequence of Reeb domains and consider the corresponding sequence of Reeb chords $\vec{\rho}(S)$. One can see there is a unique sequence $U(S)$ of Reeb domains that is composable-equivalent to S, and such that $\vec{\rho}(U(S))$ is the 123-reduction of $\vec{\rho}(S)$.

Indeed, $\vec{\rho}(S)$ is not 123-reduced if and only if there are two adjacent chords $\rho_k, \rho_{k+1} \in \vec{\rho}(S)$ that are either equal to the torus algebra chords ρ_1, ρ_{23} or ρ_{12}, ρ_3. In the first case, if $B_j^i = B^i(\rho_1)$ and $D_j^{i+1} = n_{i+1} B^{i+1}(\rho_{23})$ are the corresponding domains in S then one can see that the sequence S' given by replacing B_j^i by $B^i(\rho_{123}) = B^i(\rho_1) + B^i(\rho_{23})$ and D_j^{i+1} by $(n_{i+1}-1)B^{i+1}(\rho_{23})$ is composable-equivalent to S, and the corresponding $\vec{\rho}(S')$ differs from $\vec{\rho}(S)$ by replacing ρ_k, ρ_{k+1} with $\rho_k \cdot \rho_{k+1}$. Similar holds when $(\rho_k, \rho_{k+1}) = (\rho_{12}, \rho_3)$. Repeating this process for any 123-composable chords yields the $U(S)$ required.

It follows that S and T are composable-equivalent if and only if $U(S) = U(T)$, which holds if and only if the 123-reductions of $\vec{\rho}(S)$ and $\vec{\rho}(T)$ agree, and so the result follows. □

Proposition 3.28 *Suppose that $B \in \pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x}, \boldsymbol{y})$ is a domain which carries a holomorphic representative, and is $\vec{\rho}_{\clubsuit}$-compatible. Then for every $j = 1, \dots, h$, the domain $B_j - B(\vec{\rho}_{\clubsuit,j})$ may be written uniquely, up to composable-equivalence, as a sequence of Reeb domains.*

Proof We construct an explicit sequence of Reeb domains S inductively. If there is only one set of $\boldsymbol{\gamma}$-curves, B_j is an element of the group $H_2(N_j, \partial N_j \cup \delta_j^1)$, which, as discussed, is equal to $n_1 B_j^1(\rho)$ for some nonjumping $\rho \in \mathcal{T}$.

Consider a small disc neighbourhood M_j^1 of $e_j^{\infty,1}$, disjoint from all other $e_j^{\infty,i}$ and e_j^{∞}. The restriction of the domain of B to M_j^1 is modelled upon $n_1 B(\rho_1) + n_2 B(\rho_2) + n_3 B(\rho_3)$ for some $n_1, n_2, n_3 \geq 0$. There are three cases:

- If γ_j^2 is meridional, write $D_j^1 := n_3 B_j^1(\rho_{23})$ and $B_j^1 := B_j^1(\rho_2)$.
- If γ_j^2 is longitudinal and $n_1 > n_2$, write $D_j^1 := n_2 B_j^1(\rho_{12})$ and $B_j^1 := B_j^1(\rho_1)$.
- If γ_j^2 is longitudinal and $n_1 \leq n_2$, write $D_j^1 : n_1 B_j^1(\rho_{12})$ and $B_j^1 := B_j^1(\rho_3)$.

It follows that the domain $B_j' = B_j - (D_j^1 + B_j^1)$ has the same multiplicity upon either side of γ_j^2. As such, B_j' is an element of $H_2(N^j, \partial N^j \cup (\delta_j^3 \cup \cdots \cup \delta_j^k))$, and the result follows by induction. □

With this in mind, we may make a definition:

Definition 3.29 Let $B \in \pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x}, \boldsymbol{y})$ be a domain which carries a holomorphic representative u, and let S_j be a decomposition of each B_j into Reeb domains given by the construction in Proposition 3.28. We denote by $[\vec{\rho}(B)_j]$ the 123-equivalence class of Reeb multichords defined by $[\vec{\rho}(B)_j] := [\vec{\rho}(S_j)]$. We often collect these together, writing $[\vec{\rho}(B)] = ([\vec{\rho}(B)_1], \dots, [\vec{\rho}(B)_h])$.

By construction, $[\vec{\rho}(B)_1], \dots, [\vec{\rho}(B)_h]$ satisfies that for any set of representatives $\vec{\rho}(B)_1, \dots, \vec{\rho}(B)_h$ of each corresponding 123-equivalence class, the sequence of idempotents $\vec{\iota}_j := \iota_j^1, \dots, \iota_j^{m+1}$ has

$$|\{\iota_j^i \in \vec{\iota}_j \mid \iota_j^i \neq \iota_j^{i-1}\}| = |\mathrm{jump}(\vec{\rho}_j)|,$$

as in Definition 2.9. As such:

Definition 3.30 If $B \in \mathrm{Dom}(\mathcal{D}(\varepsilon,\vec{\iota}))$ is a domain, then there is a well-defined splicing of $\vec{\rho}(B)_1,\dots,\vec{\rho}(B)_h$ given by $\vec{\sigma}(B) := \vec{\sigma}(\vec{\rho},\vec{\iota})$.

We shall later be interested in domains which look similar between splayed diagrams. A useful construction to bear in mind (cf [13, Definition 3.34]) is:

Definition 3.31 Fix an h-splaying sequence of idempotents $\vec{\iota}$, and consider an additional tuple of idempotents $\iota = (\iota_1,\dots,\iota_h)$. Let $\vec{\iota}' = (\iota,\vec{\iota})$ and consider the partially splayed diagrams $\mathcal{D}^\partial(\varepsilon,\vec{\iota})$ and $\mathcal{D}^\partial(\varepsilon,\vec{\iota}')$; we label the curves in each by $\boldsymbol{\gamma}^0,\dots,\boldsymbol{\gamma}^m$ and $\boldsymbol{\delta}^0,\dots,\boldsymbol{\delta}^{m+1}$ respectively. For every $i>1$, the curves $\boldsymbol{\gamma}^i$ and $\boldsymbol{\delta}^{i+1}$ are small approximations of one another (with associated approximation region R^i, and the curves $\boldsymbol{\gamma}^2$ are small approximations of some subset of the curves $\boldsymbol{\gamma}^1_{\bar{e}}$ in the east-compactification $\mathcal{D}^\partial(\varepsilon,\vec{\iota})|_{\bar{e}}$ (with associated approximation region R^1).

For every curve $\gamma^i_j \in \boldsymbol{\gamma}^i$, choose a pair of basepoints on either side of γ^i_j such that the basepoints are disjoint from each of the approximation regions R^i. For any

$$B' \in \pi_2(\boldsymbol{x},\boldsymbol{\theta}^{1,2},\dots,\boldsymbol{\theta}^{m-1,m},\boldsymbol{y}) \subset \mathrm{Dom}(\mathcal{D}^\partial(\varepsilon,\vec{\iota}')),$$

we let $\phi(B')$ denote the unique domain

$$B \in \pi_2(\boldsymbol{x},\boldsymbol{\theta}^{2,3},\dots,\boldsymbol{\theta}^{m-1,m},\boldsymbol{y}) \subset \mathrm{Dom}(\mathcal{D}^\partial(\varepsilon,\vec{\iota}))$$

with the same multiplicities at each of these basepoints as B' (here we abuse notation and write the same symbols for generators which are identified by the nearest-point map). We sometimes call $\phi(B')$ the domain obtained from B' by *taking multiplicities away from the approximation region*, and shall say that B' *splays* B if $\phi(B') = B$.

4 Moduli spaces of holomorphic curves and polygons

In this section, we define a series of moduli spaces which are required in the definition of the invariants $\mathrm{Poly}(\mathcal{D})$; we also examine the structure of these moduli spaces and establish results which restrict the possible elements which may occur in the boundaries of their compactifications, in order to show that $\mathrm{Poly}(\mathcal{D})$ satisfies the $\mathcal{A}_\infty$ relations, and also to understand the behaviour of $\mathrm{Poly}(\mathcal{D})$ under a neck-stretching argument which we use in Section 5 to identify $\mathrm{Poly}(\mathcal{D})$ with the bordered-sutured Heegaard Floer invariants of the bordered Heegaard diagram.

4.1 Complex structures on polygons

Before we describe the moduli spaces that the results of the paper will be based upon, we need to fix notation for moduli spaces of polygons. Most of this section follows [24, Chapter 9].

Let D_m denote a disc, with m punctures upon the boundary (we often view these as marked vertices). Choose a distinguished edge $e^0 \subset \partial D_m$, and label the remaining edges $e^1,\dots,e^{m-1}$ anticlockwise

from e^0; also label the vertices by $p^{i,i+1}$, so that the vertex $p^{i,i+1}$ lies between e^i and e^{i+1} (here, superscripts are taken modulo m).

Let $\mathrm{Conf}(D_m)$ denote the moduli space of equivalence classes of positively oriented complex structures upon the disc D_m, up to biholomorphism that preserves the labelling of the edges (and vertices). The moduli space $\mathrm{Conf}(D_m)$ has a natural compactification by trees of holomorphic polygons which we summarise now.

Definition 4.1 An *m-leaved tree* is a properly embedded planar tree $T \subset \mathbb{R}^2$ with $m+1$ semi-infinite edges, one of which is distinguished and called the *root* of T, and the others we call the *leaves*. We denote by $V(T)$ the set of vertices of T, and $\mathrm{Ed}^{\mathrm{int}}$ the set of interior (finite) edges of T.

A *flag* in T is an oriented pair of a vertex and adjacent edge, (V, e), oriented so that the edge e begins at V. We orient T so that every path from the root to a leaf is positively oriented. To every vertex v of T we can associate a series of flags, given by the pair (V, e) and oriented accordingly. One of these flags is *negative* with respect to V, in the sense that the orientation on e induced by T differs from that stemming from the flag (ie the edge points "into" V), and the others are *positive*. It follows that we can order the flags originating at any given vertex V by saying $f_0(V)$ is the unique negative flag at V, and ordering the remainder $f_1(V), \dots, f_{|V|-1}(V)$ so that they are encountered in order from $f_0(V)$, anticlockwise. We call this the *root-ordering* of the flags.

Suppose we are given an m-leaved tree T, and a collection of Riemann surfaces $S = \{S(V)\}_{V \in V(T)}$ with boundary and punctures on the boundary, such that the boundary punctures of $S(V)$ are in one-to-one correspondence with the flags $f(V)$. To each interior edge e in T connecting V_1 and V_2 there are thus two related boundary punctures: one in $S(V_1)$ corresponding to the flag (V_1, e), and one in $S(V_2)$ corresponding to the flag (V_2, e). There is an operation upon the surface S given by deleting a small open neighbourhood of each of these punctures, and identifying the two new resultant boundary arcs to yield a new surface. Performing this operation iteratively over all interior edges gives a connected (topological) surface, $S^{\cup}(T)$.

For disks, we require more predictable behaviour near the punctures.

Definition 4.2 Let D be a representative for some equivalence class in $\mathrm{Conf}(D_n)$. A choice of *strip-like ends* for D is a choice of open neighbourhood $N^{i,i+1}$ of each boundary puncture $p^{i,i+1}$ — such that $N^{i,i+1} \cap N^{j,j+1} = \varnothing$ for $i \neq j$ — together with a choice of a biholomorphism between each $N^{i,i+1}$ and the strip $\mathbb{R}_+ \times [0,1]$ for every $i \neq 0$, and between $N^{0,1}$ and the strip $\mathbb{R}_- \times [0,1]$.

Suppose T is a d-leaved tree, and that we are given a collection of Riemann surfaces $S = \{S(V)\}_{V \in V(T)}$, where each $S(V)$ represents some equivalence class in $\mathrm{Conf}(D_{|V|})$ and is endowed with a choice of strip-like ends (ie a tree of holomorphic discs with strip-like ends). We shall say that a map $\rho : \mathrm{Ed}^{\mathrm{int}}(T) \to (0, \infty)$ is a *length assignment* for T — thus giving T the structure of a metric graph. We often omit ρ from the notation.

We refine the construction of $S^{\cup}(T)$. There is a natural correspondence between the flags of each vertex V and the punctures of $D_{|V|}$, between the puncture $p^{i,i+1} \in D_{|V|}$ and the flag $f_i(V)$. When we delete neighbourhoods of punctures corresponding with a negative flag (V_1, e) and positive flag (V_2, e), we take these to be the open neighbourhoods identified by the choices of strip-like ends with $(-\infty, -\log(\rho(e))) \times [0,1] \subset \mathbb{R}_- \times [0,1]$ and $(\log(\rho(e)), \infty) \times [0,1] \subset \mathbb{R}_+ \times [0,1]$ respectively.

This construction yields an element D_T in some equivalence class of $\mathrm{Conf}(D_m)$ — an m-pointed disc with Riemann surface structure, with vertices labelled $p^{0,1}, \dots, p^{m-1,0}$ starting from the one corresponding with the root of T (and edges labelled accordingly). Moreover, there is a natural choice of strip-like ends upon D_T induced by those upon the punctures corresponding with the semi-infinite flags of T.

Allowing edge lengths to degenerate and take the value 0 allows us to describe the compactification of the moduli space $\mathrm{Conf}(D_m)$. If a given edge length is zero, we omit this edge from the cutting and pasting procedure described above — effectively cutting the edge into two semi-infinite edges, and thus the tree into more components, upon which we perform the gluing process individually. The resulting collection of discs is described by a tree $\overline{T}$, which is given by collapsing all internal edges of nonzero length, and assigning to each vertex $\bar{v}$ of $\overline{T}$ a $|\bar{v}|$-pointed disc in $\mathrm{Conf}(D_{|V|})$.

There is a canonical way to label the edges of $\overline{T}$ in a way compatible with those of T. We first label the semi-infinite flags of $\overline{T}$ by $(0,1), (1,2), \dots, (m,0)$ anticlockwise from the root of $\overline{T}$. The task now is to label the interior flags of $\overline{T}$ by labels compatible with these; in such a way that the edges of the corresponding disks D_V can be labelled by $0, 1, \dots, m$, so that an edge between punctures (i,k) and (k,j) is labelled by k. The way to do this is to first label the vertices which just have semi-infinite flags and one in-coming flag — in this case, the semi-infinite flags are already labelled anticlockwise by some contiguous subsequence $(j, j+1), (j+1, j+2), \dots, (j+k-1, j+k)$, so the root flag is forced to be labelled by $(j+k, j)$. We can then delete this vertex from $\overline{T}$ and replace the root flag with a semi-infinite flag labelled by $(j+k, j)$, and repeat the process until all edges of $\overline{T}$ are labelled.

Definition 4.3 Suppose that $\mathcal{D}$ is a diagram with m sets of attaching curves $\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$, and let T be a (possibly degenerate) m-leaved metric tree. Then a *generator assignment* for T is an assignment of a generator $\boldsymbol{x}^{i,j} \in \mathcal{G}(\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^j)$ to each edge of $\overline{T}$ labelled by (i,j). If $\vec{\boldsymbol{x}}$ is a generator assignment for T, we refer to the sequence of generators corresponding to the edges adjacent to a vertex $V \in \overline{T}$, ordered by root-ordering, by $\vec{\boldsymbol{x}}(V)$. Note that a generator assignment for T has a well-defined overall sequence of generators $\vec{\boldsymbol{x}}(T) \in \mathcal{G}(\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m)$, given by the set of generators labelling the exterior edges of T.

Given a generator assignment $\vec{\boldsymbol{x}}$, a *family of compatible domains* is a set $\{B_V\}_{V \in V(\overline{T})}$ of domains such that B_V is a member of the group $\pi_2(\vec{\boldsymbol{x}}(V))$. The juxtaposition $B(T) = \sum_{V \in V(\overline{T})} B_V$ is, accordingly, a member of the group $\pi_2(\vec{\boldsymbol{x}}(T))$.

We would like to work with bona fide representatives of each equivalence class for $\mathrm{Conf}(D_m)$, with strip-like ends. Once these are chosen, the compactification of the moduli space $\mathrm{Conf}(D_m)$ is then described by

the set $\bigcup_{T,\rho} D_T$, indexed over all d-leaved trees and all (possibly degenerate) length assignments ρ. As edge lengths in this space degenerate to zero, the surface D_T degenerates to a disjoint union of surfaces $\{D_V\}_{V\in V(\overline{T})}$.

To make this make sense, one must choose representatives D and choices of strip-like ends for every $d \le m$ and each equivalence class of $\mathrm{Conf}(D_d)$, so that at these boundary strata the strip-like ends chosen agree with those induced by those chosen upon D_d. This is possible; see, for instance, [24, Section 9g].

From this point, we fix such a set of choices of representatives for equivalence classes with strip-like ends such that this works (ie we have consistency on the boundary strata), and abuse notation by identifying elements of $\mathrm{Conf}(D_m)$ with these representatives. We also fix a choice of symplectic forms $\{\omega_j\}_{j\in\mathrm{Conf}(D_m)}$ where ω_j is compatible with j, so that if $(j_n)_{n=1}^{\infty} \in \mathrm{Conf}(D_m)$ is a sequence of complex structures converging to the collection $\{j_{\bar v}\}_{\bar v\in V(\overline{T})} \in \partial\overline{\mathrm{Conf}(D_m)}$, then $\omega_{j_n} \to \{\omega_{j_{\bar v}}\}_{\bar v\in V(\overline{T})}$.

We will also have need to fix identifications η^i of a small neighbourhood $N(e^i)$ of each edge e^i in the polygons $(D_m, j) \in \mathrm{Conf}(D_m)$ with $\mathbb{R}\times[0,\varepsilon)$, so that on each of these $N(e^i)$ the symplectic form ω_j agrees with the pullback of the standard symplectic form on $\mathbb{R}^2$. To make this make sense, we require that these identifications are continuous, consistent with the choice of strip-like ends, and consistent across boundary strata of $\mathrm{Conf}(D_m)$ (see [13, Section 3.3; 24, Section 10e], for instance).

We will later need to single out some vertices of our trees:

Definition 4.4 Let T be a tree. Then the set of *comb vertices* of $\overline{T}$ is the set of vertices of $\overline{T}$ visited by the path in $\overline{T}$ connecting the exterior edge labelled by $(0,1)$ (ie the root of $\overline{T}$) with the exterior edge labelled by $(1,2)$. The set of *spinal vertices* of $\overline{T}$ is the set of vertices of $\overline{T}$ which have a flag f_V labelled by (i,j) where one of i or j is equal to 0. The set of *easterly vertices* of $\overline{T}$ is the set of nonspinal vertices of $\overline{T}$.

If V is spinal, we associate to it its *storied tree* $T' \subset T$, given by the maximal connected subtree of T which contains V and no other spinal vertices (so, it contains V and a series of easterly vertices).

4.2 Sources

In this section, we define what shall be the sources of the various types of holomorphic curves considered in the remainder of this chapter. We first fix positive integers m and h, and a disc $D_m \in \mathrm{Conf}(D_m)$.

Definition 4.5 A *topological* (h,m)*-source* is a surface S, with boundary and punctures $\mathrm{punc}(S)$ together with a labelling $\lambda\colon \mathrm{punc}(S) \to \mathrm{punc}(D_m)\cup\mathcal{T}$ of the boundary punctures of S by either a puncture in ∂D_m or a Reeb chord in $\mathcal{T}$. We will write $E(S)$ for the subset of "east" punctures $\lambda^{-1}(\mathcal{T})$, and also require that $E(S)$ is partitioned into h subsets $E_1(S),\dots,E_h(S)$.

An (h,m)*-source* is a topological (h,m)-source together with a choice of complex structure j_S which gives S the structure of a smooth (not nodal) Riemann surface.

It will be useful to have some more adjectives to describe various types of source.

Definition 4.6 We give names to the following specific varieties of (h, m) source:

- A (topological) *bordered* source is a (topological) $(h, 2)$-source.
- A (topological) *polygonal* m-source is a (topological) $(0, m)$-source such that $E(S) = \varnothing$.
- A (topological) *tooth* source is a (topological) (h, m)-source together with a further labelling of the punctures $E(S)$ by either "e" or "w". We will usually denote these by T, and shall abuse notation to write $E_j(S)$ (resp. $W_j(S)$) for the set of punctures labelled by "e" (resp. "w") in part $E_j(S)$.
- In the degenerate case where $m = 1$, ie $D_m = \mathbb{H}$ is the upper-half plane, with one boundary puncture at ∞, we call a (topological) $(h, 1)$ source an *anchor source*.

Later, we will need to fix the behaviour of holomorphic curves near a certain subset of noneast punctures. The relevant kind of source is:

Definition 4.7 A (topological) *splayed* (h, m)-source is a (topological) (h, m) source together with a distinguished set $\rho_{123}(S) \subset \mathrm{punc}(S) - E(S)$, and a partition Q of $\rho_{123}(S)$ into three sets, $Q = (Q_{123}, Q_{1,23}, Q_{12,3})$.

If we are given a metric tree T, then we can associate to every vertex V of the corresponding tree $\overline{T}$ a (topological) $(h, |V|)$ source S_V. If, for each V and V' connected by an edge of $\overline{T}$, we are given a bijection between the punctures of S_V and $S_{V'}$ corresponding with this edge, we can associate to T a corresponding topological *preglued* (h, m) source S_T, where m is the number of noninterior edges of $\overline{T}$. This is formed by deleting small semi-infinite strip neighbourhoods of punctures of S corresponding with an edge of T, and gluing the resultant ends together according to the chosen bijection. There are natural decorations of S_T induced by those upon S_V: each east puncture in a source S_V is preserved by the gluing process, and thus has a corresponding puncture in S_T; we label the resultant punctures by the corresponding element of $\mathcal{T}$, and partition them according to $E_j(S_T) = \bigcup_{V \in V(T)} E_j(S_V)$. If some sources S_V are also endowed with the extra data of a distinguished set $\rho_{123}(S_V)$ and partition Q_V as in Definition 4.7, the resultant source has a natural distinguished set $\rho_{123}(S_T)$ given by the union of all of the $\rho_{123}(S_V)$ after the gluing procedure — where we omit any elements of $\rho_{123}(S_V)$ which are glued along, and regard a (topological) (h, m) source as a splayed source with empty distinguished set.

Remark 4.8 The definition of the source S_T differs from the definition of the surface $S^{\cup}(T)$ defined in Section 4.1 in that (in addition to the extra data of labellings, etc) each of the sources $S(V)$ can have multiple punctures corresponding to the same edge of T. It is also worth noting that we have not specified any complex structure on the surface underlying S_T; the construction gives a purely topological source.

If $\overline{T}$ only has two vertices, V_1 and V_2, then we sometimes denote the corresponding source by $S_{V_1} \natural S_{V_2}$. A similar construction can also be made for tooth sources: if S is a bordered source and T is a tooth source, with a bijection between $E(S)$ and $W(T)$, we can form a glued source $S \natural T$ given by deleting small neighbourhoods of $E(S)$ and $W(T)$ and identifying the resultant "stubs" via the bijection.

4.3 Moduli spaces of holomorphic curves

We will now define a series of moduli spaces of holomorphic curves. Suppose that $\mathcal{D} = (\Sigma; \boldsymbol{\gamma}; \boldsymbol{b}; \boldsymbol{p})$ is a diagram, where $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m\}$. We shall be concerned with holomorphic maps from an m-source into the four-manifold $X_{\mathcal{D},m} := \Sigma \times D_m$. We assume that Σ is endowed with a symplectic form ω_Σ, and choose a complex structure j_Σ upon Σ which is compatible with ω_Σ.

We also assume that j_Σ is cylindrical near the boundary, in that we fix small neighbourhoods $N_j \subset \Sigma_{\bar{e}}$ of each e_j^∞, together with a symplectic identification of each $U_j = N_j - e_j^\infty$ with $S^1 \times (0,1)$. Under this identification we assume that the complex structures $(j_\Sigma)|_{U_j}$ and arcs $\boldsymbol{\gamma} \cap U_j$ are also invariant in the $(0,1)$-direction. It follows from this assumption that the complex structure j_Σ can be extended to a complex structure $j_{\bar{e}}$ upon $\Sigma_{\bar{e}}$.

We denote by

$$\pi_\Sigma : \Sigma \times D_m \to \Sigma \quad \text{and} \quad \pi_D : \Sigma \times D_m \to D_m$$

the relevant projection maps.

Definition 4.9 An *admissible* family of almost-complex structures for $\mathcal{D}$ is a collection $\boldsymbol{J} = \bigcup_{m \geq 1} \boldsymbol{J}_m$ of families of almost-complex structures $\boldsymbol{J}_m = \{J_j\}_{j \in \mathrm{Conf}(D_m)}$ upon $X_{\mathcal{D},m}$ satisfying:

(1) The projection map $\pi_D : \Sigma \times D_m \to D_m$ is (J_j, j)-holomorphic for each j.

(2) The fibres of the maps π_D are J_j-holomorphic for every j.

(3) Each J_j is adjusted to the split symplectic form $\omega_\Sigma \oplus \omega_j$ on $\Sigma \times D_m$.

(4) Near the punctures $e_1^\infty \times D_m, \dots, e_h^\infty \times D_m$, the almost complex structures J_j are split.

For fixed m, we shall study moduli spaces of maps into $X_{\mathcal{D},m}$ which are $J \in \boldsymbol{J}_m$-holomorphic. We will require some form of consistency for the family $\boldsymbol{J}$, so that for any family of elements of $\mathrm{Conf}(D_m)$ converging to some point in $\partial\overline{\mathrm{Conf}(D_m)}$, the corresponding family of almost-complex structures upon $X_{\mathcal{D},m}$ converges to the correct tree of almost-complex structures. Namely:

Definition 4.10 Suppose $\boldsymbol{J}$ is a family of admissible almost-complex structures. Recall that for any m and sequence of m-leaved metric trees $(T_n)_{n \geq 0}$ converging to some degenerate tree T, the chosen family of equivalence classes $\mathrm{Conf}(D_m)$ satisfies that the corresponding almost-complex structures $j_n := j_{T_n} \in \mathrm{Conf}(D_m)$ converge to a tree of complex structures $\{j_V\}_{V \in V(T)}$, where each $j_V \in \mathrm{Conf}(D_{|V|})$. We say that $\boldsymbol{J}$ is *consistent* if the family J_{j_n} converges to the corresponding tree $\{J_{j_V}\}_{V \in V(T)}$ — where each J_{j_V} is an almost-complex structure on $X_{\mathcal{D},|V|}$.

Given a diagram $\mathcal{D}$, there are corresponding Lagrangian "cylinders" inside the four-manifold $X_{\mathcal{D}}$. At their most basic, these are given by the sets of cylinders and half-cylinders $C^i := \boldsymbol{\gamma}^i \times e_i \subset \Sigma \times e_i$. In Section 4.7, we will perturb these cylinders slightly, replacing C^i with cylinders $C^i(\varepsilon)$ associated to some Hamiltonian perturbation of Σ. In this case, the corresponding definition is to be made by replacing C^i with $C^i(\varepsilon)$ below.

With these definitions in mind, we specify the types of holomorphic curves we shall be interested in.

Definition 4.11 Let $\mathcal{D} = (\Sigma; \boldsymbol{\gamma}; \boldsymbol{b}; \boldsymbol{p})$ be a diagram with attaching curves $\boldsymbol{\gamma} = \boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$, and write $\boldsymbol{\gamma}^i = \boldsymbol{\gamma}^{i,a} \cup \boldsymbol{\gamma}^{i,c}$ for the partition of each $\boldsymbol{\gamma}^i$ into arcs and closed curves, respectively. Suppose also that we are given a partition of the points $\boldsymbol{p} = \boldsymbol{w} \cup \boldsymbol{z}$, where either part may be empty.

Fix a consistent, admissible family $\boldsymbol{J}$ of almost complex structures for $\mathcal{D}$, and a source S, endowed with a complex structure j_S. We consider pairs (u, j), where $j \in \mathrm{Conf}(D_m)$ and

$$u\colon (S, \partial S) \to (X_{\mathcal{D},m}, C^1 \cup \cdots \cup C^m)$$

is a map that satisfies the following conditions:

(M-1) u is (j_S, J_j)-holomorphic.

(M-2) $u\colon S \to \Sigma \times D_n$ is proper.

(M-3) u has finite energy.

(M-4) The image of u is disjoint from the sets $\boldsymbol{b} \times D_m$ and $\boldsymbol{z} \times D_m$.

(M-5) u extends to a proper map $u_{\bar{e}}\colon S_{\bar{e}} \to \Sigma_{\bar{e}} \times D_n$. (This is only meaningful when $\Sigma_{\mathcal{D}}$ has boundary.)

(M-6) $\pi_{\mathbb{D}} \circ u_{\bar{e}}$ is a g-fold branched cover.

(M-7) u is an embedding.

(M-8) At each puncture q of S labelled by $p^{i,i+1}$, $\lim_{z \to q}(\pi_{\boldsymbol{D}} \circ u)(z) = p^{i,i+1}$.

When $\Sigma_{\mathcal{D}}$ has boundary, we require three further conditions:

(BOR-1) At each puncture $q \in E(S)_j$, $\pi_\Sigma \circ u$ is asymptotic to the Reeb chord labelling q upon the boundary component $\partial\Sigma_j$.

(BOR-2) For every $i = 1, \dots, m$ and $j = 1, \dots, g+h-1$, and every $t \in e^i$, we require that $u^{-1}(\gamma_j^{i,c} \times \{t\})$ consists of a single point.

(BOR-3) **Strong boundary monotonicity** For every $i = 1, \dots, m$ and $j = 1, \dots, 2k_i$, and every $t \in e^i$, we require that $u^{-1}(\gamma_j^{i,a} \times \{t\})$ consists of at most one point.

From these conditions, it is clear that near each puncture q of S which is labelled by $p^{i,i+1}$, the map u is asymptotic to

$$I^{i,i+1}(q) := x^{i,i+1}(q) \times p^{i,i+1},$$

where $x^{i,i+1}(q)$ is a point in $\boldsymbol{\gamma}^i \cap \boldsymbol{\gamma}^{i+1}$. We collect these into sets of points

$$\boldsymbol{x}^{i,i+1} := \{x^{i,i+1}(q) \mid \lambda(q) = p_{i,i+1}\},$$

for each i, and collect the points $I^{i,i+1}(q)$ into corresponding sets $I^{i,i+1}(\boldsymbol{x}^{i,i+1})$.

We will be concerned with a few specific instances of this construction, which we spend the next sections giving more thorough names to.

4.4 For bordered diagrams

This section briefly recounts [14, Section 5.2].

Suppose that $\mathcal{D}$ is a toroidal bordered diagram, $\mathcal{D} = (\Sigma; \boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{b})$. We relabel $C^1 := C^{\boldsymbol{\alpha}}$ and $C^2 := C^{\boldsymbol{\beta}}$ for clarity's sake, and consider the case when S is a bordered source.

Given generators $\boldsymbol{x}$ and $\boldsymbol{y}$, a choice of almost-complex structure J upon $X_{\mathcal{D},2}$, a fixed decorated source S and a homology class $B \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$, we shall write $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S)$ for the moduli space of curves u with source S which satisfy (M-1)–(M-8) and (BOR-1)–(BOR-3) with respect to J, in the homology class B, with asymptotics $(\boldsymbol{x}, \boldsymbol{y})$. Sometimes, to make explicit which almost-complex structure we are using, we will write $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; J)$.

If $u \in \mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S)$ is a holomorphic curve, then there is a corresponding *evaluation map* for each puncture q of S labelled by a Reeb chord, $\mathrm{ev}_q(u) := t \circ u_{\bar{e}}(q)$. Sometimes, if u is implied, we refer to $\mathrm{ev}_q(u)$ as the *height* of q. There are evaluation maps

$$\mathrm{ev}_j = \prod_{q \in E_j(S)} \mathrm{ev}_q \colon \mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; J) \to \mathbb{R}^{|E_j(S)|},$$

which fit together into a single map $\mathrm{ev} \colon \mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; J) \to \mathbb{R}^{|E(S)|}$ given by $\mathrm{ev} := \mathrm{ev}_1 \times \cdots \times \mathrm{ev}_h$.

If we fix a partition P_j of $E_j(S)$, there is a corresponding subspace $\Delta_{\boldsymbol{P}} \subset \mathbb{R}^{|E_j(S)|}$ given by

$$\{x_p = x_q \mid P^i_j \in P_j, p, q \in P^i_j\}.$$

This allows us to further split the moduli space above by requiring that the heights of punctures in each part coincide; namely:

Definition 4.12 Let $\boldsymbol{x}, \boldsymbol{y}$ be generators, $B \in \pi_2(\boldsymbol{x}, \boldsymbol{y})$, S be a decorated source, and $P := (P_1, \ldots, P_h)$ partitions of the set $E_1(S), \ldots, E_h(S)$. Then we write

$$\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; P_1, \ldots, P_h) := \mathrm{ev}^{-1}(\Delta_{P_1} \times \cdots \times \Delta_{P_h}) \subset \mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S).$$

Again, we sometimes write $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; P_1, \ldots, P_h; J)$ to make explicit which almost-complex structure we are using, and often abbreviate to $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; P)$.

A further refinement is obtained by endowing each partition P_j with an ordering of its parts, $\vec{P}_j$. We can then write $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P})$ for the subset of the set of curves $u \in \mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; P_1, \ldots, P_h)$ for which the ordering of each $\vec{P}_j$ agrees with that given by dictating $P^i_j < P^{i'}_j$ if $\mathrm{ev}_{q^i}(u) < \mathrm{ev}_{q^{i'}}(u)$ for all (any) $q^i \in P^i_j$ and $q^{i'} \in P^{i'}_j$. Often we abbreviate $\vec{P} := (\vec{P}_1, \ldots, \vec{P}_h)$.

If we are given a set of sequences of Reeb chords $\vec{\rho}_{\clubsuit} = (\vec{\rho}_{\clubsuit,1}, \ldots, \vec{\rho}_{\clubsuit,h})$, then we shall say the pair $(\vec{P}, S)$ is *compatible* with $\vec{\rho}_{\clubsuit}$ if the sequence of sets of Reeb chords gotten from $\vec{P}_j$ by replacing each part P^i_j with the set of Reeb chords labelling the punctures $q \in P^i_j \subset E_j(S)$ is equal to $\rho^i_{\clubsuit,j}$, and we

shall often write $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit})$ for the moduli space $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{P})$, where $(\vec{P},S)$ are compatible with $\vec{\rho}_{\clubsuit}$. If $(\vec{P},S)$ are $\vec{\rho}_{\clubsuit}$-compatible, for any $\rho\in\vec{\rho}_{\clubsuit,j}\in\vec{\rho}_{\clubsuit}$ we shall write $q(\rho)$ for the corresponding puncture of S.

Remark 4.13 The moduli space $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit})$ has an action by $\mathbb{R}$, induced by translating in the $\mathbb{R}$-direction of $D_2=[0,1]\times\mathbb{R}$. Unless S is a disjoint union of g strips, each with two boundary punctures, and the map u has trivial homology class, this translation is free — and we say u is *stable*. We will have cause to consider the quotient moduli space $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit})/\mathbb{R}$ which we term $\widetilde{\mathcal{M}}^B(\boldsymbol{x},\boldsymbol{y};S;\vec{P})$.

In [14], what we call $\mathcal{M}_{\leftrightarrow}$ is called $\widetilde{\mathcal{M}}$, and the moduli space we call here $\widetilde{\mathcal{M}}$ is simply called $\mathcal{M}$. We reserve the latter for certain moduli spaces of bordered polygon with constraints on behaviour near certain corners; however, in the case where the bordered polygon is in fact a bigon, this does not present a further constraint and thus $\mathcal{M}=\widetilde{\mathcal{M}}$, agreeing with the previous notation in the literature.

In [14], a number of properties of these moduli spaces are explored. We presently restate those that shall be of use to us.

Proposition 4.14 [14, Proposition 5.6] *For each $\boldsymbol{x},\boldsymbol{y},S$, there is a dense set of admissible J such the moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;J)$ are transversally cut out by the $\bar{\partial}$-equations. Moreover, for any countable set $\{M_i\}$ of submanifolds of $\mathbb{R}^{|E(S)|}$, there is a dense set of admissible J which satisfy that $\mathrm{ev}\colon\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;J)\to\mathbb{R}^{|E(S)|}$ is transverse to all of the M_i; ie there is a dense set of admissible J such that the moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit};J)$ are also transversely cut out.*

*In particular, the set of **sufficiently generic** admissible almost complex structures — those for which the moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit};J)$ are transversally cut out for **all** choices of $\boldsymbol{x},\boldsymbol{y},S,\vec{\rho}_{\clubsuit}$ — is nonempty.*

By the implicit function theorem, it follows that for sufficiently generic J the moduli space of $u\in\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;J;\vec{\rho}_{\clubsuit})$ satisfying $\mathrm{ind}(D\bar{\partial}(u))=k$ is a k-dimensional manifold. The index at any given u satisfies a formula in terms of the topology of S and B, contingent on the *Euler measure* of a domain:

Definition 4.15 Let $B\in\pi_2(\boldsymbol{x}^{0,1},\dots,\boldsymbol{x}^{k,0})$ be a domain. Then its *Euler measure*, $e(B)$, is given by

$$e(B):=\chi(B)-\tfrac{1}{4}\mathrm{ac}(B)+\tfrac{1}{4}\mathrm{ob}(B),$$

where $\chi(B)$ denotes the Euler characteristic, and $\mathrm{ac}(B)$ and $\mathrm{ob}(B)$ are the number of acute and obtuse corners in B respectively.

The relevant index formula is:

Proposition 4.16 [14, Proposition 5.8] *The expected dimension of $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\clubsuit};J)$ is given by*

$$\mathrm{ind}(B,S,P)=g-\chi(S)+2e(B)+|\vec{\rho}_{\clubsuit,1}|+\cdots+|\vec{\rho}_{\clubsuit,h}|,$$

where $|\vec{\rho}_{\clubsuit,j}|$ is the number of Reeb chords in $\vec{\rho}_{\clubsuit,j}$, and $e(B)$ is the Euler measure of B.

It follows from this that for fixed B and S, the dimension of the moduli space $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}; J)$ is well defined. There is a further interpretation of this formula solely in terms of data about the domain $[u]$, though we do not pursue this here; the important related fact for us will be:

Proposition 4.17 [14, Proposition 5.69] *Whether or not a holomorphic curve $u \in \mathcal{M}_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}; J)$ is embedded is determined solely by the Euler characteristic of the source S: the curve u is embedded if and only if $\chi(S) = \chi_{\mathrm{emb}}(S)$ for some well-defined quantity $\chi_{\mathrm{emb}}(S)$.*

Though the moduli spaces $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}; J)$ form manifolds, they are not compact except when the moduli space $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}; J)$ is zero-dimensional. They admit certain compactifications in terms of *holomorphic combs*, which we consider in Section 4.6.3.

4.4.1 Splicings and moduli spaces Suppose that S is a bordered source and $\vec{P}$ is a sequence of partitions $\vec{P}_1, \dots, \vec{P}_h$, such that $(S, \vec{P})$ is compatible with a sequence of sets of Reeb chords $\vec{\rho} = (\vec{\rho}_1, \dots, \vec{\rho}_h)$.

The definitions of the moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; P)$ and $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P})$ stipulate relations on the relative heights of punctures from each set $E_j(S)$, as compatible with the sequence $\vec{\rho}$. More generally, we could consider constraints upon relative heights of punctures which belong to $E_j(S)$ and $E_{j'}(S)$ for $j \neq j'$.

Suppose we are given an m-splicing $\vec{\sigma} = \vec{\sigma}_1, \dots, \vec{\sigma}_h$ of $\vec{\rho}$, and that $(S, \vec{P})$ is $\vec{\rho}$-compatible. There are m corresponding sets

$$Q^i := \{q(\rho) \mid \rho \in \sigma^i_1 \cup \cdots \cup \sigma^i_h\}$$

of punctures. In turn, these yield corresponding submanifolds of $\mathbb{R}^{|E(S)|}$ specified by

$$\Delta^i_\sigma := \{x_p = x_q \mid p, q \in Q^i\}.$$

This allows us to define moduli spaces $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; P, \sigma)$ and $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P}, \sigma)$, given by

$$\mathrm{ev}^{-1}(\Delta^1_\sigma, \dots, \Delta^m_\sigma)$$

for the evaluation maps from $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; P)$ and $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P})$, respectively, to $\mathbb{R}^{|E(S)|}$. These are transversally cut out for sufficiently generic J by Proposition 4.14.

The corresponding index formula is given by

$$\mathrm{ind}(B, S, P, \sigma) = g - \chi(S) + 2e(B) + |\vec{\rho}_{\clubsuit,1}| + \cdots + |\vec{\rho}_{\clubsuit,h}| - \mathrm{Col}(\sigma),$$

and is proved in a similar manner as [14, Proposition 5.8] — each height constraint coming from σ drops the expected dimension by one.

These moduli spaces only depend upon the unordered partitions in σ. The extra information of the order of the partitions in $\vec{\sigma}$ allows us to define even more restricted moduli spaces $\widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P}, \vec{\sigma})$ as the set of $u \in \widetilde{\mathcal{M}}^B(\boldsymbol{x}, \boldsymbol{y}; S; \vec{P}, \sigma)$ such that the ordering of the parts in σ according to $\vec{\sigma}$ agrees with the ordering of the parts in σ by dictating that $Q^i < Q^{i'}$ if $\mathrm{ev}_{q_i}(u) < \mathrm{ev}_{q_{i'}}(u)$ for every (any) $q^i \in Q^i$ and $q^{i'} \in Q^{i'}$.

4.5 Heegaard multidiagrams

We now turn our attention to the relevant moduli spaces when $\mathcal{D}$ is a Heegaard m-diagram, $\mathcal{D} = (\Sigma, \boldsymbol{\gamma}^0, \dots, \boldsymbol{\gamma}^{m-1}, \boldsymbol{p})$, where $\boldsymbol{p}$ splits as $\boldsymbol{w} \cup \boldsymbol{z}$.

Fix a collection of admissible compatible almost-complex structures $\boldsymbol{J}$. As Σ is closed, the conditions (BOR-1)–(BOR-3) do not apply, and we are concerned with pairs (u, j) where $j \in \mathrm{Conf}(D_m)$ and u is a (j_S, J_j)-holomorphic curve satisfying (M-1)–(M-8) disjoint from the set $\boldsymbol{z} \times D_m$.

The conditions imply that $\boldsymbol{x}^{i,i+1}$ is a generator for the pair $(\boldsymbol{\gamma}^i, \boldsymbol{\gamma}^{i+1})$. In particular, the sequence $\vec{\boldsymbol{x}} = (\boldsymbol{x}^{0,1}, \dots, \boldsymbol{x}^{m-1,0})$ constitutes a sequence of generators for $\mathcal{D}$.

For $m > 2$, we collect the holomorphic curves (u, j) with source S and given asymptotics $\vec{\boldsymbol{x}}$ in homology class $B \in \pi_2(\vec{\boldsymbol{x}})$ satisfying (M-1)–(M-8) into moduli spaces $\widetilde{\mathcal{M}}^B(\vec{\boldsymbol{x}}; S)$. If $m = 2$, as in the previous section the relevant moduli space has a free $\mathbb{R}$-action, and so we denote by $\mathcal{M}^B(\vec{\boldsymbol{x}}; S)$ the result of taking the quotient by this. We call an element of one of these moduli spaces a *holomorphic m-gon*. As in the previous section, we sometimes include the family $\boldsymbol{J}$ in the notation to draw attention to it; writing things like $\mathcal{M}^B(\vec{\boldsymbol{x}}; S; \boldsymbol{J})$, for instance.

For any subset $U \subset \{0, \dots, m-1\}$ with $|U| = l$, we can apply the above construction to $\mathcal{D}|_U$ to define corresponding moduli spaces $\mathcal{M}^B(\boldsymbol{x}^{i_0,i_1}, \dots, \boldsymbol{x}^{i_{l-1},i_0}; S; J)$.

The relevant transversality result for holomorphic m-gons is:

Proposition 4.18 [13, Proposition 3.8] *There is an admissible, compatible family of almost-complex structures $\boldsymbol{J}$ for $\mathcal{D}$, such that for every source, every l-subdiagram of $\mathcal{D}$, every sequence of generators for such a subdiagram, and every almost-complex structure $J \in \boldsymbol{J}_l$, the corresponding moduli spaces are all transversally cut out.*

There is a corresponding index formula:

Proposition 4.19 *Suppose that $(u, j) \in \mathcal{M}^B(\boldsymbol{x}^{0,1}, \dots, \boldsymbol{x}^{k-1,k}, \boldsymbol{x}^{k,0}; S)$ is a holomorphic polygon. Then the index of the $D\bar{\partial}$ operator at (u, J) is given by*

$$\mathrm{ind}(B, S) = \tfrac{1}{2}(3-k)g - \chi(S) + 2e(B).$$

Once more, it follows from this and the implicit function theorem that these moduli spaces are manifolds of dimension $\mathrm{ind}(B, S) + k - 2$ (the $k - 2$ stems from the space $\mathrm{Conf}(D_k)$ having moduli), though in general noncompact.

4.6 Bordered polygons

In this section, we construct relevant moduli spaces of polygons for bordered multidiagrams, generalising both of the previous two sections. For our purposes, it will be enough to consider bordered polygons that only have east asymptotics during the first edge of D_n, which we have called e^1 above.

We fix a toroidal bordered Heegaard m-diagram $\mathcal{D}$. Supposing that $\boldsymbol{J}$ is a collection of consistent, admissible almost-complex structures, we again consider pairs (u, j) of $j \in \mathrm{Conf}(D_m)$ and (j_s, J_j)-holomorphic curve u with source S satisfying (M-1)–(M-8) and (BOR-1)–(BOR-3), mapping into

$$(\Sigma \times D_m, C^0 \cup \cdots \cup C^{m-1}).$$

The conditions imply that the sequence of points $\vec{\boldsymbol{x}} = (\boldsymbol{x}^{0,1}, \dots, \boldsymbol{x}^{m-1,0})$ form a sequence of generators for $\mathcal{D}$. As per usual, such (u, j) have a homology class B in the group $\pi_2(\vec{\boldsymbol{x}})$, and so we collect these maps into moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S)$.

Once again, if we wish to draw attention to the family $\boldsymbol{J}$ in question, we include it in the notation; eg $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; \boldsymbol{J})$. We often refer to a holomorphic curve (u, j) by solely u, and hope that this does not cause undue confusion.

As in Section 4.4, there is an evaluation map at each east puncture q of S, $\mathrm{ev}_q(u) : \mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S) \to e^1 \subset \partial D_m$. These fit together to form evaluation maps

$$\mathrm{ev}_j : \mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S) \to (e^1)^{|E_j(S)|}$$

for each $j = 1, \dots, h$, which in turn fit together to give a map

$$\mathrm{ev} : \mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S) \to (e^1)^{|E(S)|}.$$

As such, for any partitions $P := (P_1, \dots, P_h)$ of $E_1(S), \dots, E_h(S)$ there is once more a corresponding moduli space

$$\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; P) := \mathrm{ev}^{-1}(\triangle_{P_1} \times \cdots \times \triangle_{P_h}) \subset \mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S).$$

Furthermore, ordering the parts of the P_j as $\vec{P}_j = (P_j^1, \dots, P_j^{k_j})$ yields a corresponding moduli space

$$\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; \vec{P}),$$

where we further require that the orderings of the punctures induced by the evaluation maps ev_j agree with that of $\vec{P}_j$.

Now consider the set of sequences of Reeb chords $\vec{\rho}_{\triangleq} = \vec{\rho}_1, \dots, \vec{\rho}_k$. We will say that a map u is *compatible* with $\vec{\rho}_{\triangleq}$ if for each i and j, we have that $\{\rho(q)\}_{q \in P_j^i} = \rho_j^i$. Sometimes we shall write

$$\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\triangleq})$$

for the moduli space of curves in $\mathcal{M}^B_{\leftrightarrow}(\boldsymbol{x}; S; \vec{P})$ which is compatible with $\vec{\rho}_{\triangleq}$.

As in Section 4.4.1, if we are given an m-splicing $\vec{\sigma} = \vec{\sigma}_1, \dots, \vec{\sigma}_h$, there are corresponding submanifolds $\Delta^1_{\vec{\sigma}}, \dots, \Delta^m_{\vec{\sigma}}$ of $\mathbb{R}^{|E(S)|}$, and thus we may define the moduli spaces $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; P; \sigma)$ and $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}, ; S; \vec{P}; \sigma)$ by $\mathrm{ev}^{-1}(\Delta^1_\sigma, \dots, \Delta^h_\sigma)$ for the evaluation maps from $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; P)$ and $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}; S; \vec{P})$ to $\mathbb{R}^{|E(S)|}$ respectively. We also write $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}, ; S; \vec{P}; \vec{\sigma})$ for the subset of $\mathcal{M}^B_{\leftrightarrow}(\vec{\boldsymbol{x}}, ; S; \vec{P}; \sigma)$ such that the ordering upon jumping punctures induced by the evaluation maps agrees with the ordering dictated by $\vec{\sigma}$.

As the group of Möbius transformations acts three-transitively upon boundary points of the disc, we can assume that the greatest height of an east puncture of S labelled by a jumping chord of $\mathcal{T}$ is 0. When $m \neq 2$, this does not alter the moduli spaces we study; when $m = 2$, these moduli spaces are precisely the bordered moduli spaces $\mathcal{M}^{\boldsymbol{B}}_{\leftrightarrow}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\blacktriangle}; \vec{\sigma})$ described in Section 4.4, and the effect of fixing the height to be zero is to instead consider the quotient moduli spaces $\widetilde{\mathcal{M}}^{\boldsymbol{B}}(\boldsymbol{x}, \boldsymbol{y}; S; \vec{\rho}_{\blacktriangle}; \vec{\sigma})$. In general, we denote the relevant moduli space of curves with greatest height of an east puncture fixed to be zero by replacing any instances of $\mathcal{M}_{\leftrightarrow}$ with $\widetilde{\mathcal{M}}$ in the above definitions.

There is a corresponding transversality result:

Proposition 4.20 *There is an admissible, compatible family of almost-complex structures $\boldsymbol{J}$ for $\mathcal{D}$, such that for every l-fold subdiagram of $\mathcal{D}$, every source, every sequence of generators for the subdiagram, every set of sequences of Reeb chords $\vec{\rho}_{\blacktriangle}$ and every corresponding splicing $\vec{\sigma}$, the corresponding moduli spaces $\widetilde{\mathcal{M}}^{\boldsymbol{B}}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}; \vec{\sigma})$ are all transversally cut out by the $\bar{\partial}$ equations. In this instance, we shall say that $\boldsymbol{J}$ is **generic**.*

Proof This is a minor modification of [13, Proposition 4.18]. □

In fact, the proof of Proposition 4.20 contains a little more information. Namely:

Proposition 4.21 *Let $\mathcal{U}$ denote the space of all admissible, compatible families almost-complex structures, and suppose that $\boldsymbol{J}$ is a generic family of admissible, compatible almost-complex structures. For any path $\boldsymbol{J}_t : [0,1] \to \mathcal{U}$ which satisfies that $\boldsymbol{J}_0 = \boldsymbol{J}$, there is some ε such that $\boldsymbol{J}_t$ is generic for all $t \in [0, \varepsilon)$.*

This should be compared with, for instance, [16, Remark 3.2.8].

The common generalisation of Theorems 4.16 and 4.19 is below, and proved in a very similar manner; cf [13, Lemma 4.7].

Proposition 4.22 *Suppose that $(u, j) \in \widetilde{\mathcal{M}}^{\boldsymbol{B}}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}, \vec{\sigma})$ is a bordered holomorphic polygon. Then the index of the $D\bar{\partial}$ operator at u is given by*

$$\operatorname{ind}(B, S, \vec{\rho}_{\blacktriangle}, \sigma) := \tfrac{1}{2}(3-m)g - \chi(S) + 2e(B) + |\vec{\rho}_{\blacktriangle,1}| + \cdots + |\vec{\rho}_{\blacktriangle,h}| - \operatorname{Col}(\sigma).$$

And, much like Proposition 4.17, we have the following:

Proposition 4.23 [13, Proposition 4.9] *Whether or not a holomorphic curve in $\widetilde{\mathcal{M}}^{\boldsymbol{B}}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}, \vec{\sigma})$ is embedded is determined solely by the Euler characteristic of the source S.*

4.6.1 Compactifying at east infinity We now consider the effect of compactifying a toroidal bordered diagram at east infinity upon the corresponding moduli spaces. If $\mathcal{D} = (\Sigma; \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^m; \boldsymbol{p}; \boldsymbol{b})$ is a toroidal bordered multidiagram, recall that there is a corresponding compactified diagram $\mathcal{D}_{\bar{e}}$, where $\partial\Sigma_j$ is

replaced by a single point e_j^∞, and completing the arcs $\boldsymbol{\gamma}^{1,a}$ to curves $\boldsymbol{\gamma}_{\bar{e}}^{1,a}$. There is a corresponding complex structure $j_{\Sigma_{\bar{e}}}$ (resp. symplectic form) upon $\Sigma_{\bar{e}}$ which agrees with j_Σ (resp. ω_Σ) outside of arbitrarily small neighbourhoods of e_j^∞, and is such that the compactified $\gamma_i^{1,a}$ are smooth curves (ie the angles between them are correct). Similarly, to every family of (generic, admissible, consistent) almost-complex structures $\boldsymbol{J} = \bigcup_m \boldsymbol{J}^m$ there is a corresponding family $\boldsymbol{J}_{\bar{e}} = \bigcup_m \boldsymbol{J}_{\bar{e}}^m$, compatible with the corresponding split symplectic form $\omega_{\Sigma_{\bar{e}}} \times \omega_{D_n}$. Any sequence of generators $\vec{\boldsymbol{x}} \in \mathcal{G}(\mathcal{D})$ has an obvious counterpart in $\mathcal{D}_{\bar{e}}$, which we call $\vec{\boldsymbol{x}}'$ here but usually abuse notation and also call $\vec{\boldsymbol{x}}$. Similarly each domain $B \in \pi_2(\vec{\boldsymbol{x}})$ has a corresponding domain $B_{\bar{e}} \in \pi_2(\vec{\boldsymbol{x}}')$.

To any map $u\colon S \to X_{\mathcal{D},m} \times D_m$ there is a corresponding map $u'\colon S \to X_{\mathcal{D}_{\bar{e}}} \times D_m$. We can do slightly better, though: because of the conformal structure upon $X_{\mathcal{D}_{\bar{e}}}$ near east infinity, punctures corresponding to nonjumping chords can be removed completely from S. To any bordered polygonal source S we may associate the source $S_{\bar{e}}$ given by compactifying S at its east punctures labelled by ρ_{12} or ρ_{23}. We then consider the space of pairs $(u, J_{\bar{e}})$ where $J_{\bar{e}} \in \boldsymbol{J}_{\bar{e}}$ and u is a $J_{\bar{e}}$-holomorphic curve satisfying analogues of properties (M-1)–(M-8) and (BOR-1)–(BOR-3).

We denote the moduli space of such pairs in homology class $B_{\bar{e}}$ with asymptotics described by $\vec{\boldsymbol{x}} = (\boldsymbol{x}^{0,1}, \dots, \boldsymbol{x}^{m-1,0})$ by $\mathcal{M}_{\leftrightarrow}^{B_{\bar{e}}}{}_{\bar{e}}(\boldsymbol{x}^{0,1}, \dots, \boldsymbol{x}^{m-1,0}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$.

This is closely related to the moduli space $\mathcal{M}_{\leftrightarrow}^{B}(\vec{\boldsymbol{x}}; S; \boldsymbol{J})$. In particular, if u is a holomorphic curve in $\mathcal{M}_{\leftrightarrow}^{B}(\vec{\boldsymbol{x}}; S; \boldsymbol{J})$, there is a corresponding map $u_{\bar{e}} \in \mathcal{M}_{\leftrightarrow}{}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$ which can be constructed as follows.

In a small neighbourhood of an east puncture of S the map u takes the form $u = (u_\Sigma, u_{D_n})$, as the almost-complex structure J is split near east infinity. By the Schwarz reflection principle and Riemann removable singularities theorem, the map u_Σ can be extended holomorphically across the puncture. By the uniqueness in Schwarz and Riemann's theorems, this identifies the moduli space $\mathcal{M}_{\leftrightarrow}^{B}(\vec{\boldsymbol{x}}; S; \boldsymbol{J})$ with $\mathcal{M}_{\leftrightarrow}{}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$, and we denote the image of $\mathcal{M}_{\leftrightarrow}^{B}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\clubsuit}, \vec{\sigma}; \boldsymbol{J})$ under this identification by

$$\mathcal{M}_{\leftrightarrow}{}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\clubsuit}, \vec{\sigma}; \boldsymbol{J}_{\bar{e}}).$$

If $u \in \mathcal{M}_{\leftrightarrow}{}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$, then for $j = 1, \dots, h$, there is a sequence of partitions P_j of $\overline{E}_j(S_{\bar{e}}) := (\pi_\Sigma \circ u)^{-1}(e_j^\infty)$, determined by dictating that q and q' belong to the same part if and only if $\mathrm{ev}(q) = \mathrm{ev}(q')$. These may be ordered as $\vec{P}_j$, where for parts P_j and P_j' we say that $P_j < P_j'$ if and only if for some $q \in P_j$ and $q' \in P_j'$ we have that $\mathrm{ev}_j(q) < \mathrm{ev}(q')$.

If $q \in \overline{E}_j(S_{\bar{e}})$, then there is an associated Reeb chord $\rho(q)$ defined by first choosing a small disc neighbourhood M_j of e_j^∞ such that $S_q := (\pi_\Sigma \circ u)^{-1}(M_j)$ is a small half-disc neighbourhood of $q \in S$ which is disjoint from all other elements of $\overline{E}_j(S_{\bar{e}})$. The homology class of $(\pi_\Sigma \circ u)(S_q)$ is given by $B(\rho(q))$ for $\rho(q)$ some element of the torus algebra, as described in Section 3.1.1. It follows that u has a corresponding set of sequences of algebra elements $\vec{\rho}(u) := \vec{\rho}_1, \dots, \vec{\rho}_h$, given by replacing each $q \in \vec{P}_j$ by $\rho(q)$.

The subspace

$$\mathcal{M}_{\leftrightarrow}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\clubsuit}, \vec{\sigma}; \boldsymbol{J}_{\bar{e}}) \subset \mathcal{M}_{\leftrightarrow}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$$

can be characterised as the set of u such that $\vec{\rho}(u) = \vec{\rho}_{\clubsuit}$, and such that the ordered partition of the punctures of $S_{\bar{e}}$ (which correspond with jumping chords in $\vec{\rho}_{\clubsuit}$) induced by the evaluation maps agrees with the splaying $\vec{\sigma}$.

As with the moduli spaces $\mathcal{M}_{\leftrightarrow}^{B}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\clubsuit}; \vec{\sigma})$, we may assume that the highest height of an east puncture of S is 0 — calling the corresponding moduli spaces $\widetilde{\mathcal{M}}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$ and $\widetilde{\mathcal{M}}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\clubsuit}, \vec{\sigma}; \boldsymbol{J}_{\bar{e}})$.

If $\vec{\rho}_{\clubsuit}$ contains no jumping chords, then along the edge $\Sigma_{\bar{e}} \times e^1$, $u(\partial S_{\bar{e}})$ is contained in a set of $g + h - 1$ cylinders which are a subset of $C_{\bar{e}}^1 := \boldsymbol{\gamma}_{\bar{e}}^1 \times e^1$. It follows that:

Proposition 4.24 *If $\vec{\rho}_{\clubsuit}$ has no jumping chords, then the moduli space $\widetilde{\mathcal{M}}_{\bar{e}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\clubsuit}; \varnothing; \boldsymbol{J}_{\bar{e}})$ defined by counting (bordered) polygons associated with the partially splayed diagram $\mathcal{D}^{\partial}(\varepsilon, \vec{\iota})$ is identified with the moduli space $\widetilde{\mathcal{M}}^{B_{\bar{e}}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \boldsymbol{J}_{\bar{e}})$, defined by counting polygons with respect to the (not partially) splayed Heegaard multidiagram $\mathcal{D}(\varepsilon, \vec{\iota})$.*

4.6.2 Fixing cut behaviour Now we consider the case where $\mathcal{D}$ is a splayed or partially splayed diagram $\mathcal{D}'(\varepsilon, \vec{\iota})$ or $\mathcal{D}^{\partial}(\varepsilon, \vec{\iota})$, and suppose that $B \in \mathrm{Dom}(\mathcal{D})$ is a domain where $\vec{\sigma}(B) = \vec{\sigma}^1, \dots, \vec{\sigma}^k$ (see Definition 3.30).

If $\vec{\rho}(B) = ([\vec{\rho}_1(B)], \dots, [\vec{\rho}_h(B)])$, let $\tilde{\rho}_j = (\tilde{\rho}_j^1, \dots, \tilde{\rho}_j^{i_j})$ be the 123-reduced representative of $[\vec{\rho}_j(B)]$, and denote by $\tilde{\rho}_{\clubsuit} := (\tilde{\rho}_1, \dots, \tilde{\rho}_h)$. If S is a decorated k-source, we denote by $\rho_{123}(S, u)$ the set of punctures of S which are mapped by u to a point $e_j^{\infty,i}$ such that the i^{th} chord $\tilde{\rho}_j^{k_i}$ in $\mathrm{jump}(\tilde{\rho}_j)$ satisfies $\tilde{\rho}_j^{k_i} = \rho_{123}$.

We will impose extra conditions upon the moduli spaces $\widetilde{\mathcal{M}}^{B}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\clubsuit}; \vec{\sigma})$ based upon the behaviour of u near any $q \in \rho_{123}(S, u)$. Namely, fix such a q which is mapped to $e_j^{\infty,i}$. We shall say that u has a *cut* at q if $\pi_{\Sigma} \circ u$ has a boundary branch point along one of the arcs γ_j^i or γ_j^{i+1}. There are three cases:

(1) If there is a boundary branch point upon $\gamma^{i,i+1}$, we will say that q is of (ρ_{12}, ρ_3)-type.

(2) When there is a boundary branch point upon $\gamma^{i,i+1}$, we will say that q is of (ρ_1, ρ_{23})-type.

(3) When there is no boundary branch point we shall say that q is of ρ_{123}-type.

This has an equivalent description: if there is a branch point along γ_j^i then there is some small disc neighbourhood $S(q)$ of $e_j^{\infty,i}$ such that when restricted to the component $S(q)$ of $u|_{S(q)}$ containing q, the homology class of $\pi_{\Sigma} \circ u(S(q))$ is modelled upon $B(\rho_3)$. Similarly, if the branch point appears upon γ_j^{i+1}, there is a corresponding $S(q)$ such that the homology class of $\pi_{\Sigma} \circ u(S(q))$ is modelled upon $B(\rho_1)$. Finally, if there is no cut at q, every sufficiently small disc neighbourhood M_j^i of q satisfies that, restricted to the component $S(q)$ containing q, the map $\pi_{\Sigma} \circ u$ is modelled upon $B(\rho_{123})$.

We would like to impose conditions upon our moduli spaces that dictates which of these cases holds for each q. We first choose sensible notation, and then rewrite the conditions in a way for which transversality and index formulae are more apparent.

Definition 4.25 Fix $B \in \pi_2(\vec{x})$ with associated sets of sequences of Reeb chords $\tilde{\rho}_{\natural}$, and fix also a splayed source S with corresponding partition of $\rho_{123}(S)$ into $Q = (Q_{123}, Q_{1,23}, Q_{12,3})$.

We shall write $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q_{123}; \vec{\sigma})$ for the subset of $u \in \widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}; \vec{\sigma})$ where $\rho_{123}(S,u) = \rho_{123}(S)$ and each $q \in Q_{123}$ is of ρ_{123}-type (with respect to u). Similarly, we shall write $\mathcal{M}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q; \vec{\sigma})$ for the subset of $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q_{123}; \vec{\sigma})$ such that each $q \in Q_{1,23}$ is of (ρ_1, ρ_{23})-type, and each $q \in Q_{12,3}$ is of (ρ_{12}, ρ_3)-type.

We often use a shorthand for these moduli spaces:

Definition 4.26 Write $\vec{\rho}_{\natural}(Q,u)$ for the representative for the 123-equivalence class $[\vec{\rho}(B)]$ given by replacing each $\rho \in \tilde{\rho}_{\natural}$ which is equal to ρ_{123} by ρ_{12}, ρ_3 if $q(\rho) \in Q_{12,3}$; by ρ_1, ρ_{23} if $q(\rho) \in Q_{1,23}$; and, finally, leaving it as ρ_{123} if $q(\rho) \in Q_{123}$.

If $\vec{\rho}_{\natural}$ is a representative of $[\vec{\rho}(B)]$, define by

$$\mathcal{M}^B(\boldsymbol{x}, S, \vec{\rho}_{\clubsuit}, \vec{\rho}_{\natural}; \vec{\sigma})$$

the moduli space

$$\mathcal{M}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q; \vec{\sigma}),$$

where $\vec{\rho}_{\natural} = \vec{\rho}_{\natural}(Q,u)$.

We call an element of $\mathcal{M}^B(\boldsymbol{x}, S, \vec{\rho}_{\clubsuit}, \vec{\rho}_{\natural}; \vec{\sigma})$ a *splayed* bordered holomorphic polygon.

To ensure the corresponding moduli spaces are transversally cut out, we rephrase this condition in a way more compatible with the language of almost-complex structures (cf [18, Lemma 9.4; 1, Chapter 6]). First, a definition:

Definition 4.27 Let $u \in \mathcal{M}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q; \vec{\sigma})$, and let $q \in Q_{1,23}$ or $Q_{12,3}$. Denote the corresponding boundary branch point by b, and the arc of ∂S connecting q and $(\pi_\Sigma \circ u)^{-1}(b)$ by $\lambda_{q,b}$. We define the *length of the cut at q* to be the length of the arc $\bar{\lambda}_{q,b} := (\pi_\Sigma \circ u)(\lambda_{q,b})$ with respect to the metric induced upon Σ by the uniformisation theorem.

We focus initially upon curves where the cut behaviour near each $q \in \rho_{123}(S)$ is fixed so that either there is no cut, or such that the length of the cut at q is short enough that the corresponding boundary branch point b is contained in a disc neighbourhood N_j^i of $e_j^{\infty,i}$ which is holomorphically identified with the unit disc in $\mathbb{C}$. We also require that this disc is small enough that for each q, the component $S(q)$ of $(\pi_\Sigma \circ u)^{-1}(N_j^i)$ containing q is a half-disc with q its only boundary puncture. Denote by (u_j^i, j') the restriction of (u, j') to $(\pi_\Sigma \circ u)^{-1}(N_j^i)$ (we recall that we have been writing u for the pair (u, j') where $j' \in \mathrm{Conf}(D_m)$).

We can modify the complex structure upon N_j^i so that the angle between the arcs in $\partial B_j^i(\rho_{123})$ is π (by composing with the map $\varphi\colon z \mapsto z^{2/3}$, say), and modify the (split) almost-complex structure $J|_{N_j^i} \times j'$

over $N_j^i \times D_m$ accordingly, to a corresponding almost-complex structure $J^0_{N_j^i} \times j'$. To u_j^i we may associate a corresponding v_j^i with respect to this new almost-complex structure: moreover, if u_j^i had no boundary branch point then this new map does not either, and so there is a corresponding holomorphic map from the half-disk $S(q)$ with boundary puncture q removed to the manifold $X_{\mathcal{D},m}$ (by the removable singularities theorem and Schwarz reflection principles).

Similarly, if u_j^i had a boundary branch point then so does v_j^i. We further modify the complex structure by composing the chart on N_j^i with the Riemann mapping which takes the open disc with slit of length t along $\varphi(\gamma_j^{i+1})$ (resp. slit of length $-t$ along $\varphi(\gamma_j^i)$) to the open half-disc. These Riemann mappings yield a family of almost-complex structures $J^t_{N_j^i}$ upon N_j^i for each $t \in (-1,1)$, such that for every $(J_{N_j^i} \times j')$-holomorphic curve u_j^i with a cut of length t along γ_j^{i+1} (resp. length $-t$ along γ_j^i), the corresponding curve (which we also denote by v_j^i) with respect to $J^t_{N_j^i} \times j'$ has no cut, and thus can be completed to a holomorphic map with source given by the half-disk.

We would like to modify the almost-complex structures $J_{j'}$ to agree with this $J^t_{N_j^i} \times j'$ in a neighbourhood of $u(q)$ and then apply the transversality result of Proposition 4.20, but this does not make sense with respect to other components of $u^{-1}(N_j^i)$. The solution, as in [14, Proposition 6.36] (see also [25]) is to instead consider, for $\vec{t} = \{t_q\}_{q \in \rho_{123}(S)}$, related moduli spaces of $\boldsymbol{J}^{\vec{t}}$- holomorphic sections $u \colon S \to S \times \Sigma_{\vec{e}} \times D_n$, where each $\boldsymbol{J}^{\vec{t}}$ is an almost-complex structure on $S \times \Sigma_{\vec{e}} \times D_n$ equal upon each (s,z,w) to some $j_S \times \boldsymbol{J}_s^{\vec{t}}$. We choose $\boldsymbol{J}_s^{\vec{t}}$ so that for s near any puncture $q \in \rho_{123}(S)$, the almost-complex structure $\boldsymbol{J}_s^{\vec{t}}|_{N_j \times D_m}$ agrees with the split almost-complex structure $J^{t_q}_{N_j^i} \times j'$ described in the previous paragraph, and away from this agrees with $\boldsymbol{J}$.

The proof of transversality for such families follows closely along the lines of [10, Chapter 3]. As such, the moduli spaces $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma})$ are transversally cut out. Moreover the moduli space

$$\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, \rho_{123}(S); \vec{\sigma}) \subset \widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma})$$

of curves with no cut at any member of $\rho_{123}(S)$ is also transversally cut out and, along the lines of Proposition 4.21, for some ε, the moduli spaces defined with respect to $\boldsymbol{J}^{\vec{t}}$ for $\vec{t} \in (-\varepsilon, \varepsilon)^{|\rho_{123}(S) - Q_{123}|}$ are also transversally cut out — thus each curve $u \in \widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, \rho_{123}(S), \vec{\sigma})$ has a neighbourhood in $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma})$ homeomorphic to $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma}) \times (-\varepsilon, \varepsilon)^{\rho_{123}(S) - Q_{123}}$.

Suppose that $u, u' \in \widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma}_{\triangleq})$, and that with respect to u a puncture $q \in \rho_{123}(S) - Q_{123}$ is of (ρ_{12}, ρ_3)-type, and with respect to u' the same puncture is of (ρ_1, ρ_{23})-type. The topology upon the moduli spaces described in Definition 4.34 is metrizable (cf [4, Appendix B]), and for each $q \in \rho_{123}(S) - Q_{123}$ the length of the curve $\bar{\lambda}_{q,b}$ is continuous with respect to this metric. As such (by the intermediate value theorem), the curves u and u' belong to different path-components of

$$\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, Q_{123}; \vec{\sigma}) - \widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\triangleq}, \rho_{123}(S); \vec{\sigma}).$$

It follows that:

Proposition 4.28 *There is an admissible, compatible family of almost-complex structures $\boldsymbol{J}$ for $\mathcal{D}$, such that for all choices of argument, the moduli spaces $\mathcal{M}^{\boldsymbol{B}}(\vec{\boldsymbol{x}}; S; P, Q; \vec{\sigma}_{\clubsuit})$ are all transversally cut out, and of expected dimension*

$$\operatorname{ind}(B, S, P, Q, \sigma) := \operatorname{ind}(B, S, P, \sigma) - \#_{123}(\vec{\rho}_{\maltese}).$$

Proof Transversality is explained in the lead-up to the statement of the proposition. The index formula follows along similar lines to [10, Section 4.1]: we do not count "corners" stemming from punctures of ρ_{123}-type, as with respect to the relevant $\boldsymbol{J}^{\vec{t}}$ they are not corners at all. □

Convention 4.29 In the course of this paper, we will be considering a number of different partial splayings of a bordered Heegaard diagram, which has the potential to become confusing. By Definition 2.7, a splicing $\vec{\sigma}$ of a set of sequences of Reeb chords $\vec{\rho}_{\maltese}$ determines a sequence of sets of idempotents $\iota_{\vec{\sigma}}$; as such, we may indicate the diagram being considered for a specific moduli space by using the notation

$$\mathcal{M}^{\boldsymbol{B}}(\boldsymbol{x}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\clubsuit}, \vec{\sigma}_{\maltese})$$

to refer to the moduli space

$$\mathcal{M}^{\boldsymbol{B}}(\boldsymbol{x}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\clubsuit})$$

defined using the diagram $\mathcal{D}^{\partial}(\varepsilon, \iota_{\vec{\sigma}_{\maltese}})$.

Similarly, to keep notation compact, when it is clear from the context we will write

$$\mathcal{M}^{\boldsymbol{B}}(\boldsymbol{x}; S; \vec{\rho}_{\maltese}; \vec{\sigma}_{\maltese})$$

for the moduli space

$$\mathcal{M}^{\boldsymbol{B}}(\boldsymbol{x}; S; \varnothing, \vec{\rho}_{\maltese}; \varnothing, \vec{\sigma}_{\maltese})$$

defined using the closed multidiagram $\mathcal{D}(\varepsilon, \iota_{\vec{\sigma}})$, and

$$\operatorname{ind}(B, S, Q_{\maltese}) := \operatorname{ind}(B, S, \varnothing, Q_{\maltese}, \varnothing),$$

where B is a domain in $\mathcal{D}(\varepsilon, \iota_{\vec{\sigma}})$.

4.6.3 Compactifications As in [14], the aforementioned moduli spaces admit natural compactifications in terms of *holomorphic combs*.

To discuss these, we need the notion of a holomorphic tooth; this is a map

$$v\colon (T, \partial T) \to (\mathbb{R} \times (\partial\Sigma - \boldsymbol{b}) \times D_n, \mathbb{R} \times \boldsymbol{a} \times e^1)$$

where T is a tooth source, and the codomain is endowed with the obvious split symplectic form and the split complex structure $J = j_\Sigma \times j_{D_n}$. We denote by $\mathcal{N}(T)$ the moduli space of such maps satisfying:

(1) v is (j, J)-holomorphic for some almost complex structure j upon T.

(2) v is proper.

(3) $\pi_{\boldsymbol{D}} \circ v$ maps ∂T to e^1.

(4) Near each west puncture $q \in W_j(T)$ labelled by a Reeb chord ρ,

$$\lim_{z \to q} \pi_\Sigma \circ v(z) = \rho \subset \{-\infty\} \times \partial \Sigma_j.$$

(5) Near each east puncture $q \in E_j(T)$ labelled by a Reeb chord ρ,

$$\lim_{z \to q} \pi_\Sigma \circ v(z) = \rho \subset \{+\infty\} \times \partial \Sigma_j.$$

These moduli spaces can once more be refined by requiring the evaluation maps $\mathrm{ev}_q := \lim_{z \to q} t \circ v(z)$ partition the east and west punctures of T in a prescribed manner. Explicitly, if $P_w := (P_{w,1}, \dots, P_{w,h})$ and $P_e := (P_{e,1}, \dots, P_{e,h})$ are partitions of the sets $W(T) = (W_1(T), \dots, W_h(T))$ and $E(T) = (E_1(T), \dots, E_h(T))$ respectively, denoting

$$\mathrm{ev}_{w,j} := \prod_{q \in W_j(T)} \mathrm{ev}_q : \mathcal{N}(T) \to (e^1)^{|W_j(T)|},$$
$$\mathrm{ev}_{e,j} := \prod_{q \in E_j(T)} \mathrm{ev}_q : \mathcal{N}(T) \to (e^1)^{|E_j(T)|},$$

and $\mathrm{ev}_w := \mathrm{ev}_{w,1} \times \cdots \times \mathrm{ev}_{w,h}$ and $\mathrm{ev}_e := \mathrm{ev}_{e,1} \times \cdots \times \mathrm{ev}_{e,h}$, we write $\mathcal{N}(T; P_w, P_e)$ for the moduli space

$$(\mathrm{ev}_w \times \mathrm{ev}_e)^{-1}(\Delta_{P_{w,1}} \times \cdots \times \Delta_{P_{w,h}} \times \Delta_{P_{e,1}} \times \cdots \times \Delta_{P_{e,h}});$$

omitting P_w from the notation if it is the discrete partition. If we are given a splicing σ_e of the corresponding sequence of Reeb chords given by replacing each $q \in P_e$ by the Reeb chord labelling it, we can restrict these moduli spaces even further, writing $\mathcal{N}(T; P_w, P_e; \sigma_e)$ for the subset $\mathrm{ev}^{-1}(\Delta^1_\sigma, \dots, \Delta^h_\sigma)$ for the evaluation maps from $\mathcal{N}(T; P_w, P_e; \sigma_e) \to e^1$.

Transversality is known in a simple case:

Proposition 4.30 [14, Proposition 5.16] *If all components of T are topological disks, then $\mathcal{N}(T)$ is transversally cut out by the $\bar{\partial}$-equations for any choice of split almost-complex structure J.*

This will allow us to describe the compactifications of our moduli spaces, as in [14, Section 5.4]. We firstly consider small constituent pieces:

Definition 4.31 A simple holomorphic m-comb is a pair (u, v) of a bordered holomorphic polygon $(u, j) \in \mathcal{M}^B_{\leftrightarrow}(\vec{x}; S)$ and holomorphic teeth $v \in \mathcal{N}(T)$ for some bordered polygon source S and tooth source T, together with a one-to-one correspondence between $E(S)$ and $W(T)$ that preserves labelling by Reeb chords, such that $\mathrm{ev}_w(v) = \mathrm{ev}(u)$.

In more generality, we must consider curves with even more degenerate behaviour at east infinity:

Definition 4.32 A holomorphic m-storey is a sequence $U = (u, v_1, \dots, v_k)$ for some $k \geq 0$, where, for some B, bordered polygonal source S and collection of tooth sources $T_1, \dots, T_k$,

- $u \in \mathcal{M}^B_{\leftrightarrow}(\vec{x}; S)$;

- $v_i \in \mathcal{N}(T_i)$;
- (u, v_1) is a simple holomorphic m-comb;
- there is a correspondence between $E(T_i)$ and $W(T_{i+1})$ for $i = 1, \dots, k-1$ which preserves labelling by Reeb chords; and
- $\mathrm{ev}_e(v_i) = \mathrm{ev}_w(v_{i+1})$ for every i.

If $k = 0$, we say that U is *toothless*.

In full generality, we need to consider trees of holomorphic stories.

Definition 4.33 A *tree of holomorphic combs* is:

(1) A metric tree (T, ρ) with corresponding reduced tree $\overline{T}$.

(2) A generator assignment $\vec{\boldsymbol{x}}(\overline{T})$ for $\overline{T}$; a tree of $\vec{\boldsymbol{x}}(\overline{T})$-compatible domains $B(\overline{T})$; a comb-tree of sources $S(\overline{T})$; and a tree of admissible, generic, compatible tree of almost-complex structures $\boldsymbol{J}_T \subset \boldsymbol{J}$.

(3) A set of holomorphic stories $\{U_V\}_{V \in V(\overline{T})}$, such that each U_V is a $J(V)$-holomorphic $|V|$-storey in the moduli space $\mathcal{M}^{B(V)}_{\leftrightarrow}(\vec{\boldsymbol{x}}(V); S(V))$, and for every noncomb vertex $V \in \overline{T}$, the storey U_V is toothless.

If $|V(\overline{T})| = k$, we often say that U has *height* k, and we often refer to a height k tree of holomorphic combs by its vertices: ie $T = (u_1, \dots, u_k)$.

A priori, we must allow holomorphic curves to have nodal sources to ensure compactness holds. As in [14, Section 5.4], we refer to a tree of elementary holomorphic combs as *nodal* or *smooth* if we want to explicitly allow or forbid nodes.

Recall that to each spinal vertex $V_1, \dots, V_k$ of $\overline{T}$ we may associate its storied tree $T(V_i)$. Each of these has an associated domain $B(V_i)$, equal to the sum of the domains $\{B(V')\}_{V' \in T(V_i)}$, and so we may associate to any T-tree U the domain B given by the sum of the $B(V_i)$. We can also associate to U a set of well-defined Reeb chords which is given by the Reeb chords of the right-most tooth of each storey in the spine.

As in [14, Section 5.4], we call a component C of a T-tree U Σ-*stable* (resp. D-stable) if $(\pi_\Sigma \circ U)|_C$ (resp. $(\pi_D \circ U)|_C$) is a stable map, and Σ-unstable (resp. D-unstable) otherwise. We modify the definition of convergence from [14] to general T-trees of elementary combs:

Definition 4.34 We shall say that a sequence (u_n, j_n) of holomorphic n-gons in $\Sigma_{\bar{e}} \times D_m$ *converges* to a holomorphic tree U if:

(1) Let S_Σ denote the result of collapsing all Σ-unstable components of U. Then the map $\{\pi_\Sigma \circ u_n\}$ converges to $\pi_\Sigma \circ U|_{S_\Sigma}$, in the sense of [4, Section 7.3].

(2) The complex structures j_n on D_m converge to the tree of almost complex structures $\{j_v\}_{v\in T}$ over $\{\Sigma_{\bar{e}} \times D_{|v|}\}_{v\in T}$.

(3) Let S_D be the result of collapsing all D-unstable components of U. Then $\{\pi_D \circ u_n\}$ converges to $\pi_D \circ U|_{S_D}$ in the sense of [4, Section 7.3].

(4) For n sufficiently large, all u_n represent the same homology class $B = [U]$.

As is tradition, to avoid drowning in notation, we leave the more general definition of convergence of T-trees of elementary holomorphic combs to the reader.

With this in hand, we make the following definition:

Definition 4.35 Denote by

- $\overline{\overline{\mathcal{M}}}^B(\vec{x}; S)$ the space of all trees of elementary holomorphic combs with corresponding preglued source S in homology class B, asymptotic to the sequence of generators $\vec{x} = (x^{0,1}, \dots, x^{m-1,0})$, compatible with the almost complex structures $\boldsymbol{J}$;
- $\overline{\mathcal{M}}^B(\vec{x}; S)$ the closure of the moduli space $\mathcal{M}^B(\vec{x}; S)$ in $\overline{\overline{\mathcal{M}}}^B(\vec{x}; S)$;
- $\overline{\overline{\mathcal{M}}}^B(\vec{x}; S; P; \sigma)$ the space of all relevant holomorphic trees respecting the partition P;
- $\overline{\mathcal{M}}^B(\vec{x}; S; P; \sigma)$ the closure of $\mathcal{M}^B(\vec{x}; S; P; \sigma)$ in $\overline{\overline{\mathcal{M}}}^B(\vec{x}; S; P)$.

To save on ink and patience, we observe the convention that for all other moduli spaces $\mathcal{M}^B(*)$, we denote by $\overline{\mathcal{M}}^B(*)$ the closure of $\mathcal{M}^B(*)$ in the corresponding moduli space $\overline{\mathcal{M}}^B(\vec{x}; S; P; \sigma)$.

Indeed, these are the required notions for compactifying the moduli spaces:

Proposition 4.36 *The spaces above are compact*; *ie if $\{U_n\}$ is a tree of holomorphic polygon combs in a fixed homology class with fixed preglued topological source S, then $\{U_n\}$ has a subsequence which converges to a tree of holomorphic polygon combs U in the same homology class.*

Proof The argument from [14, Proposition 5.24] carries over readily to this more general case. In particular, [14, Propositon 5.27] applies to a general symmetric cylindrical almost-complex manifold, and doubling both source and $X_{\mathcal{D},m}$ is still a manifold of this type. □

4.7 Holomorphic anchors and perturbations

We discuss modifying the Lagrangian cylinders used in the definition of the moduli spaces

$$\mathcal{M}^B(\vec{x}; S; \vec{\rho}_{\text{🜁}}, \vec{\rho}_{\text{🜂}}; \vec{\sigma}_{\text{🜁}}, \vec{\sigma}_{\text{🜂}}).$$

4.7.1 Interpolating cylinders Fix a partially splayed bordered diagram $\mathcal{D}' = \mathcal{D}^{\partial}(\varepsilon, \vec{\imath})$ with attaching curves $\boldsymbol{\gamma}^0, \boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$, and recall the east-compactified curves $\boldsymbol{\gamma}^1_{\bar{e}} = \boldsymbol{\gamma}^{1,a}_{\bar{e}} \cup \boldsymbol{\gamma}^{1,c}_{\bar{e}}$ with corresponding cylinders in the east-compactification of $X_{\mathcal{D}',m}$ given by

$$C^1_j := \gamma^1_{\bar{e}} \times e^1 \subset \Sigma_{\bar{e}} \times e^1.$$

We can perturb these cylinders using the Hamiltonian H as described in Section 3. Namely, we put

$$C^1(\varepsilon) = \begin{cases} H(\boldsymbol{\gamma}^1_{\bar{e}}, 0) \times t & \text{for } t \in \left(-\infty, -\frac{\varepsilon}{2}\right], \\ H\left(\boldsymbol{\gamma}^1_{\bar{e}}, \frac{2t+\varepsilon}{2(m+1)}\right) \times t & \text{for } t \in \left[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right], \\ H\left(\boldsymbol{\gamma}^1_{\bar{e}}, \frac{\varepsilon}{m+1}\right) \times t & \text{for } t \in \left[\frac{\varepsilon}{2}, \infty\right). \end{cases}$$

(recall that we fixed an identification of a neighbourhood of e^1 with $\mathbb{R} \times [0, \varepsilon)$). For $t \leq -\frac{\varepsilon}{2}$, these cylinders coincide with the cylinders $C^1 = \boldsymbol{\gamma}^1_{\bar{e}} \times e^1$, and for $t \geq \frac{\varepsilon}{2}$, they coincide with the cylinders $(C^1)' = H(\boldsymbol{\gamma}^1_{\bar{e}}, \varepsilon/(m+1)) \times e^1$. They are Lagrangian with respect to a perturbed symplectic form upon the east compactification of $X_{\mathcal{D}',m}$, as described in [13, Equation 3.25].

4.7.2 Moduli spaces with perturbed cylinders Consider again a partially splayed diagram $\mathcal{D}^\partial(\varepsilon, \vec{\iota})$, with attaching curves labelled $\boldsymbol{\gamma}^0, \dots, \boldsymbol{\gamma}^m$. We consider the moduli space of holomorphic curves

$$\mathcal{M}^{B_{\bar{e}}}_{\bar{e},\varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\text{\Large\diamond}}, \vec{\rho}_{\text{\Large\clubsuit}}; \vec{\sigma}; \boldsymbol{J}_{\bar{e}}),$$

which are defined by replacing the unperturbed cylinders $C^i_{\bar{e}} = C^i(\boldsymbol{\gamma}^i_{\bar{e}})$ in the definition of the east-compactified moduli space

$$\mathcal{M}^{B_{\bar{e}}}_{\bar{e}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\text{\Large\diamond}}, \vec{\rho}_{\text{\Large\clubsuit}}; \vec{\sigma}; \boldsymbol{J}_{\bar{e}})$$

with cylinders such that

- $C^0(\varepsilon)$ are the cylinders $\boldsymbol{\gamma}^0_{\bar{e}} \times e^0$,
- $C^1(\varepsilon)$ are the perturbed cylinders $C^1(\varepsilon)$, and
- $C^i(\varepsilon)$ are the cylinders $H\left(\boldsymbol{\gamma}^i_{\bar{e}}, \frac{\varepsilon(m-i)}{m(m+1)}\right) \times e^i$ for all $i \geq 2$.

Note that if $\iota = (\iota_1, \dots, \iota_h)$ is a set of idempotents, then for the corresponding partially splayed diagram $\mathcal{D}^\partial(\varepsilon, \iota \cup \vec{\iota})$ with attaching curves $\boldsymbol{\delta}^0, \boldsymbol{\delta}^1, \dots, \boldsymbol{\delta}^{m+1}$, we have that

- $C^0(\varepsilon)$ agree with $\boldsymbol{\delta}^0_{\bar{e}} \times e^0$,
- $C^1(\varepsilon)$ agrees with $\boldsymbol{\delta}^1_{\bar{e}}$ for $t \leq -\frac{\varepsilon}{2}$ and with $\boldsymbol{\delta}^2_{\bar{e}}$ for $t \geq \frac{\varepsilon}{2}$, and
- for every $i \geq 2$, $C^i(\varepsilon)$ agrees with the cylinders $\boldsymbol{\delta}^{i+1} \times e^i$.

Proposition 4.37 *For small enough ε, all of the moduli spaces of the form*

$$\mathcal{M}(\boldsymbol{J}_{\bar{e}}) := \mathcal{M}^{B_{\bar{e}}}_{\bar{e}}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\text{\Large\diamond}}, \vec{\rho}_{\text{\Large\clubsuit}}; \vec{\sigma}; \boldsymbol{J}_{\bar{e}})$$

are identified with the corresponding moduli space

$$\mathcal{M}_\varepsilon(\boldsymbol{J}_{\bar{e}}) := \mathcal{M}^{B_{\bar{e}}}_{\bar{e},\varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\text{\Large\diamond}}, \vec{\rho}_{\text{\Large\clubsuit}}; \vec{\sigma}; \boldsymbol{J}_{\bar{e}}).$$

Proof For every $m \geq 2$, consider a Lagrangian isotopy $\phi_t = \phi(x,t) \colon X_{\mathcal{D},m} \times [0,1] \to X_{\mathcal{D},m}$ which is supported in a ε-neighbourhood of each $\Sigma_{\bar{e}} \times e^i$, and satisfies that $\phi_t(C^i_{\bar{e}}) = C^i_{\bar{e}}(\varepsilon/t)$ for every $t \in [0,1]$.

A holomorphic curve in the moduli space $\mathcal{M}_{t\varepsilon}(\boldsymbol{J}_{\vec{e}})$ is the same as a holomorphic curve in the moduli space $\mathcal{M}(\boldsymbol{J}^t_{\vec{e}})$, where $\boldsymbol{J}^t_{\vec{e}}$ is the family of perturbed almost-complex structures $\phi_t^*\boldsymbol{J}$. Thus we may identify these two moduli spaces. Moreover, the family of almost-complex structures $\boldsymbol{J}^t_{\vec{e}}$ constitutes a smooth path in the space of families of admissible almost-complex structures. As such, by Proposition 4.21, there is some $T \in (0,1)$ such that

$$\bigcup_{t\in[0,T]} \mathcal{M}_{t\varepsilon}(\boldsymbol{J}_{\vec{e}}) = \bigcup_{t\in[0,T]} \mathcal{M}(\boldsymbol{J}^t_{\vec{e}})$$

provides a cobordism between $\mathcal{M}(\boldsymbol{J}_{\vec{e}}) = \mathcal{M}_0(\boldsymbol{J}_{\vec{e}})$ and $\mathcal{M}_{T\varepsilon}(\boldsymbol{J}_{\vec{e}})$; thus choosing $\varepsilon = T\varepsilon$ provides the result. □

Remark 4.38 There is a similar construction for subdiagrams: if $\mathcal{D}' := \mathcal{D}^\partial(\varepsilon,\vec{\imath})|_S$ is a subdiagram of $\mathcal{D}^\partial(\varepsilon,\vec{\imath})$, and $\vec{\boldsymbol{x}}'$ is a set of generators for $\mathcal{D}'$, there is a corresponding perturbed moduli space

$$\mathcal{M}^{\boldsymbol{B}_{\vec{e}}}_{\vec{e},\varepsilon}(\vec{\boldsymbol{x}}'; S'_{\vec{e}}; \vec{\rho}_{\clubsuit}; \boldsymbol{J}_{\vec{e}})$$

formed by replacing the cylinders $\{C^i \mid i \in S\}$ used in the definition of

$$\mathcal{M}^{\boldsymbol{B}_{\vec{e}}}_{\vec{e}}(\vec{\boldsymbol{x}}'; S'_{\vec{e}}; \vec{\rho}_{\clubsuit}; \boldsymbol{J}_{\vec{e}})$$

with the corresponding perturbed cylinders $\{C^i(\varepsilon) \mid i \in S\}$. The proof of Proposition 4.37 carries over to this case to identify these two moduli spaces.

Similarly, for regularisations of diagrams (Definition 3.18) we also have:

Proposition 4.39 *Let $\mathcal{D}|_S$ be a subdiagram of $\mathcal{D}$, and $(\mathcal{D}|_S)_{\mathrm{reg}}$ be its regularisation. Then the moduli space $\mathcal{M}_{\mathrm{reg}} := \mathcal{M}^{\boldsymbol{B}}_{\mathrm{reg}}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \vec{\sigma})$ defined using the diagram $(\mathcal{D}|_S)_{\mathrm{reg}}$ agrees with the moduli space $\mathcal{M}|_S := \mathcal{M}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \vec{\sigma})$ defined using $\mathcal{D}|_S$.*

Proof This follows closely along the lines of the proof of Proposition 4.37 — one perturbs the cylinders for $\mathcal{D}|_S$ to agree with those for $(\mathcal{D}|_S)_{\mathrm{reg}}$. When $\boldsymbol{\gamma}^i$ and $\boldsymbol{\gamma}^{i+1}$ are small approximations of one another, this identifies $\mathcal{M}|_S$ with a moduli space of curves with boundary contained in Lagrangians $\boldsymbol{\gamma}^i \times e^i$ and $\boldsymbol{\gamma}^i \times e^{i+1}$; applying the removable singularities theorem to extend any u' in this moduli space to a bona fide map in $\mathcal{M}_{\mathrm{reg}}$ yields the result. □

4.7.3 Holomorphic anchors and stretching the neck Later on, we will stretch the neck of the almost-complex structures $\boldsymbol{J}$ along a hypersurface, forcing degenerations of moduli spaces. In this section, we fix notation in preparation for this; although we later largely restrict possible degenerations, we begin here in full generality.

Fix an almost-complex structure $J_{\maltese}$ upon $\Sigma_{\vec{e}} \times \mathbb{H}$, where $\mathbb{H}$ denotes the upper-half plane $\{z \in \mathbb{C} \mid \Im(z) \geq 0\}$. Consider the moduli space of maps

$$u_{\maltese} : (S_{\maltese}, \partial S_{\maltese}) \to (\Sigma_{\vec{e}} \times \mathbb{H}, C^1(\varepsilon) \times \mathbb{R}),$$

where $S_{\maltese}$ is an anchor source.

We require these to satisfy:

(1) $u_{\mathfrak{t}}$ is $(j, J_{\mathfrak{t}})$-holomorphic for some almost complex structure j upon $S_{\mathfrak{t}}$;

(2) $u_{\mathfrak{t}}$ is proper;

(3) $u_{\mathfrak{t}}$ has finite energy;

(4) $u_{\mathfrak{t}}$ is disjoint from $\boldsymbol{p}, \boldsymbol{b}$;

(5) the map $\pi_{\mathbb{H}} \circ u_{\mathfrak{t}}$ is a branched covering;

(6) $u_{\mathfrak{t}}$ is an embedding;

(7) At each puncture q of $S_{\mathfrak{t}}$ labelled by ∞, $\lim_{z \to q}(\pi_D \circ u)(z) = \infty$.

(8) At each puncture q of $S_{\mathfrak{t}}$ labelled by a Reeb chord ρ, we have that upon some neighbourhood $\eta(q)$ of q, $(\pi_\Sigma \circ u)(\eta)$ is modelled upon $B(\rho)$.

(9) For every $t \in \mathbb{R}$ and each $j = 1, \dots, h$, we have that $u^{-1}(C_j^1|_{\Sigma \times \{t\}})$ consists of at most one point, and for each $j = h+1, \dots, g+h-1$, the set $u^{-1}(C_j^1|_{\Sigma \times \{t\}})$ consists of exactly one point.

From these conditions, it follows that near each puncture q of $S_{\mathfrak{t}}$ labelled by ∞, the map $u_{\mathfrak{t}}$ is asymptotic to a chord of the form $\theta^{1,2}(q) \times \infty$, such that $\boldsymbol{\theta}^{1,2} = \bigcup_q \theta^{1,2}(q)$ is a generator for the pair $(\boldsymbol{\gamma}^1_{\bar{e}}, H(\boldsymbol{\gamma}^1_{\bar{e}}, \varepsilon/(m+1)))$, or, more transparently, for the pair $(\boldsymbol{\delta}^1, \boldsymbol{\delta}^2)$ described in the previous subsection. Each such map has an associated homology class $B_{\mathfrak{t}}$ in the relative homology group $H_2(\Sigma_{\bar{e}} \times \mathbb{H}, C^1(\varepsilon))$. We call such curves *holomorphic anchors* and collect them into moduli spaces $\mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}})$. We shall later see that in fact, for the moduli spaces to be nonempty, the generator $\boldsymbol{\theta}^{1,2}$ must be the distinguished generator $\boldsymbol{\theta}^{1,2}_+$.

As with all of our other moduli spaces, we can cut these down according to various partitions of the east punctures of $S_{\mathfrak{t}}$. If $P_{\mathfrak{t}}$ is a set of partitions $P_{\mathfrak{t}} = (P_{\mathfrak{t},1}, \dots, P_{\mathfrak{t},h})$, where each $P_{\mathfrak{t},j}$ is a partition of $E_j(S_{\mathfrak{t}})$, there is a corresponding evaluation map

$$\mathrm{ev} := \prod_{q \in E(S_{\mathfrak{t}})} \mathrm{ev}_q \colon \mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}})) \to (e^1)^{|E(S_{\mathfrak{t}})|},$$

and we write

$$\mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}})$$

for the moduli space $(\mathrm{ev})^{-1}(\Delta_{P_{\mathfrak{t}}})$. If we are further provided with a splicing $\sigma_{\mathfrak{t}}$ of the set of punctures of $S_{\mathfrak{t}}$ labelled with a jumping chord, we can define the moduli spaces $\mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}}; \sigma_{\mathfrak{t}})$ for the subset $\mathrm{ev}^{-1}(\Delta^1_{\sigma_{\mathfrak{t}}}, \dots, \Delta^h_{\sigma_{\mathfrak{t}}})$, where $\mathrm{ev} \colon \mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}}) \to e^1$.

It transpires that for particularly simple sources $S_{\mathfrak{t}}$, the moduli spaces $\mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S)$ are always transversally cut out:

Proposition 4.40 *Let $S_{\mathfrak{t}}$ be a disjoint union of topological discs. Then the moduli spaces $\mathcal{M}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}, S_{\mathfrak{t}})$ are always transversally cut out by the $\bar{\partial}$-equations, for any choice of almost-complex structure $J_{\mathfrak{t}}$.*

Proof This follows along the lines of [14, Proposition 5.16] — namely, any $u = (u_\Sigma, u_{\mathbb{H}})$ doubles to a pair of maps from a punctured CP^1 to a punctured CP^1. The first Chern class of these doubled maps is equal to the Maslov index of u_Σ and $u_{\mathbb{H}}$, which is -1 in both cases (it is straightforward to see that each is a small anticlockwise disc). Thus by [16, Lemma 3.3.1], $D\bar{\partial}$ is surjective for this doubled map, which implies surjectivity for the original map — see, for instance, [8, page 158]. □

These anchors will arise as the result of "stretching the neck" of bordered holomorphic polygons. Explicitly, for every m fix an ε-radius half-disc neighbourhood η^m_ε of the point $\{0\} \in e^1 \subset \partial D_m$. We consider the family $\boldsymbol{J}^t$ obtained from $\boldsymbol{J}$ by inserting a neck of length t along the boundary of each η^m_ε, as described in [10, Section A.2].

Suppose that $\mathrm{ind}(B, S, \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}, \sigma_{\blacktriangle}) = 0$, and let

$$\boldsymbol{\mathcal{M}}^B_{\geq T}(\boldsymbol{x}; S; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J})$$

denote the set $\bigcup_{t \geq T} \mathcal{M}^B(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$. By a standard transversality argument, for generic choice of $\boldsymbol{J}$ and generic T, the moduli spaces

$$\boldsymbol{\mathcal{M}}^B_{\geq T} := \boldsymbol{\mathcal{M}}^B_{\geq T}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$$

are one-dimensional manifolds. Moreover, for small enough ε, the moduli spaces

$$\boldsymbol{\mathcal{M}}^{B_{\bar{e}}}_{\geq T, \bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$$

defined to be equal to $\bigcup_{t \geq T} \mathcal{M}^{B_{\bar{e}}}_{\bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$ are also transversally cut out, and form one-manifolds homeomorphic to $\boldsymbol{\mathcal{M}}^B_{\geq T}$ (cf Proposition 4.37). By Gromov compactness, in both instances, there is some T such that for every $t > T$, each constituent moduli space $\mathcal{M}^B(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$ (resp. $\mathcal{M}^{B_{\bar{e}}}_{\bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$) is transversally cut out.

The perturbed moduli spaces admit compactifications similar to the aforementioned trees of holomorphic polygons; but there is one more possible type of degeneration which we force in the neck-stretching process. This is an *anchored holomorphic curve*:

Definition 4.41 An *anchored holomorphic polygon* is a pair $(u, u_{\maltese})$ where

$$u \in \mathcal{M}^B(\vec{\boldsymbol{x}}, \boldsymbol{\theta}^{0,1}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \sigma_{\blacktriangle}, \sigma_{\maltese}; \boldsymbol{J})$$

is a bordered holomorphic polygon and $u_{\maltese} \in \mathcal{M}_{\maltese}(\boldsymbol{\theta}^{1,2}; S_{\maltese}; P; \sigma_{\maltese})$ is a holomorphic anchor. (See Figure 14 for a schematic.)

We write

$$\overline{\boldsymbol{\mathcal{M}}}^{B_{\bar{e}}}_{\geq T, \bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t)$$

for the compactification of

$$\boldsymbol{\mathcal{M}}^{B_{\bar{e}}}_{\geq T, \bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; \vec{\rho}_{\blacktriangle}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\blacktriangle}, \vec{\sigma}_{\maltese}; \boldsymbol{J}^t),$$

which can be defined in much the same way as in Definition 4.35, but allowing for the bordered polygons in each definition to be anchored holomorphic polygons.

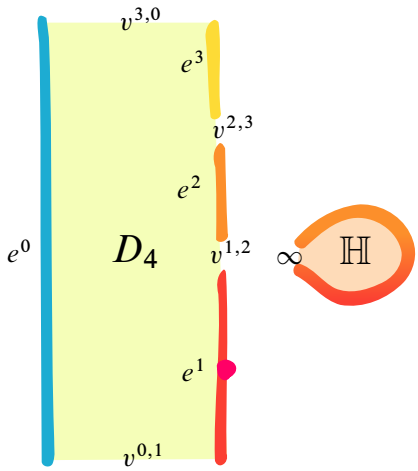


Figure 14: A schematic of an anchored holomorphic polygon. The pink dot indicates $\pi_D \circ u$ of an east puncture of the source S.

4.8 Gluing for trees of elementary holomorphic combs and anchors

We now quote the relevant gluing results, and then explore their consequences.

Proposition 4.42 *Let $U = (u_1, u_2)$ be a height 2 tree of (bordered) holomorphic polygons where $u_1 \in \widetilde{\mathcal{M}}^{B_1}(\vec{\boldsymbol{x}}(V_1), S_1; P_1, Q_1; \sigma_1)$ and $u_2 \in \widetilde{\mathcal{M}}^{B_2}(\vec{\boldsymbol{x}}(V_2), S_2; P_2, Q_2; \sigma_2)$ are spinal.*

Let $B = B_1 * B_2$, S denote the splayed source $S_T = S_1 \natural S_2$, with induced distinguished set of noneasterly punctures $Q = Q_1 \cup Q_2$, and finally let $\sigma = \sigma_1 \star \sigma_2$.

For some sufficiently small open neighbourhoods U_1 of u_1 and U_2 of u_2, there is an open neighbourhood of (u_1, u_2) in $\overline{\mathcal{M}}^B(\vec{\boldsymbol{x}}(T), S; P_1 \star P_2, Q; \sigma)$ which is homeomorphic to $U_1 \times U_2 \times [0, 1)$.

We also require gluing for two-storey combs with one easterly vertex:

Proposition 4.43 *Let $U = (u_1, u_2)$ be a height 2 tree of (bordered) holomorphic polygons where $u_1 \in \widetilde{\mathcal{M}}^{B_1}(\vec{\boldsymbol{x}}(V_1), S_1; P_1, Q; \sigma_1)$ is spinal and $u_2 \in \widetilde{\mathcal{M}}^{B_2}(\vec{\boldsymbol{x}}(V_2), S_2; P_2; \sigma_2)$ is easterly.*

Let $B = B_1 * B_2$, S denote the splayed source $S_T = S_1 \natural S_2$, with induced distinguished set of noneasterly punctures Q', and finally let $\sigma = \sigma_1 \star \sigma_2$.

For some sufficiently small open neighbourhoods U_1 of u_1 and U_2 of u_2, there is an open neighbourhood of (u_1, u_2) in $\overline{\mathcal{M}}^B(\vec{\boldsymbol{x}}(T), S; P_1 \star P_2, Q; \sigma)$ which is homeomorphic to $U_1 \times U_2 \times [0, 1)$.

We also require gluing for simple combs:

Proposition 4.44 *Let (u, v) be a simple m-comb, with $u \in \widetilde{\mathcal{M}}^B(\vec{\boldsymbol{x}}; S; Q; J)$, and $v \in \mathcal{N}(T, P_e; \sigma)$, and suppose the corresponding moduli spaces are transversally cut out at u and v respectively, and the evaluation maps $\mathrm{ev}\colon \widetilde{\mathcal{M}}^B(\vec{\boldsymbol{x}}; S; Q; J) \to e^1$ and $\mathrm{ev}_w\colon \mathcal{N}(T; P_e; \sigma) \to e^1$ are transverse at (u, v).*

If we denote by $S' = S \natural T$, and Q' the corresponding set of distinguished noneasterly punctures induced by S, then for sufficiently small open neighbourhoods U_u of u and U_v of v, there is an open neighbourhood of (u, v) in $\overline{\mathcal{M}}^B(\vec{x}; S'; P, Q; \sigma)$ homeomorphic to $(U_u \times_{\mathrm{ev}} U_v) \times [0, 1)$.

Finally, we require the following result for anchored holomorphic curves:

Proposition 4.45 *Let u' be a splayed holomorphic polygon in*

$$\mathcal{M}_{\vec{e}}^{B'}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S'; P', Q'; \sigma'; \boldsymbol{J}),$$

and $u_{\text{ŧ}} \in \mathcal{M}_{\text{ŧ}}^{B_{\text{ŧ}}}(\boldsymbol{\theta}^{1,2}; S_{\text{ŧ}}; P_{\text{ŧ}}; \sigma_{\text{ŧ}}; J_{\text{ŧ}})$ a holomorphic anchor, and suppose that these moduli spaces are transversally cut out at u and $u_{\text{ŧ}}$ respectively.

Denote by $B = B' * B_{\text{ŧ}}$, S the surface formed by gluing S' and $S_{\text{ŧ}}$ at punctures labelled by $v^{1,2}$, $P = P' \star P_{\text{ŧ}}$, $\sigma = \sigma' \star \sigma_{\text{ŧ}}$, and Q the set of distinguished noneast punctures of S induced by Q'. Then, for sufficiently small open neighbourhoods U and $U_{\text{ŧ}}$ containing u and $u_{\text{ŧ}}$ respectively, there is an open neighbourhood of $(u, u_{\text{ŧ}})$ in the moduli space

$$\overline{\boldsymbol{\mathcal{M}}}_{\geq T}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k-1), \boldsymbol{y}; S; P, Q; \sigma)$$

homeomorphic to $U \times U_{\text{ŧ}} \times [0, 1)$.

Proof of Propositions 4.42, 4.43, 4.44 and 4.45 A straightforward combination of [14, Proposition 5.30] and [13, Proposition 4.21] gives the first three of these results when dropping the relevant instances of the letter Q from the statements, and the last follows similarly. In each case, as the punctures of each relevant source are mapped either to east infinity or to some boundary puncture of the relevant disc, the fact that the manifolds $X_{\mathcal{D},m}$ have two types of infinity is irrelevant, so the result follows by standard gluing techniques (the provision of splicings behaves no differently than the other partitions of east punctures). To include the provision of cut behaviour, one sees that the nearby families of holomorphic curves constructed by the implicit function theorem in the proof of [14, Proposition 5.30] and [13, Proposition 4.21] must have small cuts near the relevant punctures themselves — thus, by the discussion in Section 4.6.2 there are nearby families of curves with no cut at these punctures.

For example, for Proposition 4.42, one constructs preglued maps $u_1 \natural_R u_2$ with sources $S_1 \natural S_2$ which converge to the two-storey building. One shows that $D\bar{\partial}(u_1 \natural_R u_2) \to 0$ as $R \to \infty$, and applies the implicit function theorem to find a neighbourhood $N_\delta = N_\delta(u_1 \natural_R u_2)$ such that the set of zeroes of the $\bar{\partial}$-equations in N_δ form a smooth manifold of the correct dimension. As such, choosing R sufficiently large finds families of solutions to the $\bar{\partial}$-equations near to $u_1 \natural_R u_2$.

If S_1 and S_2 have distinguished sets of punctures Q_1 and Q_2 respectively, where $Q_1 \cup Q_2 = Q$, it follows that the preglued maps $u_1 \natural_R u_2$ have no cut at the corresponding punctures of $S_1 \natural S_2$. As such, any holomorphic curve in the neighbourhood N_δ has (a priori) at least a very small cut at the relevant punctures. By the discussion in Section 4.6.2, it follows that the set $\bar{\partial}^{-1}(0) \cap N_\delta$ can be parameterised

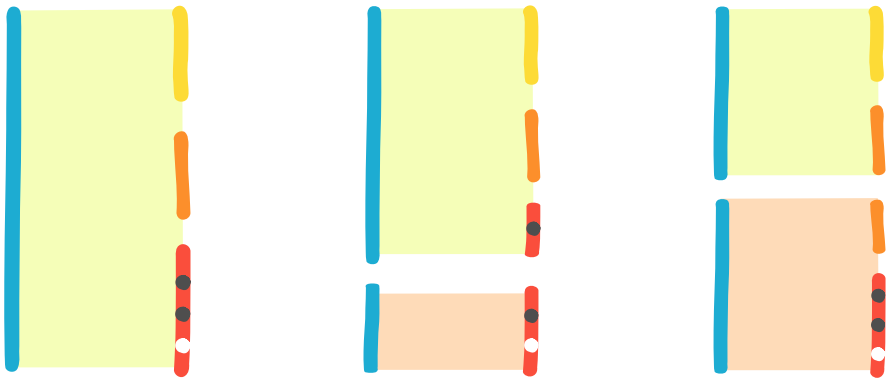


Figure 15: A schematic picture of a one-dimensional moduli space (left), with two types of spinal ends shown (centre and right). The breaks in the boundary indicate punctures of D_n, whereas the dots indicate east punctures — with punctures corresponding to one east infinity shown in grey, and a puncture corresponding to another shown in white.

by cut length as $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q; \vec{\sigma}) \times (-\varepsilon, \varepsilon)^k$ for some k. It follows by the triangle inequality that the set $\widetilde{\mathcal{M}}^B(\vec{x}; S; \vec{\rho}_{\clubsuit}, Q; \vec{\sigma})$ intersects B_δ, and thus provides a family of holomorphic curves with the correct cut behaviour near to $u_1 \natural_R u_2$ and thus near to (u_1, u_2), as in [10, Section A.4] (see also [3, Section 5.3.3]). □

4.9 Restrictions upon easterly curves and the boundary of one-dimensional moduli spaces

In this section, we prove the main result which we shall use to show that maps we define in the next section satisfy the partial $\mathcal{A}_\infty$ relations. We will need to understand the modulo-two count of the ends of moduli spaces of expected dimension 1 in more detail. As in [14, Definition 5.58], we give names to the types of ends that occur. Some of these are illustrated in Figures 15, 16 and 17 — recall that we distinguished some special types of vertices of a tree in Definition 4.4.

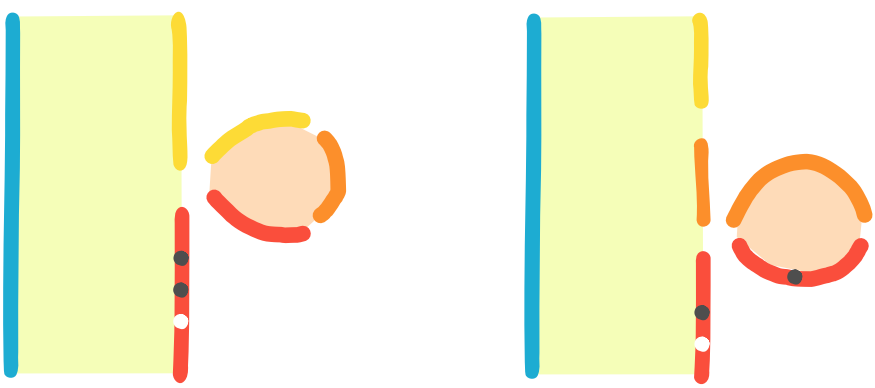


Figure 16: A schematic picture of two types of mixed ends of the one-dimensional moduli space on the left of Figure 15.

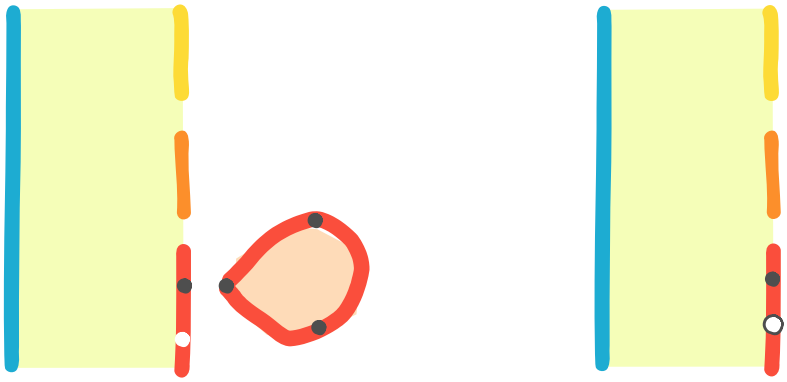


Figure 17: A schematic picture of a composition end (left) and a collision end (right) of the moduli space on the left of Figure 15.

Definition 4.46 Let $\mathcal{M} := \mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\spadesuit}; \sigma_{\clubsuit}, \sigma_{\spadesuit})$ be a one-dimensional moduli space. Then:

(1) A *spinal two-storey end* of $\mathcal{M}$ is a height two spinal tree of prescribed-cut polygons (u, v), where

$$u \in \mathcal{M}^{B_1}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+, \boldsymbol{w}; S_1; \vec{\rho}^1_{\clubsuit}, \vec{\rho}^1_{\spadesuit}; \sigma^1_{\clubsuit}, \sigma^1_{\spadesuit})$$

and

$$v \in \mathcal{M}^{B_2}(\boldsymbol{w}, \vec{\boldsymbol{\theta}}_+, \boldsymbol{y}; S_2; \vec{\rho}^2_{\clubsuit}, \vec{\rho}^2_{\spadesuit}; \sigma^2_{\clubsuit}, \sigma^2_{\spadesuit})$$

are spinal, and such that $B_1 \in \pi_2^{\vec{\boldsymbol{\theta}}_+}(\boldsymbol{x}, \boldsymbol{w})$ and $B_2 \in \pi_2^{\vec{\boldsymbol{\theta}}_+}(\boldsymbol{w}, \boldsymbol{z})$ are such that $B_1 * B_2 = B$, $S_1 \natural S_2 = S$, $\vec{\rho}^1_{\clubsuit} \star \vec{\rho}^2_{\clubsuit} = \vec{\rho}_{\clubsuit}$, $\vec{\rho}^1_{\spadesuit} \star \vec{\rho}^2_{\clubsuit} = \vec{\rho}_{\spadesuit}$, $\sigma^1_{\clubsuit} \star \sigma^2_{\clubsuit} = \sigma_{\clubsuit}$ and $\sigma^1_{\spadesuit} \star \sigma^2_{\spadesuit} = \sigma_{\spadesuit}$.

(2) A *mixed two-storey end* of $\mathcal{M}$ is a height two tree (u, v) such that

$$u \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}_+^{1,2}, \dots, \boldsymbol{\theta}_+^{\kappa-1,\kappa}, \boldsymbol{\theta}^{\kappa,\kappa+\lambda}, \boldsymbol{\theta}_+^{\kappa+\lambda,\kappa+\lambda+1}, \dots, \boldsymbol{\theta}_+^{k-1,k}, \boldsymbol{y}; S'; \vec{\rho}'_{\clubsuit}, \vec{\rho}_{\spadesuit}; \sigma'_{\clubsuit}, \sigma_{\spadesuit})$$

is spinal and

$$v \in \mathcal{M}^{B_E}(\boldsymbol{\theta}_+^{\kappa,\kappa+1}, \dots, \boldsymbol{\theta}_+^{\kappa+\lambda-1,\kappa+\lambda}, \boldsymbol{\theta}^{\kappa+\lambda,\kappa}; S_E; \vec{\rho}_E)$$

is easterly, and such that $B' * B_E = B$ and $S' \natural S_E = S$.

(3) An (i, j)-*composition end* is an element of $\mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S'; \vec{\rho}'_{\clubsuit}, \vec{\rho}_{\spadesuit}; \sigma'_{\clubsuit}, \sigma_{\spadesuit})$, where $\vec{\rho}'_{\clubsuit}$ is given by $\bar{\mu}^i_j(\vec{\rho}_{\clubsuit})$ for some i, j; $\sigma'_{\clubsuit}$ is given by σ^i_j; and S' is obtained from S by contracting the arc of ∂S connecting the punctures corresponding with ρ^i_j and ρ^{i+1}_j, and labelling the corresponding puncture $\rho^i_j \cdot \rho^{i+1}_j$.

(4) A κ-*collision end* is an element of $\mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\spadesuit}; \sigma_{\clubsuit}(\kappa), \sigma_{\spadesuit})$.

(5) A *cut-vanishing end* of $\mathcal{M}$ is a prescribed-cut polygon u in the moduli space

$$\mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S'; \vec{\rho}'_{\clubsuit}, \vec{\rho}'_{\spadesuit}; \sigma'_{\clubsuit}, \sigma'_{\spadesuit})$$

such that $\vec{\rho}'_{\clubsuit} \star \vec{\rho}'_{\diamondsuit}$ is a 123-reduction of $\vec{\rho}_{\clubsuit} \star \vec{\rho}_{\diamondsuit}$ formed by replacing an adjacent pair ρ_1, ρ_{23} or ρ_{12}, ρ_3 with ρ_{123}, where ρ_1 (resp. ρ_3) is an element of $\vec{\rho}_{\diamondsuit}$. Here, S' is given by contracting the arc in ∂S between the punctures corresponding with ρ_1, ρ_{23} (resp. ρ_{12}, ρ_3) and labelling the resultant puncture by ρ_{123}.

(The condition on a cut-vanishing end essentially is the same as $\vec{\rho}'_{\diamondsuit}$ being a 123-reduction of $\vec{\rho}_{\diamondsuit}$, except there is a possibility for the first chord of $\vec{\rho}_{\diamondsuit,j}$ to have been ρ_3 and the final chord of $\vec{\rho}_{\clubsuit,j}$ to have been ρ_{12}.)

With this in mind, we state the first main theorem of this section:

Theorem 4.47 *Suppose that*

$$\mathcal{M} = \mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\diamondsuit}; \sigma_{\clubsuit}, \sigma_{\diamondsuit})$$

is one-dimensional. Then the total number of spinal two-storey ends of $\mathcal{M}$, mixed two-storey ends of $\mathcal{M}$, (i,j)-composition ends of $\mathcal{M}$, κ-collision ends of $\mathcal{M}$, and cut-vanishing ends of $\mathcal{M}$ is even.

Proof This follows readily along the same lines as the proof of [14, Theorem 5.61]. In particular, trees of height greater than two are ruled out in a very similar way to [14, Proposition 5.43]. In our situation, however, there are two types of two-storey ends. Height one combs still degenerate, but due to the simplicity of the diagram near the boundary, join or shuffle curves cannot appear, just sequences of split curves (corresponding with composition and collision ends of $\mathcal{M}$). A priori, nodal curves can still degenerate, but are ruled out along the lines of [14, Section 5.6.3].

One may truncate $\overline{\mathcal{M}}$ near composition and collision ends, as in the proof of [14, Theorem 5.61] to give a moduli space $\overline{\mathcal{M}}_{\text{cropped}}$ which is a compact 1-manifold with boundary. This has spinal two-storey ends if the complex structure on D_{k+2} degenerates along an arc with one end on the edge e^0, mixed two-storey ends if an arc without an end upon e^0 degenerates, a composition or collision end if a holomorphic tooth degenerates at east infinity, and finally a cut-vanishing end if the length of a cut at some corner of $\rho_{123}(S)$ degenerates to zero. Ends where more than one cut degenerates, or a cut degenerates at the same time as another type of degeneration, are ruled out by the index formula in Proposition 4.28 — each degenerated length subtracts one from the expected dimension; thus if more than one cut degenerated the corresponding moduli spaces would have negative dimension and so be empty. □

We now examine the types of mixed two-storey ends which may occur. Fix a partially splayed diagram $\mathcal{D} = \mathcal{D}'(\varepsilon, \vec{\imath})$ with attaching curves $\boldsymbol{\gamma}^0, \boldsymbol{\gamma}^1, \dots, \boldsymbol{\gamma}^m$, where we recall that $\boldsymbol{\gamma}^1$ splits as $\boldsymbol{\gamma}^1 = \boldsymbol{\gamma}^{1,a} \cup \boldsymbol{\gamma}^{1,c}$. Fix also a subdiagram $\mathcal{D}|_S$ where S is a contiguous subsequence of $\{1, \dots, m\}$. For clarity's sake, relabel the attaching curves of this subdiagram by $\boldsymbol{\delta}^1, \dots, \boldsymbol{\delta}^\lambda$, and the corresponding sequence of generators by $(\boldsymbol{\theta}_+^{\kappa,\kappa+1}, \dots, \boldsymbol{\theta}_+^{\kappa+\lambda-1,\kappa+\lambda}) := \vec{\boldsymbol{\theta}}_+(\lambda)$ and $\boldsymbol{\theta} = \boldsymbol{\theta}^{\kappa+\lambda,\kappa} \in \mathcal{G}(\boldsymbol{\delta}^\lambda, \boldsymbol{\delta}^1)$. We call the resultant generator for $\mathcal{D}|_S$ by $\vec{\boldsymbol{\theta}}(\lambda) := (\vec{\boldsymbol{\theta}}_+(\lambda), \boldsymbol{\theta})$. We first restrict B_E to some fairly manageable possibilities:

Proposition 4.48 *Let $B_E \in \pi_2(\vec{\boldsymbol{\theta}}(\lambda)) \subset \mathrm{Dom}(\mathcal{D}|_S)$. Then it is supported within the approximation region for the curves $\boldsymbol{\delta}^1, \dots, \boldsymbol{\delta}^\lambda$.*

Proof This follows from the homological linear independence of the bordered diagram $\mathcal{D}'$. The corresponding domain $B'_E \in \mathrm{Dom}(\mathcal{D}')$ given by taking multiplicities away from the approximation region has boundary contained solely in $\boldsymbol{\alpha}$ and multiplicity zero at $\boldsymbol{b}$ — contradicting homological linear independence unless all multiplicities away from the approximation region are zero. □

It follows from this that the domain B_E decomposes as a sum $B_E = B_1 + \cdots + B_{g+h-1}$, where each B_j is supported in the approximation region for $\delta^i_j, \delta^{i+1}_j$.

The next step is to separate the curves into the more widely studied case where they are each approximations of one another, and the case where they change in a given index. As such, we let

$$\Gamma := \left\{ j \in \{1,\dots,h\} \mid \mathrm{jump}(\vec{\rho}_{E,j}) = \varnothing \text{ and } \delta^1_j \approx \cdots \approx \delta^l_j \right\} \cup \{h+1,\dots,g+h-1\},$$

and $\overline{\Gamma} := \{1,\dots,h\} - (\Gamma \cap \{1,\dots,h\})$.

We now show that the domains of the less widely studied case are supported within the handles.

Proposition 4.49 *Suppose that $B_E \in \pi_2(\vec{\boldsymbol{\theta}}(\lambda))$, and let $j \in \overline{\Gamma}$. Then the domain B_j is supported within the handle $\mathfrak{H}_j$.*

Proof Any domain B_j not supported within $\mathfrak{H}_j$ but supported within the approximation region can only have corners upon points of the form $\theta^{i_k,i_{k+1}}_{+,j}$ where $\delta^{i_k}_j \approx \delta^{i_{k+1}}_j$. As $j \in \overline{\Gamma}$, either one coordinate $\theta(\lambda)^i_j$ is equal to some $e^{\infty,i}_j$, or B_j is required to have a corner at e^∞_j — neither of which are of this form. □

For each $j \in \Gamma$, then, if δ^1 is a set of $2h$ arcs, any map $u \in \mathcal{M}^{B_E}(\boldsymbol{\theta}(\lambda); S_E; \vec{\rho}_E)$ has boundary in just one of the cylinders $\delta^{1,m}_j \times e^1$ or $\delta^{1,l}_j \times e^1$. We label the corresponding curve δ^1_j. Let $\dot{\Sigma}_\Gamma$ denote the surface given by removing the handles $\bigcup_{j\in\overline{\Gamma}} \mathfrak{H}_j$ from the east-compactified splayed surface $\Sigma(\varepsilon,\vec{\imath})_{\bar{e}}$, and Σ_Γ the result of collapsing the remaining boundary components $\mathfrak{H}_{L,j}$ and $\mathfrak{H}_{R,j}$ in $\dot{\Sigma}_\Gamma$ to $2|\overline{\Gamma}|$ single points $\boldsymbol{p}_{\overline{\Gamma}}$. This contains the curves $\boldsymbol{\delta}^i_\Gamma = \{\delta^i_{\bar{e},j} \mid j \in \Gamma\}$, which approximate one another and basepoints $\boldsymbol{p}_\Gamma := \{p_j \mid j \in \Gamma\}$ and thus we form the closed multidiagram $\mathcal{D}_\Gamma$ given by the tuple $(\Sigma_\Gamma; \boldsymbol{\delta}^1_\Gamma, \dots, \boldsymbol{\delta}^l_\Gamma; \boldsymbol{p}_\Gamma \cup \boldsymbol{p}_{\overline{\Gamma}})$.

It follows from the above proposition that for $j \in \Gamma$ the domains B_j are contained in $\dot{\Sigma}_\Gamma$, and thus have a corresponding domain $B_\Gamma \in \mathrm{Dom}(\mathcal{D}_\Gamma)$, given by the sum of their images in Σ_Γ. Moreover, the domain B_Γ lives in the homology group $\pi_2(\vec{\boldsymbol{\theta}}(\lambda)|_\Gamma)$, where $\vec{\boldsymbol{\theta}}(\lambda)|_\Gamma$ is the restriction of $\vec{\boldsymbol{\theta}}(\lambda)$ to the curves $\boldsymbol{\delta}^1_\Gamma, \dots, \boldsymbol{\delta}^\lambda_\Gamma$.

Considering the surface $\dot{\Sigma}_\Gamma$ as a subset of both surfaces $\Sigma(\varepsilon,\vec{\imath})_{\bar{e}}$ and Σ_Γ, we choose a family of almost-complex structures $\boldsymbol{J}'$ upon $\Sigma_\Gamma \times D_l$ such that their restriction to $\dot{\Sigma}_\Gamma \times D_l$ agrees with the restriction of $\boldsymbol{J}$.

For every $j \in \overline{\Gamma}$, we identify the handle $\mathfrak{H}_j$ with the punctured torus $\Sigma_j := T^2 - e^\infty_j$ formed by gluing the ends $\mathfrak{H}_{L,j}$ and $\mathfrak{H}_{R,j}$ (more properly, we identify the handle with its end circles removed with an open subset of the torus). The curves and arcs $\delta^1_j \cap \mathfrak{H}_j, \dots, \delta^\lambda_j \cap \mathfrak{H}_j$ have natural images in T^2, for

which we abuse notation and call $\delta_j^1,\dots,\delta_j^\lambda$. Together with the image of the basepoint b_j, the tuple $\mathcal{D}_j := (\Sigma_j;\delta_j^1,\dots,\delta_j^\lambda;b_j)$ forms a toroidal bordered Heegaard multidiagram. The domain B_j has a natural image in $\mathrm{Dom}(\mathcal{D}_j)$ which we also denote by B_j, and lives in the group $\pi_2(\theta(\lambda)_j^1,\dots,\theta(\lambda)_j^\lambda)$. We choose a family of almost-complex structures $\boldsymbol{J}_j$ upon $\Sigma_j\times D_\lambda$ which agree with the almost-complex structure $\boldsymbol{J}$ upon $\mathfrak{H}_j\times D_\lambda$.

It follows that the moduli space $\mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda);S_E;\vec{\rho}_{\clubsuit})$ splits as a fibred product

$$\mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda);S_E;\vec{\rho}_E)=\mathcal{M}_\Gamma\times_{\mathrm{Conf}(D_\lambda)}\mathcal{M}_1\times_{\mathrm{Conf}(D_\lambda)}\cdots\times_{\mathrm{Conf}(D_\lambda)}\mathcal{M}_{|\overline{\Gamma}|},$$

where $\mathcal{M}_\Gamma=\mathcal{M}^{B_\Gamma}(\vec{\boldsymbol{\theta}}(\lambda)|_\Gamma;S_\Gamma;\boldsymbol{J}')$ is the moduli space defined by counting holomorphic polygons according to $\mathcal{D}_\Gamma$ as defined in Section 4.5, and each $\mathcal{M}_j$ is the moduli space

$$\mathcal{M}^{B_j}(\theta(\lambda)_j^{1,2},\dots,\theta(\lambda)_j^{\lambda,1};S_j;\vec{\rho}_{\clubsuit,j};\boldsymbol{J}_j)$$

defined by counting bordered polygons according to $\mathcal{D}_j$, as defined in Section 4.6.

Proposition 4.50 *There is a unique homology class $D_\Gamma\in\mathrm{Dom}(\mathcal{D}_\Gamma)$ which carries a holomorphic representative in the moduli space $\mathcal{M}_\Gamma$. This lives in the homology group $\pi_2(\vec{\boldsymbol{\theta}}_+(\lambda)|_\Gamma)$. Moreover, the projection of the moduli space $\mathcal{M}_\Gamma$ onto $\mathrm{Conf}(D_\lambda)$ has degree one for this homology class; thus $\mathcal{M}^{B_E}(\boldsymbol{\theta}(\lambda);S_E;\vec{\rho}_{\clubsuit})$ is nonempty only if the corresponding $B_\Gamma=D_\Gamma$, and $\vec{\boldsymbol{\theta}}(\lambda)|_\Gamma=\vec{\boldsymbol{\theta}}_+(\lambda)|_\Gamma$. In this case, $\mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda);S_E;\vec{\rho}_E)$ has the same number of elements as*

$$\mathcal{M}_1\times_{\mathrm{Conf}(D_\lambda)}\cdots\times_{\mathrm{Conf}(D_\lambda)}\mathcal{M}_{|\overline{\Gamma}|}.$$

Proof This is well known; see, for example, [13, Lemma 3.50] or [15, Lemma 6.17]. □

Proposition 4.51 *Let B_j, S_j and $\vec{\rho}_{\clubsuit,j}$ be such that $\mathrm{ind}(B_j,S_j,\vec{\rho}_{\clubsuit,j})=2-\lambda$. For each j, the moduli space $\mathcal{M}^{B_j}(\theta(\lambda)_j^{1,2},\dots,\theta(\lambda)_j^{\lambda,1};S_j;\vec{\rho}_{\clubsuit,j};\boldsymbol{J}_j)$ (of expected dimension zero) is nonzero only if either*

- *$\lambda=2$ and $\vec{\rho}_{\clubsuit,j}$ contains a single jumping chord, or*
- *$\lambda=3$ and $\vec{\rho}_{\clubsuit,j}$ is empty.*

In these cases there is only a single domain for which the moduli space is nonempty, and in these cases it contains a unique element.

Proof We assume that $\vec{\rho}_{\clubsuit,j}$ is empty — the case where it is nonempty follows similarly, requiring only the minor modification of changing which index formulae are used. Suppose initially that $\lambda>3$.

If, for some (noncyclically) consecutive $i,i+1$, the curves $\delta_j^i,\delta_j^{i+1}$ are approximations of one another, then by Proposition 4.45 and [10, Lemma 11.8] (cf Proposition 4.61 below), there is a suitable holomorphic anchor which may be glued to any curve $u\in\mathcal{M}^{B_j}(\theta(\lambda)_j^{1,2},\dots,\theta(\lambda)_j^{\lambda,1};S_j;\vec{\rho}_{\clubsuit,j})$ of index $2-\lambda$ to yield an element u' of the moduli space

$$\mathcal{M}'=\mathcal{M}^{B_j'}(\theta(\lambda)_j^{1,2},\dots,\theta(\lambda)^{i-1,i},\theta(\lambda)^{i+1,i+2},\dots,\theta(\lambda)_j^{\lambda,1};S_j';\vec{\rho}_{\clubsuit,j})$$

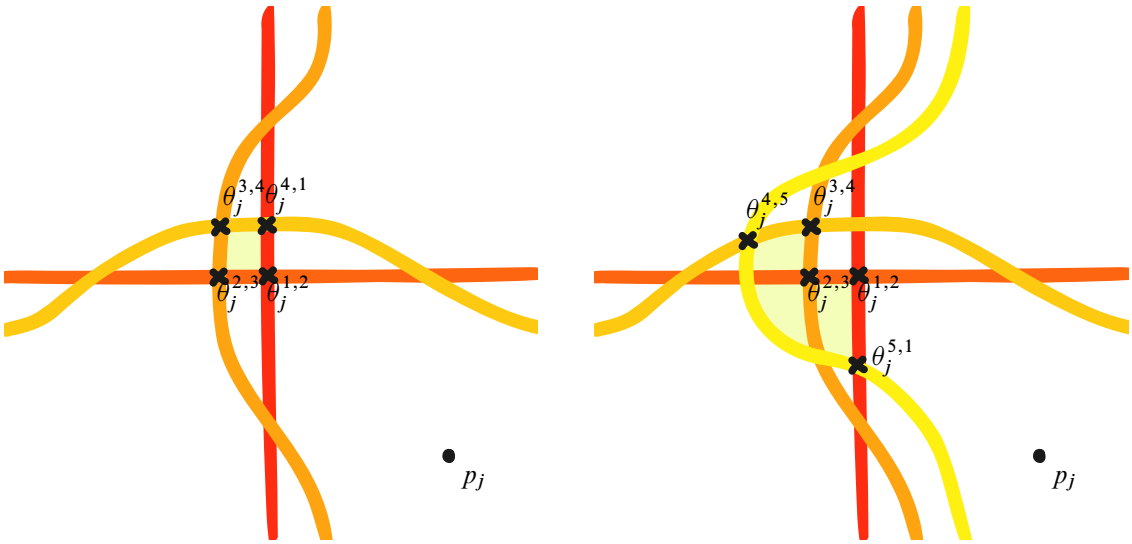


Figure 18: Ruling out $\lambda > 3$. On the left is shown the case where $\lambda = 4$, and the only possible domain connecting the intersection points shown has the incorrect orientation. On the right is shown the case where $\lambda = 5$: the domain shown has index -3; thus the corresponding moduli space is of negative expected dimension.

where the domain B_j' has Euler measure $e(B_j') = e(B_j) - 1/4$. It follows from the index formula that

$$\operatorname{ind}(B_j', S_j', \vec{\rho}_{\clubsuit,j}) = \frac{3-(\lambda-1)}{2} - \chi(S_j') + 2e(B_j') = \operatorname{ind}(B_j, S_j, \vec{\rho}_{\clubsuit,j}),$$

which implies that the moduli space $\mathcal{M}'$ has negative expected dimension. As we are working with a generic family of almost-complex structures, this implies that such a u' (and thus such a u) does not exist — thus if any (noncyclically) adjacent $\delta_j^i, \delta_j^{i+1}$ approximate one another, the corresponding moduli space is empty.

So we may assume that the family $\delta_j^1, \dots, \delta_j^\lambda$ satisfies:

(1) For every $i = 1, \dots, \lambda - 1$, δ_j^i and δ_j^{i+1} intersect transversally in a single point.

(2) For every $i = 1, \dots, \lambda - 2$, the curves $\delta_j^i, \delta_j^{i+2}$ are small approximations of one another.

(3) If λ is even, then the curves $\delta_j^\lambda, \delta_j^1$ intersect transversally in a single point; if λ is odd, then the curves $\delta_j^\lambda, \delta_j^1$ are small approximations of one another.

If λ is even, then it follows that $\theta_j \in \delta_j^l \cap \delta_j^1$ is uniquely determined and it is straightforward to see that the corresponding homology group is empty (see Figure 18 for a picture when $\lambda = 4$).

If $\lambda > 3$ is odd, then θ_j is also uniquely determined for nonempty moduli spaces; if $\theta_j = \theta_{j,+}$ then a similar gluing argument to above shows that the corresponding moduli space is empty. We can see that for general odd λ, there is a unique element of $\pi_2(\theta(\lambda)_j^{1,2}, \dots, \theta(\lambda)_j^{\lambda,1})$ — an embedded λ-gon with $\frac{1}{2}(\lambda+3)$ acute corners and $\frac{1}{2}(\lambda-3)$ obtuse corners. It follows from the index formula that the expected

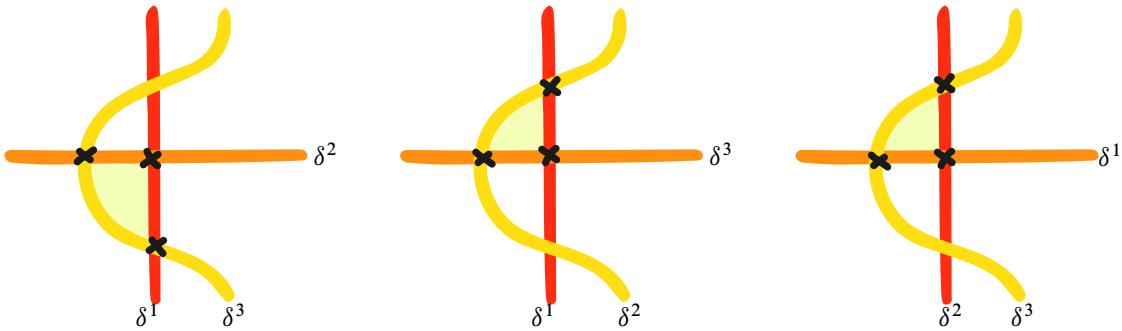


Figure 19: Three possible configurations for the curves, and the corresponding domains are shown (assuming that $\lambda = 3$) — the other three are obtained by rotating these diagrams by 90 degrees clockwise. On the left is a (ρ_1, ρ_2)-*composition piece*, and on the right are two *collision pieces*, a *precollision* and *postcollision* respectively. Rotating these gives a (ρ_2, ρ_3)-*composition piece* and two collision pieces respectively.

dimension of $\mathcal{M}^{B_j}_{\vec{e}}((\theta(\lambda)^{1,2}_j, \dots, \theta(\lambda)^{\lambda,1}_j; S_j)$ is equal to $\frac{1}{2}(3-\lambda)$. As such, this domain has index $3-\lambda$ if and only if $\lambda = 3$, as claimed in the proposition. (See Figure 18 for the case where $\lambda = 5$.)

There are six different possible configurations for the curves, with corresponding unique domain — as shown in Figure 19. □

We give things names:

Definition 4.52 We separate the possible domains near e^∞_j described in Proposition 4.51 into types, as indicated in Figure 19: (ρ_1, ρ_2)-*composition pieces*, (ρ_2, ρ_3)-*composition pieces* and *collision pieces*.

Suppose that $\lambda = 3$ and $B_E \in \pi_2(\vec{\boldsymbol{\theta}}(\lambda))$ is a domain with no Reeb chords. We say B_E is a (ρ_1, ρ_2, j)-*composition domain* (resp. (ρ_2, ρ_3, j)-*composition domain*) if the set $\overline{\Gamma}$ has two elements, j and j', and B_j is a (ρ_1, ρ_2)-composition piece (resp. (ρ_2, ρ_3)-composition piece) and $B_{j'}$ is a collision piece.

Similarly, we say B_E is a *collision domain* if $\overline{\Gamma}$ has two elements, j and j', and B_j and $B_{j'}$ are both collision pieces.

A similar definition may be made for the case $\lambda = 2$, where the corresponding domain has at most one jumping Reeb chord upon each boundary component.

It follows almost immediately that these are all of the possibilities for easterly domains:

Proposition 4.53 *Suppose B_E is such that $\mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda); S; \vec{\rho}_{\clubsuit})$ is nonempty. Then $\lambda = 3$ or 2, and either*

- *B_E is a (ρ_1, ρ_2, j)-composition domain for some j, and $\theta^{\kappa,\kappa+\lambda}_j = \theta^{\kappa,\kappa+\lambda}_{-,j}$ and $\theta^{\kappa,\kappa+\lambda}_{j+1} = \theta^{\kappa,\kappa+\lambda}_{+,j+1}$;*
- *B_E is a (ρ_2, ρ_3, j)-composition domain for some j, and $\theta^{\kappa,\kappa+\lambda}_j = \theta^{\kappa,\kappa+\lambda}_{-,j}$ and $\theta^{\kappa,\kappa+\lambda}_{j+1} = \theta^{\kappa,\kappa+\lambda}_{+,j+1}$;*
- *B_E is a collision domain, and we have that $\theta^{\kappa,\kappa+\lambda}_j = \theta^{\kappa,\kappa+\lambda}_{+,j}$ and $\theta^{\kappa,\kappa+\lambda}_{j+1} = \theta^{\kappa,\kappa+\lambda}_{+,j+1}$ for some j.*

(and similar for $\lambda = 2$). In these cases, $\mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda); S; \vec{\rho}_{\clubsuit})$ contains a unique holomorphic curve.

Proof By Proposition 4.50, the domain is uniquely determined by its restriction to handles $\mathfrak{H}_j$ where $j \in \overline{\Gamma}$, and by Proposition 4.51 we know that $\lambda = 3$ or $\lambda = 2$.

As we assumed $\sigma_{\triangle} \star \sigma_{\text{t}}$ was interleaved, the set $\overline{\Gamma}$ having more than two elements is impossible: only one curve in $\boldsymbol{\delta}^2$ is not an approximation of one in $\boldsymbol{\delta}^1$, and only one curve in $\boldsymbol{\delta}^3$ is not an approximation of one in $\boldsymbol{\delta}^2$. As such, we only need check the possibilities for B_j near each element j, j' of $\overline{\Gamma}$. There are two cases which may occur for this: either (without loss of generality), B_j is a composition piece and $B_{j'}$ is forced to be a collision piece (else too many curves from $\boldsymbol{\delta}^1$ are not approximations of curves from $\boldsymbol{\delta}^2$); or B_j is a precollision piece, forcing $B_{j'}$ to be a postcollision piece (for the same reason). The case where one piece is a composition piece splits depending upon whether δ_j^1 is meridional or longitudinal.

A similar argument applies for the case where $\lambda = 2$. □

We use this to give names to types of mixed two-storey trees of holomorphic curves:

Definition 4.54 Let

$$\mathcal{M} := \mathcal{M}^B(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\triangle}, \vec{\rho}_{\text{t}}; \sigma_{\triangle}, \sigma_{\text{t}})$$

be a one-dimensional moduli space, and $U = (u, v)$ be a height two tree of prescribed-cut polygons in a mixed end of the boundary of $\overline{\mathcal{M}}$, where

$$u \in \mathcal{M}^{B'}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k-\lambda), \boldsymbol{w}; S'; \vec{\rho}_{\triangle}, \vec{\rho}'_{\text{t}}; \sigma_{\triangle}, \sigma'_{\text{t}})$$

is spinal and $v \in \mathcal{M}^{B_E}(\vec{\boldsymbol{\theta}}(\lambda); S_E; \vec{\rho}_E)$ is easterly.

We shall say that U is

- a (κ, j)-composition end if B_E is a (ρ_1, ρ_2, j)-composition or (ρ_2, ρ_3, j)-composition domain, or
- a κ-collision end if B_E is a collision domain.

By Proposition 4.53, any mixed $U \in \partial M$ is either an (i, j)-composition end or a κ-collision end.

We conclude this section by identifying composition and collision ends with recognisable moduli spaces:

Proposition 4.55 *Suppose that*

$$\mathcal{M}' := \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}_+^{1,2}, \dots, \boldsymbol{\theta}_+^{\kappa-1,\kappa}, \boldsymbol{\theta}^{\kappa,\kappa+3}, \boldsymbol{\theta}_+^{\kappa+3,\kappa+4}, \dots, \boldsymbol{\theta}_+^{k-1,k}, \boldsymbol{y}; S'; \vec{\rho}_{\triangle}, \vec{\rho}'_{\text{t}}; \sigma_{\triangle}, \sigma'_{\text{t}}),$$
$$\mathcal{M}_E := \mathcal{M}^{B_E}(\boldsymbol{\theta}_+^{\kappa,\kappa+1}, \boldsymbol{\theta}_+^{\kappa+1,\kappa+2}, \boldsymbol{\theta}^{\kappa+2,\kappa}; S_E; \vec{\rho}_E),$$

are such that $\mathcal{M}' \times \mathcal{M}_E \subset \partial\overline{\mathcal{M}}$.

If B_E is a (ρ_1, ρ_2, j)-composition domain (resp. (ρ_2, ρ_3, j)-composition domain), then $\mathcal{M}' \times \mathcal{M}_E$ contains the same number of elements modulo two as the moduli space

$$\mathcal{M}_{\text{comp}} := \mathcal{M}^{B'}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k-1), \boldsymbol{y}; S'; \vec{\rho}_{\triangle}, \bar{\mu}_j^{\kappa}(\vec{\rho}_{\text{t}}); \sigma_{\triangle}; \sigma_{\text{t}}, (\sigma_{\triangle})_j^{\kappa}).$$

Similarly, if B_E is a collision domain, then the moduli space $\mathcal{M}' \times \mathcal{M}_E$ contains the same number of elements modulo two as

$$\mathcal{M}_{\text{coll}} := \mathcal{M}^{B'}(\boldsymbol{x}, \vec{\boldsymbol{\theta}}_+(k-1), \boldsymbol{y}; S'; \vec{\rho}_{\triangle}, \vec{\rho}_{\text{t}}; \sigma_{\triangle}, \sigma_{\text{t}}(\kappa)).$$

Proof By Proposition 4.53, the projection of each $\mathcal{M}' \times \mathcal{M}_E$ onto $\mathcal{M}'$ has degree one; hence we only have to show that, in each case, $\mathcal{M}'$ has the same number of elements as the moduli spaces in question.

By Proposition 4.39, as the diagram $\mathcal{D}_{\text{comp}}$ used for the definition of $\mathcal{M}_{\text{comp}}$ (resp. $\mathcal{D}_{\text{coll}}$ and $\mathcal{M}_{\text{coll}}$) is the regularisation of the diagram $\mathcal{D}'$ used in the definition of $\mathcal{M}'$, we know that if B_E is a composition domain, $\mathcal{M}'$ is identified with a subset of the moduli space

$$\mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}_+^{1,2}, \dots, \boldsymbol{\theta}_+^{\kappa-1,\kappa}, \boldsymbol{\theta}_+^{\kappa+3,\kappa+4}, \dots, \boldsymbol{\theta}_+^{k-1,k}, \boldsymbol{y}; S'; \vec{\rho}_{\clubsuit}, Q; (\sigma_{\clubsuit})_j^{\kappa}),$$

and it remains to be seen that this subset is precisely $\mathcal{M}_{\text{comp}}$.

Assume first that B_E is a (ρ_1, ρ_2, j)-composition domain, and suppose that for each $j' = 1, \dots, h$, the sequences

$$R_{j'} := (D^1_{j'}, B^1_{j'}, D^2_{j'}, B^2_{j'}, \dots, D^k_{j'}, B^k_{j'}, D^{k+1}_{j'})$$

form a sequence of Reeb domains for B. For every $j' \neq j$, the corresponding domain $B'_{j'}$ (as discussed in Section 3.1.1) is equal to $B_{j'}$ — it follows that in this case $R_{j'}$ is also a sequence of Reeb domains for B_j. For $j' = j$, we have that B^κ_j decomposes as the sum $C^\kappa_j + \widehat{B}^\kappa_j$, where C^κ_j is a (ρ_1, ρ_2)-composition domain. It follows that, by definition, $B^\kappa_j = B^\kappa_j(\rho_1)$, and that $\widehat{B}^\kappa_j = B^\kappa_j(\rho_1) - C^\kappa_j$ has multiplicity one in the south-west region adjacent to $e^{\infty,\kappa+1}_j$.

The domain B' has no corner at the intersection between γ^κ_j and $\gamma^{\kappa+1}_j$ — it has the same multiplicity upon either side of $\gamma^{\kappa+1}_j$. As such, $\widehat{B}^\kappa_j + B^{\kappa+1}_j$ has no corners at $e^{\infty,\kappa+1}_j$ and so $B^{\kappa+1}_j = B^{\kappa+1}_j(\rho_2)$ — and we have that $D^{\kappa+1}_j = D^{\kappa+1}_j(\varnothing)$. It follows (by comparing multiplicities away from the approximation region between γ^κ_j and $\gamma^{\kappa+2}_j$) that if $D^i_j := n_i B^i(\rho_i)$, $B^i_j := B^i(\rho'_i)$ and we denote by

$$E^i_j := \begin{cases} n_i B^i(\rho_i), & \text{for } i = 1, \dots, \kappa-1, \\ (n_\kappa + n_{\kappa+2} + 1) B^i(\rho_i), & \text{for } i = \kappa, \\ n_{i+2} B^i(\rho_{i+2}), & \text{for } i = \kappa+1, \dots, k-1, \end{cases}$$

and

$$F^i_j := \begin{cases} B^i(\rho'_i) & \text{for } i = 1, \dots, \kappa-1, \\ B^i(\rho'_{i+2}) & \text{for } i = \kappa, \dots, k-2, \end{cases}$$

then

$$(E^1_j, F^1_j, E^2_j, \dots, E^{k-2}_j, F^{k-2}_j, E^{k-1}_j)$$

is a sequence of Reeb domains for B'_j and the result follows. Similar holds for B_E a (ρ_2, ρ_3, j)-composition domain.

If B_E is a collision domain then it is completely supported within the approximation regions between $\boldsymbol{\gamma}^l$ and $\boldsymbol{\gamma}^{l+1}_k$. As such, subtracting the domain B_E has no effect upon the sequence $\vec{\rho}_{\clubsuit}(B)$, and thus $\vec{\rho}_{\clubsuit}(B'') = \vec{\rho}_{\clubsuit}(B)$. The jumping chords previously in partitions P^κ_j and $P^{\kappa+1}_{j'}$ are now determined by corners at the copies of east infinity, e^κ_j and $e^\kappa_{j'}$, and thus the corresponding partition is given by $\sigma_{\clubsuit}(\kappa)$, as required. □

4.10 Behaviour of holomorphic anchors and stretching the neck

In this section, we will try and understand the moduli spaces $\mathcal{M}^{B_\text{ⅆ}}(\boldsymbol{\theta}^{1,2}; S_\text{ⅆ})$.

Proposition 4.56 *Suppose that $(u_n) \in \mathcal{M}^{B_{\bar{e}}}_{\geq T,\bar{e},\varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; P, Q; \sigma_\text{◭}, \sigma_\text{ⅆ}; \boldsymbol{J}^t)$, is a sequence of splayed holomorphic polygons, converging to some two-storey building*

$$(u', u_\text{ⅆ}) \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{0,1}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; P', Q', \sigma'_\text{◭}, \vec{\sigma}'_\text{ⅆ}; \boldsymbol{J}) \times \mathcal{M}^{B_\text{ⅆ}}(\vec{\boldsymbol{\theta}}(m_\text{ⅆ}); S_\text{ⅆ}; P_\text{ⅆ}; \sigma).$$

Then $u_\text{ⅆ}$ is a holomorphic anchor.

Proof For notational convenience, we let $g' = g + h - 1$. We note that $\chi(S') + \chi(S_\text{ⅆ}) - g' = \chi(S)$ and $m_\text{ⅆ} + m' - 1 = m$, so $5 - m = 6 - (m_\text{ⅆ} + m')$. Denoting by $e(B_\text{ⅆ}) := e(B) - e(B')$, $\mathrm{Col}_\text{ⅆ} := \mathrm{Col}(\sigma_\text{◭}) - \mathrm{Col}(\sigma'_\text{◭})$ and $\#_\text{ⅆ} := |Q| - |Q'|$, we have that

$$\begin{aligned}
\mathrm{ind}(B, S, P, Q, \sigma_\text{◭}) &= \tfrac{1}{2}(3-m)g' - \chi(S) + 2e(B) + |P| - \mathrm{Col}(\sigma_\text{◭}) - |Q| \\
&= \tfrac{1}{2}(3-m)g' - (\chi(S') + \chi(S_\text{ⅆ}) - g') + 2e(B) + |P| - \mathrm{Col}(\sigma_\text{◭}) - |Q| \\
&= \tfrac{1}{2}(5-m)g' - \chi(S') - \chi(S_\text{ⅆ}) + 2e(B) + |P| - \mathrm{Col}(\sigma_\text{◭}) - |Q| \\
&= \tfrac{1}{2}(3-m')g' - \chi(S') + 2e(B') + |P'| - \mathrm{Col}(\sigma'_\text{◭}) - |Q'| \\
&\qquad + \tfrac{1}{2}(3-m_\text{ⅆ})g' - \chi(S_\text{ⅆ}) + 2e(B_\text{ⅆ}) + |P_\text{ⅆ}| - \mathrm{Col}_\text{ⅆ} - \#_\text{ⅆ} \\
&= \mathrm{ind}(B', S', P', Q', \sigma'_\text{◭}) + \tfrac{1}{2}(3-m_\text{ⅆ})g' - \chi(S_\text{ⅆ}) + 2e(B_\text{ⅆ}) + |P_\text{ⅆ}| - \mathrm{Col}_\text{ⅆ} - \#_\text{ⅆ}.
\end{aligned}$$

Denote by ind' the first term in this final sum, and $\mathrm{ind}_\text{ⅆ}$ the remainder. If $u_\text{ⅆ}$ were not an anchor (ie $m_\text{ⅆ} > 1$), by Proposition 4.37 we can unperturb the Lagrangian $C^i(\varepsilon)$ to give a corresponding bordered holomorphic polygon $u'_\text{ⅆ}$ which lives in a moduli space of curves of index $\mathrm{ind}_\text{ⅆ}$.

It follows that $2 - m - \mathrm{ind}' = \mathrm{ind}_\text{ⅆ}$. Noting that $\mathrm{ind}' \geq 2 - m'$, it follows that

$$\mathrm{ind}_\text{ⅆ} \leq 2 - m - (2 - m') = m' - m = 1 - m_\text{ⅆ},$$

and so $u'_\text{ⅆ}$ lives in a moduli space of negative expected dimension, so does not exist — and thus nor does $u_\text{ⅆ}$. Therefore, $u_\text{ⅆ}$ is a holomorphic anchor as required. □

Proposition 4.57 *The domain $B_\text{ⅆ}$ is supported within the approximation region for the curves, and B' splays B.*

Proof The first half of the statement follows along similar lines to Proposition 4.48. If we fix points away from the approximation region between $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ in the diagram corresponding with the moduli space for u', then each curve u_n has the same multiplicity at these points and thus so does u' — so B' has the same multiplicity as B everywhere away from this approximation region, which is precisely the definition of B' splaying B. □

We need a lemma about the acute corners of anchored curves:

Lemma 4.58 *Suppose that*

$$(u', u_{\mathfrak{t}}) \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{0,1}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; P', Q', \sigma'_{\clubsuit}, \vec{\sigma}'_{\mathfrak{t}}; \boldsymbol{J}) \times \mathcal{M}_{\mathfrak{t}}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}}, \sigma)$$

is an anchored holomorphic curve in the boundary of the moduli space

$$\boldsymbol{\mathcal{M}}^{B_{\bar{e}}}_{\geq T, \bar{e}, \varepsilon}(\vec{\boldsymbol{x}}; S_{\bar{e}}; P, Q; \sigma_{\clubsuit}, \sigma_{\mathfrak{t}}).$$

If A denotes the number of acute corners of B' at the points $e^{\infty,1}$, then the difference between the number of acute corners of B and B' satisfies

$$\mathrm{ac}(B') - \mathrm{ac}(B) \geq A - 2(|P| - |P'|),$$

with equality holding only if every component of $S_{\mathfrak{t}}$ has at most one east puncture.

Proof The acute corners of B and B' consist of two types: *east acute* corners and *main acute* corners. Here, a main acute corner is an acute corner corresponding with a noneast puncture of S (resp. S'), and east acute corners correspond with intersections of arcs in $\boldsymbol{\gamma}^1$ with $\partial\Sigma$. The main corners of B' split into two sets: those A at $e^{\infty,1}$ and all others. As B' approximates B, it follows that the number of main acute corners of B' not at $e^{\infty,1}$ is precisely the number of main acute corners of B. As such,

$$\mathrm{ac}(B') - \mathrm{ac}(B) = A + A'_E - A_E,$$

where A'_E and A_E are the number of easterly acute corners of B' and B respectively.

The quantities A_E and A'_E may be read from the multiplicity of B and B' in a neighbourhood of each e^{∞}_j. Indeed, each puncture in $E_j(S)$ (resp. S') contributes two acute corners (at the start and end of the corresponding Reeb chord) unless there is some other puncture labelled with the subsequent chord. If we choose a small neighbourhood N_j of east infinity and three points p_1, p_2, p_3 in the corresponding regions of $N_j - \boldsymbol{\gamma}^1_j$, this translates as

$$\begin{aligned} A_E &= n_{p_1}(B) + |n_{p_1}(B) - n_{p_2}(B)| + |n_{p_2}(B) - n_{p_3}(B)| + n_{p_3}(B), \\ A'_E &= n_{p_1}(B') + |n_{p_1}(B') - n_{p_2}(B')| + |n_{p_2}(B') - n_{p_3}(B')| + n_{p_3}(B'). \end{aligned}$$

We extend this to $B_{\mathfrak{t}}$, setting

$$A_{\mathfrak{t},E} := n_{p_1}(B_{\mathfrak{t}}) + |n_{p_1}(B_{\mathfrak{t}}) - n_{p_2}(B_{\mathfrak{t}})| + |n_{p_2}(B_{\mathfrak{t}}) - n_{p_3}(B_{\mathfrak{t}})| + n_{p_3}(B_{\mathfrak{t}}).$$

This final quantity is less than or equal to $2 \cdot |E_j(S_{\mathfrak{t}})|$, with equality holding only if $|E_j(S_{\mathfrak{t}})| = 1$. By the triangle inequality, $A'_E + A_{\mathfrak{t},E} \geq A_E$ and hence $A_E - A'_E \leq A_{\mathfrak{t},E} \leq 2 \cdot |E_j(S_{\mathfrak{t}})|$, with equality holding only if $|E_j(S_{\mathfrak{t}})| = 1$.

Summing over all j, it follows that

$$\mathrm{ac}(B') - \mathrm{ac}(B) = A + A'_E - A_E \geq A - 2(|P| - |P'|)$$

as required; with equality only if $|E_j(S_{\mathfrak{t}})| = 1$ for each j. □

Proposition 4.59 *Suppose that $u_n \in \mathcal{M}^B_{\geq T}(\boldsymbol{x}; S; P, Q; \sigma_{\text{🙂}}, \vec{\sigma}_{\text{☹}})$ is a sequence of splayed holomorphic polygons, converging to some anchored holomorphic curve*

$$(u', u_{\text{☹}}) \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{0,1}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; P', Q', \sigma'_{\text{🙂}}, \sigma'_{\text{☹}}; \boldsymbol{J}) \times \mathcal{M}^{B_{\text{☹}}}_{\text{☹}}(\boldsymbol{\theta}^{1,2}; S_{\text{☹}}; P_{\text{☹}}, \sigma).$$

Then $S_{\text{☹}}$ a disjoint union of topological discs, where $|E_j(S_{\text{☹}})| \leq 1$ for every j, and either $A = 1$ and Q' consists of the punctures of S' induced by Q, or $A = 0$ and Q' consists of the punctures of S' induced by Q, together with a single puncture labelled by $v^{1,2}$.

Proof Let $g' = g + h - 1$, $e(B_{\text{☹}}) := e(B) - e(B')$, $\mathrm{Col}_{\text{☹}} := \mathrm{Col}(\sigma_{\text{🙂}}) - \mathrm{Col}(\sigma'_{\text{🙂}})$ and $\#_{\text{☹}} := |Q| - |Q'|$ as in the proof of Proposition 4.56, and observe that the quantities $\mathrm{Col}_{\text{☹}}$ and $\#_{\text{☹}}$ are negative. Note also that $\chi(S_{\text{☹}}) \leq g'$, with equality if and only if $S_{\text{☹}}$ is a disjoint union of topological disks.

The argument in Proposition 4.56 tells us that $\mathrm{ind}(B, S, P, Q, \sigma_{\text{🙂}}) - \mathrm{ind}'$ is equal to

$$\tfrac{1}{2}(3 - m_{\text{☹}})g' - \chi(S_{\text{☹}}) + 2e(B_{\text{☹}}) + |P_{\text{☹}}| - \mathrm{Col}_{\text{☹}} - \#_{\text{☹}}\,. \tag{4.60}$$

As $u_{\text{☹}}$ is an anchor, it follows that $\mathrm{ind}' \geq 2 - (m+1) = 1 - m$. By assumption, $\mathrm{ind}(B, S, P, Q, \sigma_{\text{🙂}}) = 2 - m$ and so the formula (4.60) is less than or equal to 1.

We examine the quantity $2(e(B) - e(B')) = 2e(B_{\text{☹}})$. It is clear that $\chi(B) = \chi(B')$, and thus

$$2e(B_{\text{☹}}) = \tfrac{1}{2}(\mathrm{ob}(B) - \mathrm{ob}(B') - \mathrm{ac}(B) + \mathrm{ac}(B')).$$

If we let A denote the number of acute corners at the vertices $e_1^{\infty,1}, \dots, e_h^{\infty,1}$, then as no corners at east infinity are obtuse we can see that $\mathrm{ob}(B) - \mathrm{ob}(B') = (1 - g + (A - h)) = -g' + A$. By the previous lemma, we know that $\mathrm{ac}(B') - \mathrm{ac}(B) \geq A - 2(|P| - |P'|)$ and as such $2(e(B) - e(B')) \geq -g' + A - |P_{\text{☹}}|$. Hence

$$1 \geq \tfrac{1}{2}(3 - m_{\text{☹}})g' - \chi(S_{\text{☹}}) + 2e(B_{\text{☹}}) + |P_{\text{☹}}| - \mathrm{Col}_{\text{☹}} - \#_{\text{☹}} \geq 0$$

and at most one of the terms A, $g' - \chi(S_{\text{☹}})$, $\mathrm{Col}_{\text{☹}}$ or $\#_{\text{☹}}$ is positive (and equal to one).

Assume first that $\chi(S_{\text{☹}}) = g'$, so that $S_{\text{☹}}$ is a disjoint union of topological disks. There are a series of cases for this:

- If $A = 1$, then $\mathrm{Col}_{\text{☹}} = \#_{\text{☹}} = 0$ (so $|Q| = |Q'|$) and each component of $S_{\text{☹}}$ has at most one east puncture.
- If $A = 0$, then either:
  - $\#_{\text{☹}} = 1$, so $|Q| = |Q'| - 1$, $\mathrm{Col}_{\text{☹}} = 0$ and each component of $S_{\text{☹}}$ has at most one east puncture.
  - $\mathrm{Col}_{\text{☹}} = 1$, $\#_{\text{☹}} = 0$ and each component of $S_{\text{☹}}$ has at most one east puncture: this does not occur as gluing $(u', u_{\text{☹}})$ gives a nearby curve in some $\mathcal{M}^B(\vec{\boldsymbol{x}}; S; P, Q; \sigma'_{\text{🙂}}, \vec{\sigma}_{\text{☹}}; \boldsymbol{J}^t)$, where $\mathrm{Col}(\sigma'_{\text{🙂}}) > 0$ — such moduli spaces have negative expected dimension.
  - Or, finally, $A_E - A_{E'} = 2(|P| - |P|) - 1$: this does not occur as the left hand side of this equation is even.

If $\chi(S_{\text{☹}}) = g' - 1$, then one component of $S_{\text{☹}}$ is a topological annulus without an "∞" puncture on one boundary component. Any nearby curve u in $\mathcal{M}^B(\boldsymbol{x}; S; P, Q; \sigma'_{\text{🙂}}, \vec{\sigma}_{\text{☹}}; \boldsymbol{J}^t)$ has a boundary component

with no noneast puncture upon it, so that $\pi_D \circ u$ is constant on this component or violates the maximum modulus principle — hence this case does not occur (compare [14, Proposition 5.43]). □

Proposition 4.56 tells us that if $(u', u_{\mathfrak{t}}) \in \partial \overline{\mathcal{M}}^{B}_{\geq T}(\boldsymbol{x}; S; \vec{\rho}_{\mathfrak{a}}, \vec{\rho}_{\mathfrak{t}}; \vec{\sigma}_{\mathfrak{a}}, \vec{\sigma}_{\mathfrak{t}}, \boldsymbol{J}^t)$, each component of the source $S_{\mathfrak{t}}$ has at most one puncture which is mapped to e_j^{∞} for each j. By definition, each of these punctures is labelled by some element of the torus algebra ρ_j, and we shall write $\mathcal{M}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \sigma)$ to refer to the moduli space of curves in $\mathcal{M}^{B_{\mathfrak{t}}}_{\mathfrak{t}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}}, \sigma)$ where $P_{\mathfrak{t}}$ is the discrete partition and $\vec{\rho} = \{\rho_j \mid \rho_j \in E_j(S_{\mathfrak{t}})\}$.

In a similar manner to Section 4.9, we can define

$$\overline{\Gamma} = \{j \in 1, \dots, h \mid E_j(S_{\mathfrak{t}}) \neq \varnothing\}$$

and show that the moduli space $\mathcal{M}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \sigma)$ splits as a product

$$\mathcal{M}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \sigma) = \mathcal{M}_{\mathfrak{t}} \times \mathcal{M}_{\mathfrak{t},1} \times \cdots \times \mathcal{M}_{\mathfrak{t},|\overline{\Gamma}|}.$$

Here $\mathcal{M}_{\mathfrak{t}}$ is a moduli space of one-gons in a closed Heegaard multidiagram $\mathcal{D}'_{\mathfrak{t}}$, and each $\mathcal{M}_{\mathfrak{t},j}$ is a moduli space of one-gons defined using a bordered multidiagram $\mathcal{D}_{\mathfrak{t},j} = (\Sigma_{\mathfrak{t},j}; \gamma_j^1; b_j)$, where $\Sigma_{\mathfrak{t},j}$ is a torus.

Proposition 4.61 *There is a unique homology class $B'_{\mathfrak{t}}$ which carries a holomorphic representative in the moduli space $\mathcal{M}_{\mathfrak{t}}$. Moreover, $B'_{\mathfrak{t}} \in \pi_2(\boldsymbol{\theta}^{1,2}_{\mathrm{res},+})$, where $\boldsymbol{\theta}^{1,2}_{\mathrm{res},+}$ is the restriction of the distinguished generator $\boldsymbol{\theta}^{1,2}_{+}$ to the diagram $\mathcal{D}'_{\mathfrak{t}}$. As such, the moduli space $\mathcal{M}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \sigma)$ is nonempty only if $B_{\mathfrak{t}} = B'_{\mathfrak{t}} + \sum_{j \in \overline{\Gamma}} B_{\mathfrak{t},j}$, where each $B_{\mathfrak{t},j}$ is supported within the handle $\mathfrak{H}_j$, and $\boldsymbol{\theta}^{1,2} = \boldsymbol{\theta}^{1,2}_{+}$, in which case it has the same number of elements as $\sum_{j \in \overline{\Gamma}} \mathcal{M}_{\mathfrak{t},j}$.*

Proof This is considered in [10, Lemma 11.8], though as stated there the proof is incorrect; see the corrections. □

As such, we need only to understand the moduli spaces $\mathcal{M}_{\mathfrak{t},j}$.

Proposition 4.62 *Let $S_{\mathfrak{t},j}$ be an anchor source with underlying surface a disc, and with unique east puncture q. For the evaluation map $\mathrm{ev}_q \colon \mathcal{M}^{B_{\mathfrak{t},j}}(\theta_j^{1,2}; S_{\mathfrak{t},j}; \rho_j) \to \mathbb{R}$, the moduli space $\mathcal{M}_{\mathfrak{t},j}(0) = \mathrm{ev}^{-1}(0)$ contains a unique element for every ρ_j.*

Proof This follows by a model computation for each choice of ρ_j, and Proposition 4.45. For instance, consider the bordered Heegaard diagram and Heegaard multidiagrams as shown in Figure 20.

The moduli space

$$\mathcal{M}^{B}(x, y; S; \rho_3, \varnothing; \{\rho_3\}, \varnothing)$$

clearly contains a unique element, as does the moduli space

$$\mathcal{M}^{B'}(x, \theta^{1,2}_{+}, y; S'; \varnothing, \rho_3; \varnothing, \{\rho_3\}).$$

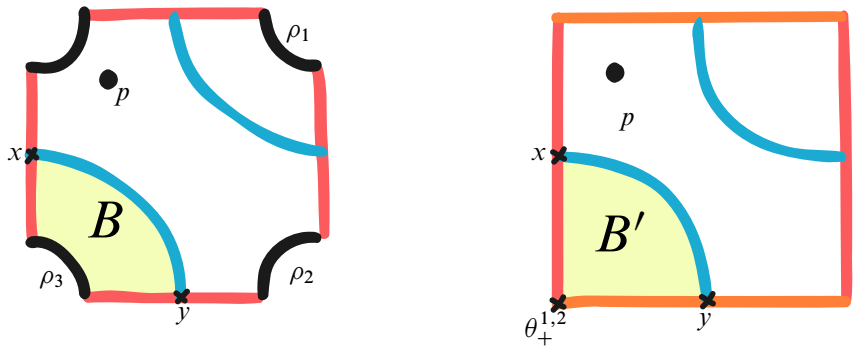


Figure 20: A model computation for Proposition 4.62.

Therefore, by Proposition 4.45,

$$|\mathcal{M}^{B}(x, y; S; \rho_3, \varnothing; \{\rho_3\}, \varnothing)| = |\mathcal{M}^{B'}(x, \theta_+^{1,2}, y; S'; \varnothing, \rho_3; \varnothing, \{\rho_3\})| \cdot |\mathcal{M}_{\mathfrak{t},j}(0)|,$$

hence $|\mathcal{M}_{\mathfrak{t},j}(0)| = 1$, as required. □

This lets us rule out anchored curves with holomorphic anchors that have more than one easterly puncture:

Proposition 4.63 *Suppose that*

$$(u', u_{\mathfrak{t}}) \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; P', Q', \sigma'_{\triangleq}, \vec{\sigma}'_{\mathfrak{t}}; \boldsymbol{J}) \times \mathcal{M}_{\mathfrak{t}}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \vec{\sigma})$$

is an anchored holomorphic curve in the boundary of the moduli space

$$\overline{\mathcal{M}}_{\geq T}^{B}(\boldsymbol{x}; S; P, Q; \sigma_{\triangleq}, \vec{\sigma}_{\mathfrak{t}}).$$

Then $\vec{\sigma}'_{\mathfrak{t}}$ is interleaved, and hence $S_{\mathfrak{t}}$ consists of a disjoint union of topological disks, only one of which has an east puncture.

Proof We already know that each component of $S_{\mathfrak{t}}$ is a disk, with at most one east puncture. Suppose that $\vec{\sigma}'_{\mathfrak{t}}$ were not interleaved. By Proposition 4.62, there is a (unique) element $v_{\mathfrak{t}}$ of $\mathcal{M}_{\mathfrak{t}}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; \vec{\rho}, \vec{\sigma})$ where the height of each east puncture is 0. As such, by Proposition 4.45, we could glue u' to $v_{\mathfrak{t}}$ to get an element of a moduli space $\mathcal{M}_{\geq T}^{B}(\boldsymbol{x}; S; \vec{\rho}_{\triangleq}, \vec{\rho}_{\mathfrak{t}}; \sigma_{\triangleq}; \boldsymbol{J}^t)$ with $\mathrm{Col}(\sigma_{\triangleq}) > 0$; but this moduli space has negative expected dimension. □

The final result we need is to understand the implied asymptotics of curves in the boundary:

Proposition 4.64 *Suppose that*

$$(u', u_{\mathfrak{t}}) \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; P', Q', \sigma'_{\triangleq}, \vec{\sigma}'_{\mathfrak{t}}; \boldsymbol{J}) \times \mathcal{M}_{\mathfrak{t}}^{B_{\mathfrak{t}}}(\boldsymbol{\theta}^{1,2}; S_{\mathfrak{t}}; P_{\mathfrak{t}}, \sigma)$$

is an anchored holomorphic curve in the boundary of the moduli space

$$\boldsymbol{\mathcal{M}}_{\geq T}^{B}(\vec{\boldsymbol{x}}; S; \vec{\rho}_{\triangleq}, \vec{\rho}_{\mathfrak{t}}; \vec{\sigma}_{\triangleq}, \vec{\sigma}_{\mathfrak{t}}).$$

Then $u' \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}^{1,2}, \dots, \boldsymbol{\theta}^{m-1,0}, \boldsymbol{y}; S'; \vec{\rho}'_{\triangleq}, \vec{\rho}'_{\mathfrak{t}}, \vec{\sigma}'_{\triangleq}, \vec{\sigma}'_{\mathfrak{t}}; \boldsymbol{J})$, where the pair $(\vec{\rho}'_{\triangleq}, \sigma'_{\triangleq})$ and $(\vec{\rho}'_{\mathfrak{t}}, \sigma'_{\mathfrak{t}})$ splay $(\vec{\rho}_{\triangleq}, \sigma_{\triangleq})$ and $(\vec{\rho}_{\mathfrak{t}}, \sigma_{\mathfrak{t}})$.

Proof Consider a small disc neighbourhood M_j^1 of $e_j^{\infty,1}$ which satisfies that each component of $\pi_\Sigma \circ u'^{-1}(M_j^1)$ contains a single point of the set $\Xi := \pi_\Sigma \circ u'^{-1}(e_j^{\infty,1})$ (we include the puncture q which is mapped to $e_j^{\infty,1}$ in this set) — label the components by $S(\xi)$ so that $\xi \in \Xi$ is contained in $S(\xi)$. The set Ξ splits into three sets: q, the set Ξ_L of points mapped to e^2 by $\pi_D \circ u'$, and the set Ξ_R of points mapped to e^1 by $\pi_D \circ u'$. There are corresponding restricted sources $S_L := \bigcup_{\xi\in\Xi_L} S(\xi)$ and $S_R := \bigcup_{\xi\in\Xi_R} S(\xi)$.

The homology class of $\pi_\Sigma \circ u'|_{S_L}$ is modelled upon $n \cdot B(\rho)$ for some nonjumping ρ (which is equal to ρ_{12} if γ_j^2 is meridional, and ρ_{23} if γ_j^2 is longitudinal), and the homology class of $\pi_\Sigma \circ u'|_{S(q)}$ is modelled upon $B(\rho')$ for some jumping ρ'. It follows (by considering multiplicities away from the approximation region between $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$) that if

$$(D_j^1, B_j^1, D_j^2, B_j^2, \dots, D_j^k, B_j^k, D_j^{k+1})$$

were a sequence of Reeb domains for B_j, the sequence obtained by replacing each $D_j^i = n_i B^i(\rho_i)$ with $n_i B^{i+1}(\rho_i)$ and each $B_j^i = B^i(\rho_i')$ with $B^{i+1}(\rho_i')$, adjoining $B_j^0 = B_j^1(\rho')$ and replacing $D_j^1 = B^i(\rho_1)$ by $n \cdot B^2(\rho) + n_1 B^2(\rho_1)$ is a sequence of Reeb domains for $B_j' - B(\vec{\rho}'_{\clubsuit,j})$.

It remains to show that the sequence

$$\vec{\rho} = (\rho', \underbrace{\rho, \dots, \rho}_{n \text{ times}})$$

satisfies that $\vec{\rho}'_{\clubsuit,j} \star \vec{\rho} = \vec{\rho}_{\clubsuit,j}$. To see this, we note that the domain $B(\vec{\rho}_{\maltese})$ must agree with $B^1(\vec{\rho}'_{\maltese}) + B^1(\vec{\rho})$ away from the approximation region. As such, $B(\vec{\rho}) = B(\vec{\rho}_{\maltese}) - B(\vec{\rho}'_{\maltese})$. This latter quantity, however, may be written in a more useful way.

Let $\mathcal{Q} = \pi_\Sigma \circ u'^{-1}(e_j^\infty)$, and let $\mathcal{Q}'$ denote the subset of $\mathcal{Q}$ given by $\{q \in \mathcal{Q} \mid \mathrm{ev}_j(q) < 0\}$. If we fix a small neighbourhood M_j^∞ of e_j^∞ such that each component of $\pi_\Sigma \circ u'^{-1}(M_j^\infty)$ contains a unique point in $\mathcal{Q}$, the domains $B(\vec{\rho}_{\maltese})$ and $B(\vec{\rho}'_{\maltese})$ can be written as a sum over points in $\mathcal{Q}$ or $\mathcal{Q}'$, respectively, of the quantities $B(\rho(q))$, where $\rho(q)$ is such that $(\pi_\Sigma \circ u')(S(q))$ is modelled upon $B(\rho(q)) \subset M_j^\infty$ — and $S(q)$ denotes the component of $\pi_\Sigma \circ u'^{-1}(M_j^\infty)$ containing q. It follows by considering the multiplicity of u' at basepoints in each of the four quadrants of M_j^∞ that $B(\vec{\rho}_{\maltese}) - B(\vec{\rho}'_{\maltese})$ coincides with the sum over points q in $\mathcal{Q} - \mathcal{Q}'$ of $B(q)$, and thus is equal to $B(\underline{\vec{\rho}})$, where $\underline{\vec{\rho}}$ satisfies $\vec{\rho}'_{\clubsuit,j} \star \underline{\vec{\rho}} = \vec{\rho}_{\clubsuit,j}$. As such, $B(\vec{\rho}) = B(\underline{\vec{\rho}})$ and so $\vec{\rho}$ must lie in the same 123-equivalence class as $\underline{\vec{\rho}}$. The cardinality of the sets $\mathcal{Q}$ and $\mathcal{Q}' \cup \Xi_L \cup \{q\}$ agree, so in fact $\vec{\rho} = \underline{\vec{\rho}}$. □

We finish this section by utilising the results above to understand properly the behaviour of moduli spaces under the neck-stretching operation.

Proposition 4.65 *Let B, S, $\vec{\rho}_{\clubsuit}$, $\vec{\rho}_{\maltese}$ and $\vec{\sigma}_{\clubsuit}$ be such that*

$$\mathcal{M}_{\geq T} := \mathcal{M}^{B_{\bar{e}}}_{\geq T,\bar{e},\varepsilon}(\vec{x}; S_{\bar{e}}; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \vec{\sigma}_{\clubsuit}, \vec{\sigma}_{\maltese})$$

is one-dimensional, and $\vec{\sigma}_{\clubsuit}$ is interleaved.

Say, as shorthand, that $(B';\vec{\rho}'_{\blacktriangle},\sigma'_{\blacktriangle};\vec{\rho}'_{\blacktriangleleft},\sigma'_{\blacktriangleleft})$ *splays* $(B;\vec{\rho}_{\blacktriangle},\sigma_{\blacktriangle};\vec{\rho}_{\blacktriangleleft},\sigma_{\blacktriangleleft})$ *if* B' *splays* B, *and* $(\vec{\rho}'_{\blacktriangle},\sigma'_{\blacktriangle})$ *and* $(\vec{\rho}'_{\blacktriangleleft},\sigma'_{\blacktriangleleft})$ *splay* $(\vec{\rho}_{\blacktriangle},\sigma_{\blacktriangle})$ *and* $(\vec{\rho}_{\blacktriangleleft},\sigma_{\blacktriangleleft})$.

Denote by

$$\mathcal{M}_{\text{splayed}} := \bigcup_{\substack{(B';\vec{\rho}'_{\blacktriangle},\sigma'_{\blacktriangle};\vec{\rho}'_{\blacktriangleleft},\sigma'_{\blacktriangleleft}) \\ \text{splays} \\ (B;\vec{\rho}_{\blacktriangle},\sigma_{\blacktriangle};\vec{\rho}_{\blacktriangleleft},\sigma_{\blacktriangleleft})}} \mathcal{M}^{B'}(\boldsymbol{x},\boldsymbol{\theta}_+(k-1),\boldsymbol{y};S';\vec{\rho}'_{\blacktriangle},\vec{\rho}'_{\blacktriangleleft};\sigma'_{\blacktriangleleft},\sigma'_{\blacktriangle};\boldsymbol{J}).$$

Then $\overline{\boldsymbol{\mathcal{M}}}_{\geq T}$ *is a compact one-manifold with boundary given by the disjoint union of*

$$\mathcal{M}^{B_{\bar{e}}}_{\bar{e},\varepsilon}(\boldsymbol{x},\boldsymbol{\theta}_+(k-2),\boldsymbol{y};S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\blacktriangleleft};\vec{\sigma}_{\blacktriangle},\vec{\sigma}_{\blacktriangleleft})$$

and $\mathcal{M}_{\text{splayed}} \times \mathcal{M}^{B_{\blacktriangleleft}}(\boldsymbol{\theta}^{1,2},S_{\blacktriangleleft},\rho)$, *where* ρ *is the unique jumping chord in the final part of* $\vec{\sigma}_{\blacktriangle}$.

Therefore, the moduli spaces

$$\mathcal{M}^{B}(\boldsymbol{x},\boldsymbol{\theta}_+(k-2),\boldsymbol{y};S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\blacktriangleleft};\vec{\sigma}_{\blacktriangle},\vec{\sigma}_{\blacktriangleleft};\boldsymbol{J}^T)$$

and $\mathcal{M}_{\text{splayed}}$ *contain the same number of elements modulo two.*

Proof By Proposition 4.64, we know that any curve in $\partial\overline{\boldsymbol{\mathcal{M}}}_{\geq T}$ lives in the moduli space

$$\mathcal{M}_{\text{splayed}} \times \mathcal{M}^{B_{\blacktriangleleft}}(\boldsymbol{\theta}^{1,2},S_{\blacktriangleleft},\rho).$$

By Proposition 4.56, we know that $S_{\blacktriangleleft}$ is a disjoint union of topological disks, and thus by Proposition 4.45, $\partial\overline{\boldsymbol{\mathcal{M}}}_{\geq T}$ is a compact one-manifold with boundary given.

The second part follows from the identification of

$$\mathcal{M}^{B_{\bar{e}}}_{\bar{e},\varepsilon}(\boldsymbol{x},\boldsymbol{\theta}_+(k-2),\boldsymbol{y};S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\blacktriangleleft};\vec{\sigma}_{\blacktriangle},\vec{\sigma}_{\blacktriangleleft})$$

with

$$\mathcal{M}^{B}(\boldsymbol{x},\boldsymbol{\theta}_+(k-2),\boldsymbol{y};S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\blacktriangleleft};\vec{\sigma}_{\blacktriangle},\vec{\sigma}_{\blacktriangleleft})$$

given in Propositions 4.6.1 and 4.7.2, and the fact that $\mathcal{M}^{B_{\blacktriangleleft}}(\boldsymbol{\theta}^{1,2},S_{\blacktriangleleft};\rho)$ has one element modulo two by Proposition 4.62. □

5 Statement and proof of theorems

After all this formality, we are at long last ready to state some theorems.

Let $\mathcal{D}$ be a bordered diagram with h boundary components. We shall define a right $\mathcal{A}_\infty$ multimodule over the torus algebra, which we call $\text{Poly}(\mathcal{D})$. We first define a series of auxiliary modules $\text{Poly}_k(\mathcal{D})$.

Consider the module generated over $\mathbb{F}_2$ by the set of generators $\mathcal{G}(\mathcal{D})$. This has an obvious right action by the ring of idempotents of $\mathcal{T}^h$, defined by setting

$$\boldsymbol{x}\cdot\iota = \begin{cases} \boldsymbol{x} & \text{if } \iota(\boldsymbol{x}) = \iota, \\ 0 & \text{otherwise.} \end{cases}$$

We define an $\mathcal{A}_\infty$ module structure upon $\mathrm{Poly}_k(\mathcal{D})$ by first defining maps

$$n^k_{i_1,\dots,i_h}\colon \mathrm{Poly}_k(\mathcal{D})\otimes_{\mathcal{I}}\mathcal{T}_1^{\otimes i_1}\otimes_{\mathcal{I}}\cdots\otimes_{\mathcal{I}}\mathcal{T}_h^{\otimes i_h}\times\vec{\sigma}(i_1,\dots,i_h)\to\mathrm{Poly}_k(\mathcal{D}),$$

where $\vec{\sigma}(i_1,\dots,i_h)$ denotes the set of all splicings of all sets of sequences of Reeb chords in

$$\mathcal{T}_1^{\otimes i_1}\otimes\cdots\otimes\mathcal{T}_h^{\otimes i_h}.$$

Suppose that $\vec{\rho}=\vec{\rho}_1,\dots,\vec{\rho}_h$ is a set of sequences of Reeb multichords in $\mathcal{T}^h$, where $|\vec{\rho}_j|=i_j$. For every interleaved splicing σ of $\vec{\rho}$, we define the coefficient

$$a^k(\boldsymbol{x},\boldsymbol{y},\vec{\rho},\vec{\sigma}):=\sum_{\substack{(\vec{\rho}_{\blacktriangle},\vec{\sigma}_{\blacktriangle}),(\vec{\rho}_{\bigstar},\vec{\sigma}_{\bigstar})\\ \text{is a } k\text{–shipping}\\ \text{of } (\vec{\rho},\vec{\sigma})}}\ \sum_{\substack{B\in\pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x},\boldsymbol{y})\\ \chi(S)=\chi_{\mathrm{emb}}(S)\\ \mathrm{ind}(B,S,\vec{\rho}_{\blacktriangle},\vec{\rho}_{\bigstar},\vec{\sigma}_{\blacktriangle})=2-|\vec{\sigma}_{\bigstar}|}}\#\mathcal{M}^B(\boldsymbol{x},\boldsymbol{\theta}_+(|\vec{\sigma}_{\bigstar}|-2),\boldsymbol{y}';S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\bigstar};\vec{\sigma}_{\blacktriangle},\vec{\sigma}_{\bigstar})$$

where the nearest-point map takes $\boldsymbol{y}'$ to $\boldsymbol{y}$, and hence put

$$n^k_{i_1,\dots,i_h}(\boldsymbol{x},\vec{\rho},\vec{\sigma}):=\sum_{\boldsymbol{y}\in\mathcal{G}(\mathcal{D})}a^k(\boldsymbol{x},\boldsymbol{y},\vec{\rho},\vec{\sigma})\cdot\boldsymbol{y}.$$

We define

$$m^k_{i_1,\dots,i_h}\colon \mathrm{Poly}_k(\mathcal{D})\otimes_{\mathcal{I}}\mathcal{A}_1^{\otimes i_1}\otimes\cdots\otimes\mathcal{A}_h^{\otimes i_h}\to\mathrm{Poly}_k(\mathcal{D})$$

by

$$m^k_{i_1,\dots,i_h}(x,\vec{\rho}):=\sum_{\sigma\text{ an interleaving for }\vec{\rho}}n^k_{i_1,\dots,i_h}(x,\vec{\rho},\vec{\sigma}).$$

Proposition 5.1 *For sufficiently generic choice of $\boldsymbol{J}$, the definition of $\mathrm{Poly}_0(\mathcal{D})$ above agrees with the definition of the multimodule $\mathrm{BSA}(\mathcal{D})$ described in* [26, Section 8.3].

Proof In the case where $k=0$, the above moduli spaces are formed of bigons in the bordered diagram $\mathcal{D}$, with constraints upon evaluation maps given by the interleaving σ. To see that these are precisely the bigons counted in the definition of $\mathrm{BSA}(\mathcal{D})$, we note that for any $(B,S,\vec{\rho}_{\bigstar})$ such that the moduli space $\mathcal{M}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\bigstar})$ is zero-dimensional (as counted in the definition of the BSA multiplication maps), the subspace of $\mathcal{M}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\bigstar})$ where the heights of any punctures q,q' labelled by a jumping chord coincide is given by $\mathcal{M}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\bigstar};\sigma)$ for σ some noninterleaved splicing of $\vec{\rho}_{\bigstar}$. By the formula at the end of Section 4.4.1 (or Proposition 4.22), $\mathrm{ind}(B,S,\vec{\rho}_{\bigstar},\sigma)<1$ and so the corresponding moduli space is empty; ie the multiplication maps in $\mathrm{BSA}(\mathcal{D})$ do not count curves where heights coincide. (This should be compared with [12, Section 2.4.3].)

As such, the moduli space $\mathcal{M}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\bigstar})$ splits as a union over all interleaved splicings of moduli spaces $\mathcal{M}(\boldsymbol{x},\boldsymbol{y};S;\vec{\rho}_{\bigstar};\vec{\sigma})$, which is precisely what the maps $m^0_{i_1,\dots,i_h}$ count. □

Proposition 5.2 (compare [14, Lemma 7.7; 26, Theorem 7.8]) *Suppose that the diagram $\mathcal{D}$ is provincially admissible. For any fixed generators $\boldsymbol{x}$ and $\boldsymbol{y}$, and a set of nonempty sets of Reeb chords, there are at most finitely many $B\in\pi_2^{\boldsymbol{\theta}+}(\boldsymbol{x},\boldsymbol{y})$ such that $\mathcal{M}^B(\boldsymbol{x},\boldsymbol{\theta}_+(k-2),\boldsymbol{y}';S;\vec{\rho}_{\blacktriangle},\vec{\rho}_{\bigstar})$ are nonempty — ie the sums in the definition of the maps m_n are finite.*

Moreover, if $\mathcal{D}$ is admissible, then there is a constant κ depending only upon $\mathcal{D}$ for which every map m_i^k with $i > \kappa$ is zero, independent of k.

Proof We proceed by induction. In the case $k = 0$, by Proposition 5.1 the proposition coincides with [26, Theorem 7.8.].

For higher k, we prove the first half of the proposition by noting that by Proposition 4.45, if there were an infinite number of $B \in \pi_2(\boldsymbol{x}, \boldsymbol{\theta}_+(k), \boldsymbol{y})$ with an element $u_B \in \mathcal{M}^B(\boldsymbol{x}, \boldsymbol{\theta}_+(k-2), \boldsymbol{y}'; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\maltese}; \sigma_{\clubsuit}, \sigma_{\maltese})$, we can glue u_B to a holomorphic anchor to give a curve $u_{B'} \in \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}_+(k-3), \boldsymbol{y}'; S'; \vec{\rho}'_{\clubsuit}, \vec{\rho}'_{\maltese}; \vec{\sigma}'_{\clubsuit}, \vec{\sigma}'_{\maltese})$, where $(\vec{\rho}'_{\clubsuit}, \vec{\sigma}'_{\clubsuit}), (\vec{\rho}'_{\maltese}, \vec{\sigma}'_{\maltese})$ is a $k-1$ shipping of $(\vec{\rho}, \vec{\sigma})$. By our induction hypothesis, this is impossible.

For the second half, we let κ be the corresponding constant for m_i^0. Again, if some $i > \kappa$ does not satisfy that $m_i^k = 0$, we may glue the corresponding polygon to a holomorphic anchor, implying that m_i^{k-1} is also nonzero. By our induction hypothesis, this is, again, impossible. □

We now turn to the main definition and theorem of the paper. Consider the 2^h elementary splayings $\mathcal{D}(\iota(\delta))$, where $\delta \in \{m, l\}^h$. The module $\mathrm{Poly}(\mathcal{D})$ is generated over the field $\mathbb{F}_2$ by the sum $\mathcal{G} := \bigoplus_{\delta \in \{m,l\}^h} \mathcal{G}(\mathcal{D}(\iota(\delta)))$. We endow this module with the obvious right action of the ring of idempotents of $\mathcal{T}^h$ by setting

$$\boldsymbol{x} \cdot \iota(\delta) = \begin{cases} \boldsymbol{x} & \text{if } \boldsymbol{x} \in \mathcal{G}(\mathcal{D}(\iota(\delta))), \\ 0 & \text{otherwise.} \end{cases}$$

We define an $\mathcal{A}_\infty$ module structure upon $\mathrm{Poly}(\mathcal{D})$ by first defining maps

$$n_{i_1,\dots,i_h} : \mathrm{Poly}(\mathcal{D}) \otimes_{\mathcal{I}} \mathcal{T}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{T}_h^{\otimes i_h} \times \vec{\sigma}(i_1, \dots, i_h) \to \mathrm{Poly}(\mathcal{D}),$$

where $\vec{\sigma}(i_1, \dots, i_h)$ denotes the set of all splicings of all sets of sequences of Reeb chords in

$$\mathcal{T}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{T}_h^{\otimes i_h}.$$

Suppose that $\vec{\rho}_{\maltese} = \vec{\rho}_1, \dots, \vec{\rho}_h$ is a set of sequences of Reeb multichords in $\mathcal{T}^h$, where $|\vec{\rho}_j| = i_j$. For every interleaved splicing $\vec{\sigma}_{\maltese}$ of $\vec{\rho}_{\maltese}$, we define

$$n_{i_1,\dots,i_h}(\boldsymbol{x}, \vec{\rho}_{\maltese}, \vec{\sigma}_{\maltese}) := \sum_{\boldsymbol{y} \in \mathcal{G}(\mathcal{D})} \sum_{\substack{B \in \pi_2^{\boldsymbol{\theta}_+}(\boldsymbol{x}, \boldsymbol{y}) \\ \chi(S) = \chi_{\mathrm{emb}}(S) \\ \mathrm{ind}(B, S, \vec{\rho}_{\maltese}) = 2 - |\vec{\sigma}_{\maltese}|}} \#\mathcal{M}_{\maltese}^B(\boldsymbol{x}, \boldsymbol{\theta}_+(|\vec{\sigma}_{\maltese}| - 2), \boldsymbol{y}'; S; \vec{\rho}_{\maltese}; \vec{\sigma}_{\maltese}) \cdot \boldsymbol{y},$$

where the nearest-point map takes $\boldsymbol{y}'$ to $\boldsymbol{y}$.

We define

$$m_{i_1,\dots,i_h} : \mathrm{Poly}(\mathcal{D}) \otimes_{\mathcal{I}} \mathcal{A}_1^{\otimes i_1} \otimes \cdots \otimes \mathcal{A}_h^{\otimes i_h} \to \mathrm{Poly}(\mathcal{D})$$

by

$$m_{i_1,\dots,i_h}(x, \vec{\rho}) := \sum_{\vec{\sigma} \text{ an interleaving for } \vec{\rho}} n_{i_1,\dots,i_h}(x, \vec{\rho}, \vec{\sigma}).$$

We first verify that the modules $\mathrm{Poly}_k(\mathcal{D})$ are $\mathcal{A}_\infty$ modules.

Theorem 5.3 *The maps $\{n^k_{i_1,\dots,i_h}\}_{i_1,\dots,i_h}$ satisfy the partial $\mathcal{A}_\infty$ relation. As such, the data*

$$(\mathrm{Poly}_k(\mathcal{D}), \{m^k_{i_1,\dots,i_h}\}^\infty_{i_1,\dots,i_h})$$

forms an $\mathcal{A}_\infty$ module over $\mathcal{T}^h$.

Proof Recall the relation which we need to verify is

$$0 = \sum_{\substack{\vec{\lambda}\star\vec{\delta}=\vec{\rho}\\ \vec{\sigma}_1\star\vec{\sigma}_2=\vec{\sigma}}} n(n(x,\vec{\lambda},\vec{\sigma}_1),\vec{\delta},\vec{\sigma}_2) + \sum_{i,j \text{ compatible}} n(x,\bar{\mu}^i_j(\vec{\rho}),\vec{\sigma}^i_j) + \sum_{k' \text{ collidable}}^{|\vec{\sigma}|-1} n(x,\vec{\rho},\vec{\sigma}(k')).$$

Consider a one-dimensional moduli space $\mathcal{M} = \mathcal{M}^B(\boldsymbol{x}, \boldsymbol{\theta}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\text{◭}}, \vec{\rho}_{\text{◮}}; \vec{\sigma}_{\text{◭}}, \vec{\sigma}_{\text{◮}})$, where $\vec{\sigma}_{\text{◭}}\star\vec{\sigma}_{\text{◮}} = \vec{\sigma}$, and $\vec{\sigma}_{\text{◮}}$ is a k-splicing. By Theorem 4.47, the ends of this moduli space are formed of the following.

- Spinal two-storey ends, corresponding with terms in the first sum in the statement of the theorem.
- Mixed two-storey ends, which, by Proposition 4.53 are either composition ends or collision ends (see Definition 4.54). Composition ends correspond (by Proposition 4.55) with terms in the second sum in the statement where i, j satisfy that $\rho^i_j \in P^{i'}_j \in \vec{\sigma}$ for $i' \geq k$, and ρ^i_j, ρ^{i+1}_j are equal to ρ_1, ρ_2 or ρ_2, ρ_3; collision ends correspond with terms in the third sum in the statement where $k' \geq k$.
- (i,j)-composition ends which correspond with terms in the second sum where i, j satisfy that $\rho^i_j \in P^{i'}_j$ for $i' < k$.
- k'-collision ends, which correspond with terms in the third sum where $k' < k$.
- Cut-vanishing ends, corresponding with terms in the second sum in the statement where i, j satisfy that $\rho^i_j \in P^{i'}_j \in \vec{\sigma}$ for $i' \geq k$, and $\rho^i_j \cdot \rho^{i+1}_j$ is equal to ρ_{123}.

By Theorem 4.47, the number of these is zero modulo two; it corresponds that the sum is zero, as required. □

We will not show directly that $\mathrm{Poly}_k(\mathcal{D})$ is an invariant of $\mathcal{D}$ up to equivalence — indeed, we have not discussed any notion of equivalence for diagrams. Instead, we just need to understand how $\mathrm{Poly}_k(\mathcal{D})$ depends upon the choice of generic admissible compatible almost-complex structures $\boldsymbol{J}$.

Proposition 5.4 *Let $\boldsymbol{J}_0$ and $\boldsymbol{J}_1$ be generic, admissible, compatible families of almost-complex structures. Then the $\mathcal{A}_\infty$ modules $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J}_0)$ and $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J}_1)$ are $\mathcal{A}_\infty$ homotopy equivalent.*

Proof This follows by modifying [14, Section 7.3.1] to the case of splayed polygons — combining it with, say, [13, Proposition 3.30], and the proof of Theorem 5.3 above. □

Theorem 5.5 *Let $k \geq 0$ be an integer. Then the $\mathcal{A}_\infty$ modules $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J})$ and $\mathrm{Poly}_{k+1}(\mathcal{D}, \boldsymbol{J})$ are $\mathcal{A}_\infty$ homotopy equivalent.*

Proof By Proposition 4.65, and the finiteness of the zero-dimensional moduli spaces involved in the definition of the m_i^k and m_i^{k+1} there is some $T > 0$ such that, for every $\boldsymbol{x}$, $\boldsymbol{y}$, B, S, $\vec{\rho}$ and $\vec{\sigma}$, the moduli spaces $\mathcal{M}^B(\boldsymbol{x}, \boldsymbol{\theta}_+(k), \boldsymbol{y}; S; \vec{\rho}_{\clubsuit}, \vec{\rho}_{\clubsuit}; \vec{\sigma}_{\clubsuit}, \vec{\sigma}_{\clubsuit}, \boldsymbol{J}^T)$ and

$$\mathcal{M}_{\text{splayed}} := \bigcup_{\substack{(B';\vec{\rho}'_{\clubsuit},\vec{\sigma}'_{\clubsuit};\vec{\rho}'_{\clubsuit},\vec{\sigma}'_{\clubsuit}) \\ \text{splays} \\ (B;\vec{\rho}_{\clubsuit},\vec{\sigma}_{\clubsuit};\vec{\rho}_{\clubsuit},\vec{\sigma}_{\clubsuit})}} \mathcal{M}^{B'}(\boldsymbol{x}, \boldsymbol{\theta}_+(k-1), \boldsymbol{y}; S'; \vec{\rho}'_{\clubsuit}, \vec{\rho}'_{\clubsuit}; \vec{\sigma}'_{\clubsuit}, \vec{\sigma}'_{\clubsuit}; \boldsymbol{J}).$$

are identified.

It follows that if we sum the number of elements of moduli spaces of both types over all $B \in \pi_2^{\boldsymbol{\theta}_+}(\boldsymbol{x}, \boldsymbol{y})$, we get the same number modulo 2. These are the coefficients of $\boldsymbol{y}$ in the maps $n^k_{i_1,\dots,i_h}(\boldsymbol{x}, \vec{\rho})$, defined with respect to $\boldsymbol{J}^T$, and of $\boldsymbol{y}$ in $n^{k+1}_{i_1,\dots,i_h}(\boldsymbol{x}, \vec{\rho})$, defined with respect to $\boldsymbol{J}$, respectively. It follows that the $\mathcal{A}_\infty$ modules $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J}^T)$ and $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J})$ are isomorphic as $\mathcal{A}_\infty$ modules.

By Proposition 5.4, $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J}^T)$ is $\mathcal{A}_\infty$ homotopy equivalent to $\mathrm{Poly}_k(\mathcal{D}, \boldsymbol{J})$, and so the proposition follows. □

We finally state and prove the main theorem of this paper:

Theorem 5.6 $\mathrm{Poly}(\mathcal{D})$ *is an* $\mathcal{A}_\infty$ *module, and* $\mathrm{Poly}(\mathcal{D})$ *is* $\mathcal{A}_\infty$ *homotopy equivalent to* $\mathrm{BSA}(\mathcal{D})$.

Proof Let κ denote the constant defined in Proposition 5.2. The only nonzero maps in Poly_κ are determined by moduli spaces $\mathcal{M}^B(\boldsymbol{x}, \boldsymbol{\theta}_+(k), \boldsymbol{y}; S; \varnothing, \vec{\rho}_{\clubsuit}; \varnothing, \vec{\sigma}_{\clubsuit}; \boldsymbol{J})$, which, by Proposition 4.24, are all identified with the moduli spaces used in the definition of the structure maps of $\mathrm{Poly}(\mathcal{D})$. As such, $\mathrm{Poly}(\mathcal{D})$ is an $\mathcal{A}_\infty$-module which is $\mathcal{A}_\infty$ isomorphic to $\mathrm{Poly}_\kappa(\mathcal{D})$. Moreover, by Proposition 5.1 and Theorem 5.5,

$$\mathrm{BSA}(\mathcal{D}) \equiv \mathrm{Poly}_0(\mathcal{D}) \cong \mathrm{Poly}_1(\mathcal{D}) \cong \cdots \cong \mathrm{Poly}_{\kappa-1}(\mathcal{D}) \cong \mathrm{Poly}_\kappa(\mathcal{D}) \equiv \mathrm{Poly}(\mathcal{D}).$$ □

References

[1] **L V Ahlfors**, *Conformal invariants: topics in geometric function theory*, McGraw-Hill, New York (1973) MR Zbl

[2] **D Auroux**, *Fukaya categories of symmetric products and bordered Heegaard–Floer homology*, J. Gökova Geom. Topol. 4 (2010) 1–54 MR Zbl

[3] **F Bourgeois**, *A Morse–Bott approach to contact homology*, PhD thesis, Stanford University (2002) Available at `https://www.proquest.com/docview/305591502`

[4] **F Bourgeois**, **Y Eliashberg**, **H Hofer**, **K Wysocki**, **E Zehnder**, *Compactness results in symplectic field theory*, Geom. Topol. 7 (2003) 799–888 MR Zbl

[5] **E Eftekhary**, *Floer homology and splicing knot complements*, Algebr. Geom. Topol. 15 (2015) 3155–3213 MR Zbl

[6] **J Hanselman**, *Bordered Heegaard Floer homology and graph manifolds*, Algebr. Geom. Topol. 16 (2016) 3103–3166 MR Zbl

[7] **T Hockenhull**, *Gluing bypass exact triangles*, in preparation

[8] **H Hofer**, **V Lizan**, **J-C Sikorav**, *On genericity for holomorphic curves in four-dimensional almost-complex manifolds*, J. Geom. Anal. 7 (1997) 149–159 MR Zbl

[9] **A S Levine**, *Knot doubling operators and bordered Heegaard Floer homology*, J. Topol. 5 (2012) 651–712 MR Zbl

[10] **R Lipshitz**, *A cylindrical reformulation of Heegaard Floer homology*, Geom. Topol. 10 (2006) 955–1096 MR Zbl Corrections at http://pages.uoregon.edu/lipshitz/MoreCylErat.pdf

[11] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Computing $\widehat{HF}$ by factoring mapping classes*, Geom. Topol. 18 (2014) 2547–2681 MR Zbl

[12] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bimodules in bordered Heegaard Floer homology*, Geom. Topol. 19 (2015) 525–724 MR Zbl

[13] **R Lipshitz**, **P S Ozsváth**, **D P Thurston**, *Bordered Floer homology and the spectral sequence of a branched double cover, II: The spectral sequences agree*, J. Topol. 9 (2016) 607–686 MR Zbl

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

[15] **C Manolescu**, **P Ozsváth**, *Heegaard Floer homology and integer surgeries on links* (2010) arXiv 1011.1317 To appear in Geom. Topol.

[16] **D McDuff**, **D Salamon**, *J-holomorphic curves and symplectic topology*, 2nd edition, Amer. Math. Soc. Colloq. Publ. 52, Amer. Math. Soc., Providence, RI (2012) MR Zbl

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

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

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

[20] **P Ozsváth**, **Z Szabó**, *Holomorphic disks, link invariants and the multi-variable Alexander polynomial*, Algebr. Geom. Topol. 8 (2008) 615–692 MR Zbl

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

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

[23] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) Available at `https://www.proquest.com/docview/305332635`

[24] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

[25] **P Seidel**, *Lectures on four-dimensional Dehn twists*, from "Symplectic 4-manifolds and algebraic surfaces", Lecture Notes in Math. 1938, Springer (2008) 231–267 MR Zbl

[26] **R Zarev**, *Bordered sutured Floer homology*, PhD thesis, Columbia University (2011) Available at `https://www.proquest.com/docview/868276640`

Frome, Somerset, United Kingdom

thomas.hockenhull@gmail.com

Received: 3 April 2019 Revised: 18 July 2023

msp

Algebraic & Geometric Topology **25**:6 (2025) 3225–3249
DOI: 10.2140/agt.2025.25.3225
Published: 1 October 2025

Exact Lagrangian tori in symplectic Milnor fibers constructed with fillings

ORSOLA CAPOVILLA-SEARLE

We use exact Lagrangian fillings and Weinstein handlebody diagrams to construct infinitely many distinct exact Lagrangian tori in 4-dimensional Milnor fibers of isolated hypersurface singularities with positive modality. We also provide a generalization of a criterion for when the symplectic homology of a Weinstein 4-manifold is nonvanishing given an explicit Weinstein handlebody diagram.

53Dxx; 57Kxx

1 Introduction

Given a Weinstein handlebody diagram of a Weinstein domain (X, λ, ϕ), where Λ is a link of Legendrian attaching spheres, one can construct a closed exact Lagrangian submanifold $\overline{\Sigma}$ of (X, λ, ϕ) by taking the union of an exact Lagrangian filling Σ of Λ and the core Lagrangian disks of the handles attached along Λ. For any such Lagrangian, $[\overline{\Sigma}] \in H_n(X)$ is a primitive class. It is not known whether, for every closed exact Lagrangian submanifold Σ of a Weinstein domain (X, λ, ϕ) that is primitive in homology, there exists a Weinstein handlebody diagram of (X, λ, ϕ) such that Σ can be built from an exact Lagrangian filling; see Casals and Murphy [7, Remark 2.22]. Vianna [41, Section 6.2] constructed an infinite family of exact Lagrangian tori in the 4-dimensional Milnor fibers of parabolic singularities. We show that there are infinite families of exact Lagrangian tori in a larger set of Milnor fibers of isolated hypersurface singularities using fillings:

Theorem 1.1 *For any $p, r \geq 1$ and $q \geq 3$ there exist infinitely many Hamiltonian nonisotopic exact Maslov-0 Lagrangian tori in Milnor fibers of $T_{p,q,r}$ singularities such that the Lagrangian tori are all smoothly isotopic and primitive in homology.*

We obtain the following corollary by either applying properties of Milnor fibers of adjacent singularities to Theorem 1.1 or to Vianna's result [41, Section 6.2].

Corollary 1.2 *If M_f is the Milnor fiber of a positive-modality isolated hypersurface singularity f, then M_f contains infinitely many Hamiltonian nonisotopic exact Maslov-0 Lagrangian tori that are all smoothly isotopic.*

Milnor fibers of isolated hypersurface singularities have been extensively studied in homological mirror symmetry, singularity theory and low-dimensional topology. An isolated hypersurface singularity is the equivalence class of the germ of a holomorphic function, $f : \mathbb{C}^{n+1} \to \mathbb{C}$ whose differential has an isolated

© 2025 The Author, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

zero at the origin. The Milnor fiber of such a singularity is the smooth manifold $M_f = f^{-1}(\varepsilon_\delta) \cap B_\delta(0)$ for suitably small $\delta > 0$ and $\varepsilon_\delta > 0$ [36]. The modality of a singularity f can be thought of as the dimension of a parameter space covering a neighborhood of f in the space of singularities after an appropriate holomorphic reparametrization. Isolated hypersurface singularities of modality 0 and 1 have been classified by Arnold, as summarized in [3, Chapter 2.3]. Unimodular isolated singularities are either parabolic, $T_{p,q,r}$ singularities or one of Arnold's 14 exceptional singularities. We focus on the Milnor fibers of $T_{p,q,r}$ singularities which were shown by Keating [31, Section 2] to be exact symplectomorphic to the affine variety

$$T_{p,q,r} = \{(x, y, z) \in \mathbb{C}^3 \mid x^p + y^q + z^r + xyz = 1\},$$

for $p, q, r \in \mathbb{Z}_{\geq 1}$ and $1/p + 1/q + 1/r \leq 1$.

The only possible exact Lagrangian submanifolds of 4-dimensional Milnor fibers of modality-0 singularities are spheres; see Abouzaid and Smith [1] and Ishii, Ueda and Uehara [28]. This is not the case for 4-dimensional Milnor fibers of positive-modality singularities. See Keating [33] for examples of monotone Lagrangian surfaces in Milnor fibers of positive-modality singularities, and more generally in Brieskorn–Pham hypersurfaces. Keating in [31, Theorem 5.7] constructed an exact Lagrangian torus in the 4-dimensional Milnor fibers of unimodular singularities using explicit Lefschetz fibrations of the 4-dimensional affine varieties $T_{p,q,r}$. Shende, Treumann and Williams in [39, Example 6.3] outline an argument that there are in fact infinitely many distinct Lagrangian tori that are not Hamiltonian isotopic in $T_{p,q,r}$ for $p, q, r \geq 1$. These potentially distinct tori are constructed from fillings of Legendrians in distinct Weinstein handlebodies of $T_{p,q,r}$ that are symplectomorphic but not yet shown to be Weinstein homotopic. In contrast, the tori we construct in Theorem 1.1 are obtained as fillings from a single Weinstein handlebody decomposition. See [27, Remark 3.17; 32, Section 7.4] for a description of $T_{p,q,r}$ as D^*T^2 with $(p + q + r)$ 2-handles attached along Legendrians. If $(p, q, r) \in \{(3, 3, 3), (2, 4, 4), (2, 3, 6)\}$, then $T_{p,q,r}$ compactifies to a del Pezzo surface; see [31, Proposition 5.19]. Vianna in [41, Section 6.2] used this fact to construct infinitely many distinct exact Lagrangian tori in $T_{3,3,3}$, $T_{2,4,4}$ and $T_{2,3,6}$ using different almost toric fibrations of the corresponding del Pezzo surfaces.

The Lagrangian torus that Keating constructs in $T_{p,q,r}$ [31] implies that the Fukaya category of $T_{p,q,r}$ is not split generated by any collection of vanishing cycles as it has one generator for which this is true. Analogously, a corollary of Theorem 1.1 is that the wrapped Fukaya category $T_{p,q,r}$ has an infinite number of generators which are not split generated by the vanishing cycles.

We start by producing a Weinstein handlebody diagram of the 4-dimensional Weinstein manifold $T_{p,q,r}$ using an algorithm established by Casals and Murphy [7], called the affine dictionary, that allows one to obtain the Weinstein handlebody diagram of a Weinstein manifold from particular Lefschetz fibrations, such as the Lefschetz fibration of $T_{p,q,r}$ given by Keating. A change in convention for the ordering of vanishing cycles and the Hurwitz move between the affine dictionary and Keating's Lefschetz fibration of $T_{p,q,r}$ resulted in an incorrect handlebody diagram in Figure 22 of [7]. We provide the corrected Weinstein handlebody diagram in Figure 3.

Proposition 1.3 *The Weinstein 4-dimensional domains*

$$T_{p,q,r} = \{(x, y, z) \in \mathbb{C}^3 \mid x^p + y^q + z^r + xyz = 1\}$$

have the Weinstein handlebody diagram shown in Figure 3 for $p, q, r \geq 0$.

We apply Legendrian Kirby calculus moves to the Weinstein handlebody diagram shown in Figure 3 to find another Weinstein handlebody diagram of $T_{p,q,r}$ for $p, r \geq 1$ and $q \geq 3$ that contains the Legendrian sublink $\Lambda(\beta_{22})$ given by the (-1) closure of the positive braid $(\sigma_2\sigma_3\sigma_1\sigma_2)\sigma_1^2\sigma_3^2 \in \mathrm{Br}_4^+$. The (-1) closure of a braid $\beta \in \mathrm{Br}_n^+$ is the Legendrian link given by placing β in a standard contact neighborhood of the standard max-tb Legendrian unknot in $(\mathbb{R}^3, \xi_{\mathrm{std}})$. See the last diagram on the bottom right of Figure 7 for a front of $\Lambda(\beta_{22})$ as a (-1) closure of a positive braid. Casals and Ng [8, Propositions 7.4 and 7.5] proved that $\Lambda(\beta_{22})$ has infinitely many distinct exact Lagrangian fillings of genus 1 using Floer-theoretic methods. We use their result in conjunction with one technical lemma (Lemma 3.2) to prove Theorem 1.1.

Weinstein handlebody diagrams can also allow one to determine whether certain symplectic invariants vanish without computing them explicitly. Suppose that X_Λ is a 4-dimensional Weinstein manifold whose 2-handles are attached along a Legendrian link Λ. The *symplectic homology* of a Weinstein manifold X_Λ can be computed using the Legendrian contact homology differential graded algebra of Λ, denoted by $\mathcal{A}(\Lambda; \partial)$; see Bourgeois, Ekholm and Eliashberg [4] and Ekholm [16]. We extend a criterion of Leverson [34, Corollary 1.5] for the nonvanishing of the symplectic homology of a Weinstein manifold given a Weinstein handlebody diagram. Symplectic homology, denoted by SH, is a difficult invariant to compute, and this extension gives an easy way to check whether it vanishes for a slightly broader class of Weinstein manifolds.

Theorem 1.4 *Let X_Λ be the Weinstein 4-manifold resulting from attaching 2-handles along a Legendrian link $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\mathrm{std}}\right)$ with n link components. If there is any sublink Λ' with l link components for $l \leq n$ such that its Legendrian differential graded algebra has a representation*

$$\rho\colon (\mathcal{A}(\Lambda'; \mathbb{Z}[t_1^{\pm 1}, \dots, t_l^{\pm 1}]), \partial) \to \mathrm{End}(V),$$

where V is a vector space over $\mathbb{Q}$ and $\rho(t_k) = -\mathrm{Id}$ for $k = 1, \dots, l$, then $\mathrm{SH}(X_\Lambda) \neq 0$.

An important theme in symplectic and contact topology is understanding flexibility phenomena which refers to when an object or situation is governed only by homotopy data. Flexible Weinstein manifolds abide by an h-principle (see Cieliebak and Eliashberg [13]) and it is often difficult to determine whether a Weinstein manifold is flexible. Since a flexible Weinstein manifold has vanishing symplectic homology (see Cieliebak [12]), Theorem 1.4 also gives a criterion for when a Weinstein manifold is not flexible.

Corollary 1.5 *Let X_Λ be the Weinstein 4-manifold resulting from attaching 2-handles along a Legendrian link $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\mathrm{std}}\right)$ with n link components. Suppose there is a sublink Λ' with l link components for $l \leq n$ such that its Legendrian differential graded algebra has a representation*

$$\rho\colon (\mathcal{A}(\Lambda'; \mathbb{Z}[t_1^{\pm 1}, \dots, t_l^{\pm 1}]), \partial) \to \mathrm{End}(V),$$

where V is a vector space over $\mathbb{Q}$ and such that $\rho(t_k) = -\mathrm{Id}$ for $k = 1, \dots, l$. Then X_Λ is not a flexible Weinstein manifold.

For an example of an application of Theorem 1.4 and Corollary 1.5, see work of the author and Acu, Gadbled, Marinković, Murphy, Starkston and Wu [2, Corollary 9.7], which showed that the Weinstein complement of certain divisors in toric 4-manifolds are not flexible.

Outline In Section 2 we provide definitions and background material on Weinstein 4-manifolds, exact Lagrangian fillings and the Legendrian contact homology differential graded algebra. In Section 3 we prove the technical lemmas that allow us to use results of Casals and Ng [8] in the proof of Theorem 1.1. In Section 4 we give the necessary background on Milnor fibers and Lefschetz fibrations and prove Theorem 1.1, Corollary 1.2 and Proposition 1.3. Finally, in Section 5 we prove the new nonvanishing criterion for symplectic homology, Theorem 1.4.

Acknowledgments The author thanks their adviser Lenhard Ng for many helpful conversations and suggestions. This project developed from a suggestion of Roger Casals, whom the author would also like to thank for various helpful conversations. The author also thanks James Hughes, Adam Levine, Ailsa Keating, Emmy Murphy and Daping Weng for useful conversations. The author was first introduced to the topic of Weinstein manifolds and Lefschetz fibrations in the 2017 Kylerec workshop. Finally, the author thanks the referee for valuable comments and suggestions. This work was supported by the NSF grants DGE-1644868 and DMS-1707652.

2 Background

2.1 Weinstein 4-manifolds

We consider only Weinstein 4-manifolds, which we will now briefly review. See [13] for more details on Weinstein manifolds in general. Let (X, ω) be a symplectic manifold with nonempty boundary. A 1-form λ is a Liouville 1-form if $d\lambda = \omega$. The vector field Z_λ that is ω-dual to λ ($\iota_{Z_\lambda}\omega = \lambda$) is called the Liouville vector field. A domain (X, λ, Z_λ) is a Liouville domain if the Liouville vector field Z_λ is complete and transverse to the boundary ∂X, and points outward along ∂X. For a Liouville domain (X, λ, Z_λ), there is an induced contact structure on ∂X given by $\ker(\lambda|_{\partial X})$. The completion of a Liouville domain (X, λ, Z_λ) is the noncompact manifold constructed from X by attaching to the boundary of X the symplectization of the boundary, $X \cup_\partial (\mathbb{R} \times \partial(X))$. A *Weinstein domain* $(X, \lambda, Z_\lambda, \phi)$ is a Liouville domain (X, λ, Z_λ) such that Z_λ is gradient-like with respect to an exhausting Morse function $\phi\colon X \to \mathbb{R}$. A *Weinstein homotopy* on a Weinstein cobordism or manifold is a smooth family of Weinstein structures $(X_t, \lambda_t, Z_t, \phi_t)$ for $t \in [0, 1]$, where we allow birth–death type degenerations such that the associated Liouville structures (X_t, λ_t, Z_t) form a Liouville homotopy. If two Weinstein domains $(X_0, \lambda_0, Z_0, \phi_0)$ and $(X_1, \lambda_1, Z_1, \phi_1)$ are Weinstein homotopic, then their completions are symplectomorphic.

The model Weinstein handle $\mathcal{H}$ of index k in dimension $2n$ for $k \leq n$,

$$\left(\mathbb{D}^k \times \mathbb{D}^{2n-k}, \sum_{i=1}^{n} y_i \, dx_i, Z_k, \phi_k\right) \subset (\mathbb{R}^{2n}, \omega_{\text{std}}),$$

is described in detail in [42]. In particular, $\mathcal{H}$ has convex boundary $S^{k-1} \times \mathbb{D}^{2n-k}$ (Z_k points outward to the handle), and concave boundary $\mathbb{D}^k \times S^{2n-k-1}$ (Z_k points inward to the handle). The convex boundary has a natural contact structure induced by Z_k. Then the components of the handle are

- the *core of the handle* $\mathbb{D}^k \times \{0\}$, a Lagrangian submanifold of $\mathcal{H}$,
- the *cocore of the handle* $\{0\} \times \mathbb{D}^{2n-k}$, a Lagrangian submanifold of $\mathcal{H}$,
- the *attaching sphere* $S^{k-1} \times \{0\}$, an isotropic sphere in the convex boundary ($S^{k-1} \times \mathbb{D}^{2n-k}$), and
- the *belt sphere* $\{0\} \times S^{2n-k-1}$ in the concave boundary ($\mathbb{D}^k \times S^{2n-k-1}$).

A Weinstein handlebody decomposition can then be given explicitly by attaching Weinstein handles each with locally defined Morse functions that one can assemble into a global Morse function. Weinstein handlebody decompositions only contain Weinstein k-handles for $k \leq n$ [21]. In fact, the Weinstein handlebody decomposition of any $2n$-dimensional Weinstein manifold $(X, \lambda, Z_\lambda, \phi)$ is encoded by the Legendrian submanifold $\Lambda \subset (\partial X_0, \lambda, Z_\lambda, \phi)$ corresponding to the attaching spheres of the critical n-handles, where $(X_0, \lambda, Z_\lambda, \phi)$ is the Weinstein subcritical domain given by attaching handles of index strictly less than n.

Weinstein 4-manifolds only admit handlebody decompositions with 0-, 1- and 2-handles. Attaching m Weinstein 1-handles to a Weinstein 0-handle $(\mathbb{D}^4, \omega_{\text{std}})$ gives the 4-manifold whose boundary is $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$. The attaching region of a 1-handle is $S^0 \times \mathbb{D}^3$. Diagrammatically, we will consider S^3 as $\mathbb{R}^3$ with a point at infinity and draw the attaching region of the 1-handle as a pair of 3-balls that are identified with a reflection. A 2-handle is then attached along a Legendrian knot $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$. For a 4-dimensional Weinstein 2-handle attached along a nullhomologous Legendrian link $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$, the framing is given by $\text{tb}(\Lambda) - 1$, where $\text{tb}(\Lambda)$ is the Thurston–Bennequin number of Λ. Although the Thurston–Bennequin number is not invariant under all Legendrian isotopies in $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$, the canonical framing of the 2-handle is invariant under contactomorphisms.

Gompf in [26, Theorem 2.2] showed that Legendrian links in $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$ can be isotoped into a standard form where the attaching spheres of the 1-handles are all vertically stacked and the Legendrian links can run through the 1-handles but are otherwise contained to a rectangular area between the pairs of 3-balls representing the 1-handles. Within this rectangular area, the contact structure, thanks to a contactomorphism, is $(\mathbb{R}^3, \xi_{\text{std}} = \ker(dz - y\,dx))$, so we can use the front projection within this region to draw the fronts of the Legendrian links. The resulting diagram on the plane is a Weinstein handlebody diagram in Gompf standard form; see for example Figure 1. The Thurston–Bennequin number of a

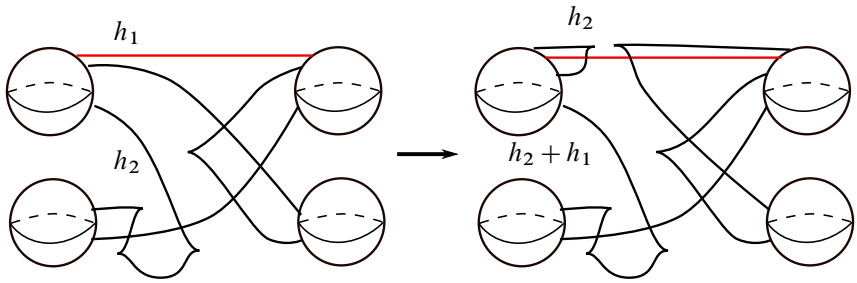


Figure 1: A Legendrian handle slide of h_2 over h_1.

Legendrian $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$ can be computed from the front projection as in the case of a Legendrian link $\Lambda \subset (\mathbb{R}^3, \xi_{\text{std}})$. Two Legendrian links in $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$ are related by Legendrian isotopies if and only if their front projections are related by a set of six moves [26], three of which are the usual Legendrian Reidemeister moves and three of which are referred to as Gompf moves. Gompf moves 4 and 5 correspond to passing cusps and crossings through 1-handles, respectively. Gompf move 6 corresponds to moving a Legendrian strand past the attaching region of a 1-handle.

Two Weinstein handlebody diagrams in Gompf standard form represent equivalent Weinstein domains if they are related by the following moves: isotropic or Legendrian isotopies of the attaching spheres, handle slides, handle cancellations and handle additions. If one has two Weinstein k-handles h_1 and h_2, a *handle slide* of h_1 over h_2 is given by isotoping the attaching sphere of h_1 and pushing it through the belt sphere of h_2, where the isotopy is an isotropic or Legendrian isotopy. In the case of a handle slide for a 4-dimensional 1-handle, one drags one of the 3-balls in the attaching region of h_1 over h_2 and back, following a Legendrian path. If h_1 and h_2 are 4-dimensional Weinstein 2-handles, with attaching spheres $\Lambda_1, \Lambda_2 \subset \left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$, then the Weinstein handlebody diagram after sliding h_1 over h_2 has attaching spheres Λ_1' and Λ_2, where Λ_1' has a front projection given by taking a connect sum of Λ_1 with a push off of Λ_2 [7, Proposition 2.14; 14, Proposition 1]. Figure 1 shows an example of a handle slide for 4-dimensional Weinstein 2-handles. One can add or cancel a pair of Weinstein handles h_1 and h_2 of indices k and $k-1$, respectively, if the attaching sphere of h_2 intersects the belt sphere of h_1 transversely at one point [7, Proposition 2.17]. We call such a pair of handles a *canceling pair*; see Figure 2 for an example of a canceling pair.

2.2 Exact Lagrangian surfaces

Definition 2.1 Let $(Y, \ker(\alpha))$ be either $(\mathbb{R}^3, \xi_{\text{std}})$ or $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$. Let Λ_+ and Λ_- be Legendrian links in $(Y, \ker(\alpha))$. An *exact Lagrangian cobordism* Σ from Λ_- to Λ_+ is an embedded oriented Lagrangian surface in the symplectization $(\mathbb{R}_t \times Y, d(e^t\alpha))$ that has cylindrical ends and is exact in the sense that, for some $N > 0$:

(1) $\Sigma \cap ([-N, N] \times Y)$ is compact.

(2) $\Sigma \cap ((N, \infty) \times Y) = (N, \infty) \times \Lambda_+$.

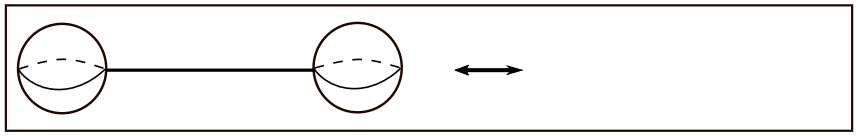


Figure 2: A canceling pair of 4-dimensional 1- and 2-handles.

(3) $\Sigma \cap ((-\infty, -N) \times Y) = (-\infty, -N) \times \Lambda_-$.

(4) There exists a function $f : \Sigma \to \mathbb{R}$ and constants $\mathfrak{c}_\pm$ such that $e^t \alpha|_\Sigma = df$, where $f|_{(-\infty,-N)\times\Lambda_-} = \mathfrak{c}_-$ and $f|_{(N,\infty)\times\Lambda_+} = \mathfrak{c}_+$. We call f the primitive on Σ.

An *exact Lagrangian filling* Σ of a Legendrian link Λ is an exact oriented Lagrangian cobordism from $\varnothing$ to Λ.

The *Maslov number* of an exact Lagrangian surface Σ is the greatest common divisor of the Maslov numbers of the closed loops in Σ. For a Legendrian Λ, the Maslov number of Λ is the Maslov number of the surface $\mathbb{R} \times \Lambda$. We will always assume that all Legendrians and Lagrangian cobordisms are Maslov-0.

Remark 2.2 A *rank-n local system* of an exact Lagrangian cobordism Σ is a representation $\pi_1(\Sigma) \to \mathrm{GL}(n, \mathbb{F})$, where $\mathbb{F}$ is a field. Local systems allow one to enrich various invariants by recording homotopic data. By abuse of notation we refer to representations $\pi_1(\Sigma) \to \mathrm{GL}(1, \mathbb{Z}) = \{+1, -1\}$ for orientable surfaces Σ as (rank-1) local systems over Σ. Furthermore, since $\mathrm{GL}(1, \mathbb{Z})$ is abelian, it suffices to consider the maps $H_1(\Sigma; \mathbb{Z}) \to \mathrm{GL}(1, \mathbb{Z})$.

One can find exact Lagrangian surfaces in a Weinstein manifold using its Weinstein handlebody decomposition as follows. If $\Sigma \subset (X_0, \lambda, Z_\lambda, \phi)$ is an exact Maslov-0 Lagrangian filling of any of the critical attaching spheres $\Lambda \subset \partial X_0$, then the union of Σ with the Lagrangian cores of the handles attached to Λ is a closed exact Maslov-0 Lagrangian $\overline{\Sigma}$. See [7, Section 2.6] for some combinatorial constructions of exact Lagrangian submanifolds obtained from fillings of Legendrian attaching spheres for high-dimensional Weinstein manifolds. We will use the fact that distinct exact Lagrangian fillings Σ_1 and Σ_2 can lead to distinct closed exact Lagrangian surfaces $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ [8, Proposition 7.11]. We first review how to distinguish fillings using Floer-theoretic invariants in the following subsection.

2.3 Augmentations of the Legendrian contact differential graded algebra

The Legendrian contact homology differential graded algebra, $(\mathcal{A}(\Lambda; R), \partial)$, also known as the Chekanov–Eliashberg dg-algebra, is a Legendrian link invariant that was first defined for Legendrian links of $(\mathbb{R}^3, \ker(dz - y\,dx))$ in [11] and fits into the larger SFT framework of [22]. For a more recent survey on this Legendrian invariant see [24].

Let $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\mathrm{std}}\right)$ be an oriented Legendrian link with n components $\Lambda = \Lambda^{(1)} \cup \cdots \cup \Lambda^{(n)}$ and with k basepoints such that every component of Λ has at least one basepoint decorated with a monomial $\pm t_i^{\pm}$. We will always assume that Λ has rotation number zero. A Reeb chord a of Λ is a

trajectory of the Reeb vector field which begins and ends on Λ. Let $\mathcal{R}(\Lambda)$ denote the set of Reeb chords. For a Reeb chord $a \in \mathcal{R}(\Lambda)$, let a^+ and a^- denote the points of Λ at the end and beginning of the Reeb chord. Then define $r(a), c(a) \in \{1, \dots, n\}$ as follows: $r(a) = i$ if $a^- \subset \Lambda^{(i)}$, and $c(a) = j$ if $a^+ \subset \Lambda^{(j)}$. We say a is a *mixed Reeb chord* if $r(a) \neq c(a)$ and a *pure Reeb chord* otherwise. We denote by $\mathcal{R}_{i,j}(\Lambda)$ the set of Reeb chords a of Λ such that $c(a) = i$ and $r(a) = j$.

The generators of $\mathcal{A}(\Lambda; R)$ The underlying algebra $\mathcal{A}(\Lambda; R)$ is a unital noncommutative graded algebra over the coefficient ring $R = \mathbb{Z}[t_1^{\pm 1}, \dots, t_k^{\pm 1}]$ that is generated by Reeb chords $a \in \mathcal{R}(\Lambda)$, and basepoints labeled by monomials $\pm t_i^{\pm}$ for $1 \leq i \leq k$. There is a surjective map from $\mathbb{Z}[s_1^{\pm}, \dots, s_k^{\pm}]$ to $\mathbb{Z}[H_1(\Lambda; \mathbb{Z})]$ where each basepoint is mapped the homology class of the link component on which the basepoint lies. We assume that the basepoints $t_i^{\pm}$ commute with the Reeb chords.

The grading on $\mathcal{A}(\Lambda; R)$ The grading is defined on the Reeb chord generators via a Conley–Zehnder index, and is defined on a word as the sum of the gradings of the letters in the word. See [17, Section 2.3] or [24, Sections 3.1 and 3.5] for more details.

The differential ∂ The differential ∂ is defined by counting rigid $\mathcal{J}$-holomorphic disks in the symplectization of $\left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$ with boundary on $\mathbb{R} \times \Lambda$. For Reeb chords $a, b_1, \dots, b_m$ of Λ, denote by $\mathcal{M}(a; b_1, \dots, b_m)$ the moduli space of $\mathcal{J}$-holomorphic disks

$$u \colon (\mathbb{D}_{m+1}, \partial \mathbb{D}_{m+1}) \to (\mathbb{R} \times \mathbb{R}^3, \mathbb{R} \times \Lambda)$$

such that

- $\mathbb{D}_{m+1}$ is the disk with $m+1$ boundary points $p, q_1, \dots, q_m$ removed and labeled counterclockwise,
- u is asymptotic to $[0, \infty) \times a$ at p, and
- u is asymptotic to $(\infty, 0] \times b_i$ at q_i.

The quotient of $\mathcal{M}(a; b_1, \dots, b_m)$ by a vertical translation of $\mathbb{R}$ is denoted by $\widetilde{\mathcal{M}}(a; b_1, \dots, b_m)$. If $\dim(\widetilde{\mathcal{M}}(a; b_1, \dots, b_m)) = 0$, then we say the disks $u \in \mathcal{M}(a; b_1, \dots, b_m)$ are rigid. One can count such rigid $\mathcal{J}$-holomorphic disks over the ring $\mathbb{Z}[H_1(\Lambda; \mathbb{Z})]$ as follows. For such a disk u, the boundary segment between q_i and q_{i+1} can be closed off on $\mathbb{R} \times \Lambda$ into a curve by concatenating with the capping paths from q_i^- to $t_{c(q_i)}$ and the capping path from $t_{r(q_{i+1})}$ to q_{i+1}^+ since $c(q_i) = r(q_{i+1})$. Let τ_i denote the homology class of this curve. Then $w(u)$ is the product of the Reeb chords and homology classes $\tau_0 \cdots \tau_m b_1 \cdots b_m$. Here we are setting $b_0 = a$. Moreover, if Λ is spin then the moduli spaces of the $\mathcal{J}$-holomorphic curves that we are considering admit a coherent orientation [18; 29], so one can assign a sign $\operatorname{sgn}(u) \in \{-1, 1\}$ to each rigid $\mathcal{J}$-holomorphic disk. For a Reeb chord $a \in \mathcal{R}(\Lambda)$, the differential is then given by

$$\partial(a) = \sum_{\dim(\widetilde{\mathcal{M}}(a; b_1, \dots, b_m)) = 0} \; \sum_{u \in \mathcal{M}(a; b_1, \dots, b_m)} \operatorname{sgn}(u) w(u). \tag{2-1}$$

Let $\partial t_i^{\pm} = 0$ for any basepoint t_i, and extend the differential to $\mathcal{A}$ using the graded Leibniz rule:

$$\partial(uv) = \partial(u)v + (-1)^{|u|} u \partial(v).$$

There is a combinatorial way to compute the differential from the Lagrangian projection of Λ. See [24, Section 3] for $\Lambda \subset (\mathbb{R}^3, \xi_{\text{std}})$ and [20] for $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\text{std}}\right)$.

One of the main properties of the Legendrian dg-algebra is that it is functorial over exact Lagrangian cobordisms. In particular, exact Lagrangian fillings of Λ induce augmentations ε of $(\mathcal{A}(\Lambda; R), \partial)$ [19; 29].

Definition 2.3 A (*graded*) *augmentation* of $(\mathcal{A}(\Lambda; R), \partial)$ to a ring of coefficients R supported in grading zero is a chain map $\varepsilon\colon (\mathcal{A}(\Lambda; R), \partial) \to (R, 0)$ such that $\varepsilon(1) = 1$, and for any element $a \in (\mathcal{A}(\Lambda; R), \partial)$ with nonzero grading $|a| \neq 0$ we have that $\varepsilon(a) = 0$.

Theorem 2.4 [19; 29] *Suppose that Σ is a spin oriented embedded Maslov-0 exact Lagrangian filling of the Legendrian link $\Lambda \subset (S^3, \xi_{\text{std}})$. Then Σ induces a dg-algebra map*

$$\varepsilon_\Sigma\colon \left(\mathcal{A}(\Lambda; \mathbb{Z}[H_1(\Sigma; \mathbb{Z})]), \partial\right) \to (\mathbb{Z}[H_1(\Sigma; \mathbb{Z})], 0),$$

where $\mathbb{Z}[H_1(\Sigma; \mathbb{Z})]$ is in grading 0. If Σ and Σ' are Lagrangian fillings of Λ such that there exists a Hamiltonian isotopy from Σ_1 to Σ_2 through exact Lagrangian fillings of Λ that fixes the boundary, then the corresponding augmentations ε_Σ and $\varepsilon_{\Sigma'}$ are chain homotopic maps.

Augmented Reeb chords $\varepsilon_\Sigma(a) \in \mathbb{Z}[H_1(\Sigma; \mathbb{Z})]$ count the number of rigid holomorphic disks u asymptotic to a with boundary on Σ and record the homology class of each disk u. See [8, Sections 3.5 and 4] for how to compute ε_Σ explicitly for decomposable exact Lagrangian fillings. Note that not all augmentations are induced by embedded Maslov-0 exact Lagrangian fillings, but they are induced from immersed exact Lagrangian fillings [37, Theorem 1.2]. We only consider embedded exact Lagrangian fillings from here on out.

For an exact embedded Lagrangian filling Σ, the induced augmentation ε_Σ over $\mathbb{Z}[H_1(\Sigma; \mathbb{Z})]$ is a family of augmentations over $\mathbb{Z}$ where each augmentation over $\mathbb{Z}$ corresponds to a choice of local system of Σ:

$$\varepsilon_\Sigma\colon \left(\mathcal{A}(\Lambda; \mathbb{Z}[H_1(\Sigma; \mathbb{Z})]), \partial\right) \to (\mathbb{Z}[H_1(\Sigma; \mathbb{Z})], 0) \to (\mathbb{Z}, 0).$$

One can also enlarge the coefficient ring $\mathbb{Z}[H_1(\Sigma; \mathbb{Z})]$ to include link automorphisms. Suppose that Λ is a link with n components and k basepoints such that each link component has at least one basepoint. Let t_i denote the product of basepoints on $\Lambda^{(i)}$ for $1 \leq i \leq n$, so that t_i is in bijection with the generator of $H_1(\Lambda^{(i)}; \mathbb{Z})$. Let $e_1, \dots, e_n$ be a choice of units in $\mathbb{Z}$. A link automorphism of Λ is an automorphism ψ of $(\mathcal{A}(\Lambda; \mathbb{Z}[H_1(\Sigma; \mathbb{Z})]), \partial)$ such that if a is a Reeb chord of Λ that starts at the i^{th} link component of Λ and ends at the j^{th} link component, then $\psi(a) = e_i e_j^{-1} a$. Now consider $\mathbb{Z}[H_1(\Sigma; \mathbb{Z})] \oplus \mathbb{Z}[t_1^{\pm}, \dots, t_{n-1}^{\pm}]$, where $t_i = e_i/e_n$ for $1 \leq i < n$. Any augmentation over $\mathbb{Z}[H_1(\Sigma; \mathbb{Z})]$ lifts to an augmentation over $\mathbb{Z}[H_1(\Sigma; \mathbb{Z}) \oplus \mathbb{Z}^{n-1}]$ by setting $\tilde{\varepsilon}(a) = e_i e_j^{-1} \varepsilon(a)$. The following definition allows one to work with such augmentations:

Definition 2.5 [8, Definition 3.9] A *k-system of augmentations* of a Legendrian link Λ is an algebra map

$$\varepsilon\colon \mathcal{A}(\Lambda; \mathbb{Z}[s_1^{\pm 1}, \dots, s_k^{\pm 1}]) \to \mathbb{Z}[s_1^{\pm 1}, \dots, s_k^{\pm 1}]$$

such that $\varepsilon\circ\partial=0$, $\varepsilon(1)=1$ and $\varepsilon(a)=0$ for any $a\in\mathcal{A}(\Lambda;\mathbb{Z}[s_1^{\pm1},\dots,s_k^{\pm1}])$ with nonzero grading. Two k-systems of augmentations

$$\varepsilon\colon\mathcal{A}(\Lambda;\mathbb{Z}[s_1^{\pm1},\dots,s_k^{\pm1}])\to\mathbb{Z}[s_1^{\pm1},\dots,s_k^{\pm1}]\quad\text{and}\quad\varepsilon'\colon\mathcal{A}(\Lambda;\mathbb{Z}[s_1^{\pm1},\dots,s_k^{\pm1}])\to\mathbb{Z}[{s'_1}^{\pm1},\dots,{s'_k}^{\pm1}]$$

are equivalent if there exists a $\mathbb{Z}$-algebra isomorphism

$$\phi\colon\mathbb{Z}[s_1^{\pm1},\dots,s_k^{\pm1}]\to\mathbb{Z}[{s'_1}^{\pm1},\dots,{s'_k}^{\pm1}]$$

such that $\varepsilon'=\phi\circ\varepsilon$. The space of such isomorphisms is parametrized by $\mathbb{Z}_2^k\times\mathrm{GL}_k(\mathbb{Z})$.

Remark 2.6 Let Σ_1 and Σ_2 be two spin oriented embedded Maslov-0 exact Lagrangian fillings of an $(n+1)$-component Legendrian link Λ. By [9, Theorem 1.4], both Σ_1 and Σ_2 realize the smooth 4-ball genus of Λ, so $\operatorname{rank}(H_1(\Sigma_1;\mathbb{Z}))=\operatorname{rank}(H_1(\Sigma_2;\mathbb{Z}))=k$. Then by Theorem 2.4 they both induce $(k+n)$-systems of augmentations $\varepsilon_1\colon\mathcal{A}(\Lambda;R_1)\to R_1$ and $\varepsilon_2\colon\mathcal{A}(\Lambda;R_2)\to R_2$, where $R_i\simeq\mathbb{Z}[H_1(\Sigma_i;\mathbb{Z})]\oplus\mathbb{Z}^n$. If Σ_1 and Σ_2 are Hamiltonian isotopic, then there exists a $\phi\in\mathbb{Z}_2^{k+n}\times\mathrm{GL}_{k+n}(\mathbb{Z})$ such that $\phi(H_1(\Sigma_1;\mathbb{Z}))=H_1(\Sigma_2;\mathbb{Z})$. Then $\phi\circ\varepsilon_1\colon\mathcal{A}(\Lambda;R_1)\to R_2$, and $\phi\circ\varepsilon_1$ is chain homotopic to ε_2 by Theorem 2.4. See [6, Definition 5.5] for when two augmentations of a Legendrian link are chain homotopic when working with coefficients over $\mathbb{Z}$.

Systems of augmentations induced by fillings that are not equivalent up to linear transformations can be used to distinguish exact Lagrangian fillings [5; 8].

2.4 Distinguishing exact Lagrangian surfaces in Weinstein 4-manifolds with augmentations

Let the Weinstein 4-manifold given by attaching 2-handles to a Legendrian link $\Lambda\subset\left(\#^m(S^1\times S^2),\xi_{\mathrm{std}}\right)$ be denoted by X_Λ. Recall that if Σ is an exact Maslov-0 Lagrangian filling of Λ, then the union of Σ with the Lagrangian cores of the handles attached to Λ is a closed exact Maslov-0 Lagrangian $\overline{\Sigma}\subset X_\Lambda$. Let the system of augmentations induced by Σ be denoted by

$$\varepsilon_\Sigma\colon(\mathcal{A}(\Lambda;\mathbb{Z}[H_1(\Sigma;\mathbb{Z})]),\partial)\to(\mathbb{Z}[H_1(\Sigma;\mathbb{Z})],0).$$

In general, composing ε_Σ with a local system on $H_1(\Sigma;\mathbb{Z})$ results in a family of $\mathbb{Z}$-valued augmentations. Such a composition is well behaved under equivalence of local systems and systems of augmentations. Since $H_1(\Sigma;\mathbb{Z})$ and $H_1(\overline{\Sigma};\mathbb{Z})$ are not in general isomorphic (namely, if Λ is a link with more than one component), not all local systems on Σ extend to local systems on $\overline{\Sigma}$. We are interested in the systems of augmentations where this extension is possible.

Definition 2.7 [8, Definition 7.9] Let Σ be an exact Lagrangian filling of a Legendrian link $\Lambda\subset(S^3,\xi_{\mathrm{std}})$ that induces a system of augmentations $\varepsilon_\Sigma\colon(\mathcal{A}(\Lambda;\mathbb{Z}[H_1(\Sigma;\mathbb{Z})]),\partial)\to(\mathbb{Z}[H_1(\Sigma;\mathbb{Z})],0)$, where Λ is equipped with the null-cobordant spin structure. The *restricted system of augmentations* associated to Σ is the composition

$$\varepsilon_{\overline{\Sigma}}\colon(\mathcal{A}(\Lambda;\mathbb{Z}[H_1(\Sigma;\mathbb{Z})]),\partial)\to(\mathbb{Z}[H_1(\Sigma;\mathbb{Z})],0)\to(\mathbb{Z}[H_1(\overline{\Sigma};\mathbb{Z})],0),$$

where the second map is induced by the quotient map $H_1(\Sigma;\mathbb{Z})\to H_1(\overline{\Sigma};\mathbb{Z})$

Suppose one starts with the Legendrian dg-algebra of Λ $(\mathcal{A}(\Lambda;\mathbb{Z}[H_1(\Lambda;\mathbb{Z})]),\partial)$ such that Λ has the Lie group spin structure. To translate to the dg-algebra where Λ has a null-cobordant spin structure we map t_i to $-t_i$, where t_i is the product of basepoints on the link component $\Lambda^{(i)}$ which represents $H_1(\Lambda^{(i)};\mathbb{Z})$. Since each homology component of Λ is nullhomologous in $\overline{\Sigma}$, any restricted augmentation defined over the null-cobordant spin structure must map t_i to -1. For Legendrian knots, $H_1(\Sigma;\mathbb{Z})$ is isomorphic to $H_1(\overline{\Sigma};\mathbb{Z})$ for nullhomologous knots Λ, so all augmentations are restricted for knots.

Definition 2.8 A Legendrian link Λ is *aug-infinite* if the collection of all $\mathbb{Z}$-valued augmentations $\varepsilon_{\overline{\Sigma}}(\mathcal{A}(\Lambda;\mathbb{Z}),\partial)\to(\mathbb{Z},0)$ induced by Maslov-0 exact Lagrangian fillings Σ of Λ, ranging over all such fillings, is infinite. A Legendrian link Λ is *restricted aug-infinite* if the collection of all $\mathbb{Z}$-valued restricted augmentations $\varepsilon_{\overline{\Sigma}}(\mathcal{A}(\Lambda;\mathbb{Z}),\partial)\to(\mathbb{Z},0)$ induced by Maslov-0 exact Lagrangian fillings Σ of Λ, ranging over all such fillings, is infinite.

Remark 2.9 To check whether a Legendrian Λ is restricted aug-infinite one must work with the null-cobordant spin structure, whereas to check whether a Legendrian Λ is aug-infinite one might have to also work with the Lie group spin structure. Also, the number of basepoints on a link component does not matter, as we always consider the product of basepoints representing the homology of that link component.

Proposition 2.10 [8, Proposition 7.11] *Let $\Lambda\subset(S^3,\xi_{\text{std}})$ be a Legendrian link and $\Sigma_1,\Sigma_2\subset(\mathbb{D}^4,\omega_{\text{std}})$ two exact Lagrangian fillings of Λ. Suppose that the two restricted systems of augmentations*

$$\varepsilon_{\overline{\Sigma}_1}:\mathcal{A}(\Lambda)\to\mathbb{Z}[H_1(\overline{\Sigma}_1;\mathbb{Z})],\quad \varepsilon_{\overline{\Sigma}_2}:\mathcal{A}(\Lambda)\to\mathbb{Z}[H_1(\overline{\Sigma}_2;\mathbb{Z})],$$

are not dg-algebra homotopic. Then the exact Lagrangian surfaces $\overline{\Sigma}_1,\overline{\Sigma}_2\subset W(\Lambda)$ are not Hamiltonian isotopic in the Weinstein 4-manifold X_Λ.

The proof of this proposition relies on the fact that for such a Weinstein 4-manifold, X_Λ, the wrapped Fukaya category of X_Λ is generated by the union of cocores of the Weinstein 2-handles of X_Λ [10]. Then, by considering the Yoneda embedding $\text{Hom}(C,-):=C\mathcal{W}(C,-)$, they show $C\mathcal{W}(C,\overline{\Sigma}_1)$ and $C\mathcal{W}(C,\overline{\Sigma}_2)$ are distinct as $C\mathcal{W}(C)$-modules and conclude that $\overline{\Sigma}_1$ and $\overline{\Sigma}_2$ are distinct objects in $\mathcal{W}(X_\Lambda)$.

3 Exact Lagrangian fillings and restricted augmentations

In this section we prove that certain Legendrians are restricted aug-infinite to conclude that Weinstein manifolds whose Weinstein handlebody diagrams contain a restricted aug-infinite Legendrian sublink then have infinitely many distinct closed exact Lagrangians. We begin by upgrading the following proposition:

Proposition 3.1 [8, Proposition 7.5] *Suppose Λ_- is an aug-infinite Legendrian link in $(\mathbb{R}^3,\xi_{\text{std}})$. If there exists a decomposable Maslov-0 exact Lagrangian cobordism Σ from Λ_- to Λ_+, then Λ_+ is also aug-infinite.*

Decomposable exact Lagrangian cobordisms are the concatenation of three types of elementary cobordisms [19]: perturbations of Legendrian isotopy traces, minimum cobordisms (essentially the unique lagrangian disk filling of the max-tb unknot U [23]), and saddle cobordisms. In the following proof we use the combinatorial framework given in [8, Section 3.5] to construct systems of augmentations of decomposable fillings. We will call a product of basepoints on a link component $\Lambda^{(i)}$ distinguished if they represent the generator of $H_1(\Lambda^{(i)};\mathbb{Z})$.

Lemma 3.2 *Suppose $\Lambda_-, \Lambda_+ \subset (\mathbb{R}^3, \xi_{\text{std}})$ are Legendrians links such that there exists a decomposable exact Maslov-0 Lagrangian cobordism from Λ_- to Λ_+. Then any restricted augmentation of Λ_- lifts to a restricted augmentation of Λ_+.*

Proof Let ε_- denote a restricted augmentation induced by a filling on Λ_- and ε_+ the augmentation on Λ_+ induced by concatenating with an elementary cobordism Σ from Λ_- to Λ_+. It suffices to check that for each link component of Λ_+ the distinguished product of basepoints is mapped to -1 by ε_+.

If Σ is induced by a Legendrian isotopy, then both Legendrians have the same number of link components which we can decorate with a single basepoint t_i so that $\varepsilon_+(t_i) = \varepsilon_-(t_i)$. By assumption, t_i is a distinguished basepoint on both Legendrians and ε_- is a restricted augmentation. So $\varepsilon_-(t_i) = -1$ and therefore $\varepsilon_+(t_i) = -1$. If Σ is a minimum cobordism, then $\Lambda_+ = U \sqcup \Lambda_-$. Decorate Λ_+ with n basepoints so each link component is decorated with a single basepoint and U is decorated with t_0. Then $\varepsilon_+(t_0) = -1$, and $\varepsilon_+(t_i) = \varepsilon_-(t_i)$ for $i \neq 0$. For $i \neq 0$, t_i is a distinguished basepoint on each link component of Λ_-, so $\varepsilon_+(t_i) = \varepsilon_-(t_i) = -1$.

Suppose that Σ is a saddle cobordism. Then Λ_+ has one more or one fewer link component than Λ_-. Suppose it is the j^{th} and k^{th} link component of Λ_+ or Λ_- which merge into one. Let σ denote the unstable manifold on Σ and mark with basepoints s and $-s^{-1}$ the two endpoints of σ on Λ_-. See for example [8, Figure 16]. Then

$$\varepsilon_+(t_j) = \varepsilon_-(t_j)s \quad \text{and} \quad \varepsilon_+(t_k) = -\varepsilon_-(t_k)s^{-1},$$

where t_j and t_k are the basepoints on the link component that is split or merged by Σ in Λ_+.

If Λ_+ has one fewer link component than Λ_-, then $t_j t_k$ is a distinguished basepoint on Λ_+ which represents the homology class of one of its link components. Since Λ_- is assumed to be restricted aug-infinite, $\varepsilon_-(t_j) = \varepsilon_-(t_k) = -1$. Therefore $\varepsilon_+(t_j t_k) = -\varepsilon_-(t_j)\varepsilon_-(t_k)ss^{-1} = -1$. If Λ_+ has one more link component than Λ_-, then $\varepsilon_-(t_j t_k) = -1$. Choose a local system such that $\eta(s) = -\varepsilon_-(t_j)$. Then $\eta \circ \varepsilon_+(t_j) = \varepsilon_-(t_j)\eta(s) = -1$, and $\eta \circ \varepsilon_+(t_k) = -\varepsilon_-(t_k)\eta(s^{-1}) = -1$. In both cases, for all other basepoints we have $\varepsilon_+(t_i) = \varepsilon_-(t_i) = -1$. Thus a restricted augmentation on Λ_- lifts to a restricted augmentation on Λ_+. □

The following corollary is an immediate consequence of Lemma 3.2 and Proposition 3.1:

Corollary 3.3 *Suppose Λ_- is a restricted aug-infinite Legendrian link in $(\mathbb{R}^3, \xi_{\text{std}})$. If there exists a decomposable Maslov-0 exact Lagrangian cobordism Σ from Λ_- to Λ_+, then Λ_+ is also restricted aug-infinite.*

Corollary 3.4 *The Legendrian links $\Lambda(\beta_{ab}) \subset (\mathbb{R}^3, \xi_{\text{std}})$ given by the (-1) closure of the positive braid $(\sigma_2\sigma_1\sigma_3\sigma_2)^4\sigma_1^a\sigma_3^b \in \text{Br}_4^+$, for $a, b \geq 1$ are restricted aug-infinite.*

Proof For $a, b > 1$ there exists a Maslov-0 exact Lagrangian cobordism from $\Lambda(\beta_{11})$ to $\Lambda(\beta_{ab})$ constructed using saddle cobordisms. In particular, this cobordism is constructed by pinching $a-1$ and $b-1$ crossings of $\Lambda(\beta_{ab})$ to obtain $\Lambda(\beta_{11})$. Casals and Ng showed that $\Lambda(\beta_{11})$ is restricted aug-infinite in the proof of [8, Corollary 1.7]. By Corollary 3.3, we can then conclude that $\Lambda(\beta_{ab})$ are restricted aug-infinite for $a, b \geq 1$. □

We apply Proposition 2.10 and Corollary 3.4 to conclude the following corollary. The fact that the Lagrangian surfaces we consider are smoothly isotopic follows from [8, Theorem 1.1 and Proposition 7.1].

Corollary 3.5 *Let X be a Weinstein manifold that has a Weinstein handlebody diagram containing as a Legendrian sublink the link $\Lambda(\beta_{ab})$ for some $a, b \geq 1$. Then X contains infinitely many Hamiltonian nonisotopic Lagrangian surfaces of genus $\frac{1}{2}(a+b-c+2)$, where $2 \leq c \leq 4$ is the number of link components of $\Lambda(\beta_{ab})$. These surfaces are all Maslov-0, smoothly isotopic and primitive in homology.*

4 Constructing infinitely many distinct Lagrangian tori in Milnor fibers

Let $f : \mathbb{C}^{n+1} \to \mathbb{C}$ be a holomorphic function such that $f(0) = 0$, $df|_0 = 0$ and $df \neq 0$ on $B_r(0)\backslash\{0\}$ for a sufficiently small $r \in \mathbb{R}_{>0}$. An *isolated hypersurface singularity* at 0 is the equivalence class of the germ of such a holomorphic function f, up to biholomorphic changes of coordinates that fix 0. For an isolated hypersurface singularity f, the *Milnor fiber* of f is the smooth manifold

$$M_f := f^{-1}(\varepsilon_\delta) \cap B_\delta(0)$$

for suitable choices of $\delta > 0$ and $\varepsilon_\delta > 0$ which depends on our choice of δ. Milnor proved that M_f is independent of the aforementioned choices [36]. As the intersection of a hypersurface with a ball, the Milnor fiber has an exact symplectic structure inherited from $\left(\mathbb{C}^{n+1}, d\left(\frac{1}{4}i\sum_{i=0}^n (z_i\, d\bar{z}_i \wedge \bar{z}_i\, dz_i)\right)\right)$. Furthermore, the negative Liouville flow is given by the gradient flow of $h(x)$ with respect to the standard Kähler metric. Choose a cut-off function

$$h_A(x) = \|Ax\|^2$$

for some $A \in \text{GL}_{n+1}(\mathbb{C})$. Then

$$f^{-1}(\varepsilon_\delta) \cap \{h_A(x) \leq \delta\}$$

is an exact symplectic manifold with contact-type boundary since the negative gradient flow of h points strictly inwards at any point of the hypersurface $\|Ax\|^2 = \delta$. Thus we can describe the Milnor fiber as a Liouville domain which we can then complete by attaching cylindrical ends. Call this the completed Milnor fiber of f and denote it by M_f. Keating [31, Lemmas 2.6 and 2.7] showed that M_f is independent of the choice of holomorphic representative of f, A, δ and ε_δ up to exact symplectomorphisms.

Isolated hypersurface singularities are equivalent to polynomials. This follows from the fact that the $(\mu+1)$-jet of a function f at an isolated critical point with Milnor number μ is sufficient. That is, for any other function g with the same $(\mu+1)$-jet, there exists a biholomorphic change of coordinates between f and g. The k-jet of a function is the Taylor series of the mapping truncated at degree k and deleting the constant term. The Milnor number μ of f is the number of isolated nondegenerate singularities near 0 in a Morsification of f.

One important property of a singularity is its modality. The group G of germs of diffeomorphisms $(\mathbb{C}^n, 0) \to (\mathbb{C}^n, 0)$ acts on the space of function germs $f\colon (\mathbb{C}^n, 0) \to (\mathbb{C}, 0)$. The modality of a singularity f is the least integer m such that a sufficiently small neighborhood of f is covered by a finite number of m-parameters of orbits. Moreover, this action of G on the space of function germs $f\colon (\mathbb{C}^n, 0) \to (\mathbb{C}, 0)$ induces an action on the k-jet space of these function germs. In fact, the *modality* of a singularity f is the modality of any of its k-jets for $k \geq \mu(f) + 1$, where μ is the Milnor number of f. Isolated hypersurface singularities of modality 0 and 1 are classified [3]. There are three families of unimodular isolated singularities: three parabolic singularities, the hyperbolic series $T_{p,q,r}$ and 14 exceptional singularities. The hyperbolic singularities $T_{p,q,r}$ are given by

$$x^p + y^q + z^r + axyz,$$

where $a \in \mathbb{C}^*$, $p, q, r \in \mathbb{Z}_{\geq 0}$ and $1/p + 1/q + 1/r < 1$. The three parabolic singularities are the $T_{p,q,r}$ singularities such that $1/p+1/q+1/r = 1$ (ie $(p, q, r) \in \{(3, 3, 3), (2, 4, 4), (2, 3, 6)\}$). If f is a weighted homogeneous function with a single isolated singularity at 0, then M_f is exact symplectomorphic to any of the hypersurfaces $f^{-1}(\varepsilon)$ for any $\varepsilon \in \mathbb{C}^*$ equipped with the standard Kähler exact symplectic form. The $T_{p,q,r}$ singularities are not in general weighted homogeneous, but Keating [31, Lemma 2.16] proved that they are independent of $a \in \mathbb{C}^*$, so we can set $a = 1$.

In summary, Keating [31] showed that the Milnor fiber of $T_{p,q,r}$ singularities are given by the affine varieties $T_{p,q,r} = \{(x, y, z) \in \mathbb{C}^3 \mid x^p + y^q + z^r + xyz = 1\} \subset \mathbb{C}^3$ for p, q and r such that $p, q, r \in \mathbb{Z}_{\geq 0}$ and $1/p + 1/q + 1/r \leq 1$. In [30] she gave a Lefschetz fibration $\Xi\colon T_{p,q,r} \to \mathbb{C}$ shown with regular fiber $\Xi^{-1}(0)$ symplectomorphic to a thrice-punctured torus and $p + q + r + 3$ distinct critical values, which we will use to construct a Weinstein handlebody diagram of $T_{p,q,r}$. We will now briefly review Lefschetz fibrations for 4-manifolds. See [38] for a comprehensive treatment of Lefschetz fibrations.

Definition 4.1 Let $\mathbb{D}^2 \subset \mathbb{C}$ denote the closed unit disk, and let X be a compact complex manifold with corners of real dimension 4. A Lefschetz fibration $\pi\colon X \to \mathbb{D}^2$ is as follows:

- The map π is a proper smooth fibration except for a finite number of critical points $\mathrm{crit}(\pi)$ in the interior of X.
- Near each critical point $p \in \mathrm{crit}(\pi)$ there are local holomorphic coordinates in which π is given by $(z_1, z_2) \to \pi(p) + z_1^2 + z_2^2$. All critical values are distinct.
- Each smooth fiber of π is a Liouville domain.

Consider a Lefschetz fibration on an affine variety $\pi : X \to \mathbb{C}$, and denote the regular fiber $\pi^{-1}(0)$ by F_π and critical values by $c_1, \dots, c_s$ in the interior of $\mathbb{D}^2$. For every critical value c_i there is an associated *vanishing cycle* $V_i \subset F_\pi$, which is the boundary of an embedded Lagrangian disk $\Delta_i \subset X$. The image of the Lagrangian embedded disk $\pi(\Delta_i)$ is an embedded path $\gamma_i : [0,1] \to \mathbb{D}^2$ with endpoints $\gamma_i(0) = 0$ and $\gamma_i(1) = c_i$. The vanishing cycle V_i is then $\Delta_i \cap F_\pi$, a Lagrangian sphere in the regular fiber. A Lefschetz fibration $\pi : X \to \mathbb{C}$ can be recuperated from the data of the regular fiber F_π and the ordered set of vanishing cycles, whose ordering is determined by the cyclic ordering of the vanishing paths.

Giroux and Pardon [25, Theorem 1.5] proved that any $2n$-dimensional Weinstein manifold (X, λ, Z, ϕ) admits a Lefschetz fibration $\pi : (X, \lambda, Z, \phi) \to \mathbb{C}$. In the case that X is an affine variety it is possible to construct a Lefschetz fibration up to isotopy by using a generic hyperplane section [35]. In fact, one can obtain a Weinstein handle decomposition of X from the data of the Lefschetz fibration, that is, from the regular fiber F_π and the set of vanishing cycles $\{V_1, \dots, V_s\}$. We will now describe this process for 4-dimensional Weinstein manifolds. Start with the Weinstein domain $F_\pi \times \mathbb{D}^2$ with contact boundary $F_\pi \times S^1 \subset \partial(F_\pi \times \mathbb{D}^2)$ and attach Weinstein 2-handles along the Legendrian lifts of the exact Lagrangian vanishing cycles $V_1, \dots, V_n$. The ordering of the vanishing cycles determines the relative Reeb height of the Legendrian lifts. Thus the data of a Lefschetz fibration, a regular fiber F and an ordered set of vanishing cycles $\{V_1, \dots, V_s\}$, determines a Weinstein domain which we denote by $\mathrm{lf}(F; V)$. Two Weinstein manifolds $\mathrm{lf}(F; V)$, and $\mathrm{lf}(F'; V')$ are Weinstein equivalent if F' is Weinstein equivalent to F and the vanishing cycles V' are Hamiltonian isotopic to those the vanishing cycles V, up to a cyclic shift. Moreover, there are two additional Weinstein equivalences given by changing the Lefschetz fibration by stabilization or Hurwitz moves. Hurwitz moves are given by

$$\mathrm{lf}(F, \{V_1, \dots, V_i, V_{i+1}, \dots, V_s\}) = \mathrm{lf}(F, \{V_1, \dots, V_{i+1}, \tau_{V_{i+1}}(V_i), \dots, V_s\}),$$
$$\mathrm{lf}(F, \{V_1, \dots, V_i, V_{i+1}, \dots, V_s\}) = \mathrm{lf}(F, \{V_1, \dots, \tau_{V_i}^{-1}(V_{i+1}), V_i, \dots, V_s\}),$$

where τ is a Dehn twist.

Casals and Murphy [7, Recipe 3.3] establish an algorithm they call the affine dictionary, which inputs certain Lefschetz fibrations $\mathrm{lf}(F; V)$ of complex affine varieties and outputs a Weinstein handlebody diagram of $\mathrm{lf}(F; V)$ by providing the fronts of the Legendrian lifts of the vanishing cycles V. We apply the affine dictionary to Keating's Lefschetz fibration $\Xi : T_{p,q,r} \to \mathbb{C}$ in order to produce a Weinstein handlebody diagram of $T_{p,q,r}$. Note that we are using the conventions for Lefschetz fibrations from [7] where we order vanishing cycles counterclockwise and also use the Hurwitz move convention in [7] where $\{V_i, V_{i+1}\} \to \{\tau_{V_i}^{-1} V_{i+1}, V_i\}$.

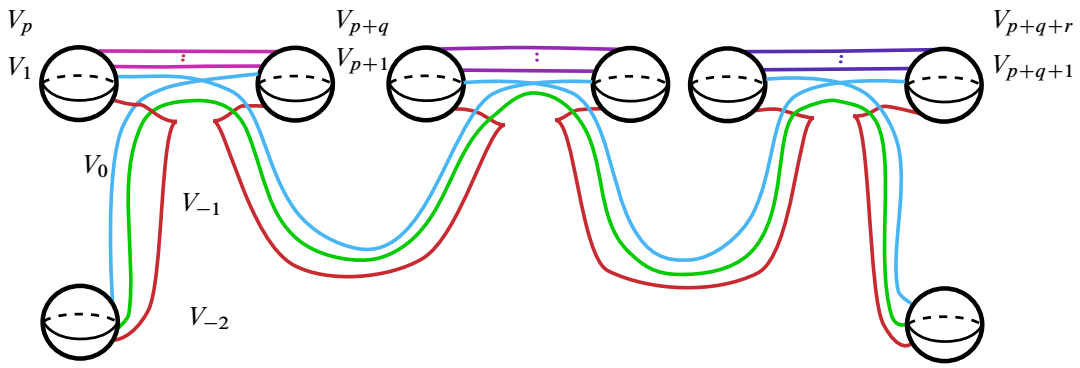


Figure 3: A Weinstein handlebody diagram of the Milnor fiber of a $T_{p,q,r}$-singularity where $p, q, r \geq 0$.

Proposition 1.3 *The Weinstein domains*

$$T_{p,q,r} = \{(x, y, z) \in \mathbb{C}^3 \mid x^p + y^q + z^r + xyz = 1\}$$

have the Weinstein handlebody diagram shown in Figure 3 for $p, q, r \geq 0$.

Proof We start with the Lefschetz fibration $\Xi : T_{p,q,r} \to \mathbb{C}$ from [30] which has vanishing cycles that can be written as Dehn twists of the four curves four curves T, P, Q and R on the regular fiber $\Xi^{-1}(0)$ shown in Figure 4. These vanishing cycles are:

$$\begin{aligned} V_{-2} &= \tau_P \tau_Q \tau_R T, & V_1 &= Q, & V_4 &= \cdots = V_{p+3} = P, \\ V_{-1} &= T, & V_2 &= R, & V_{p+4} &= \cdots = V_{p+q+3} = Q, \\ V_0 &= P, & V_3 &= T, & V_{p+q+4} &= \cdots = V_{p+q+r} = R. \end{aligned}$$

We apply three Hurwitz moves and obtain the following collection of vanishing cycles:

$$\begin{aligned} V_{-2} &= \tau_P \tau_Q \tau_R T, & V_0 &= \tau_P^{-1} \tau_Q^{-1} \tau_R^{-1} T, & V_{p+1} &= \cdots V_{p+q} = Q, \\ V_{-1} &= T, & V_1 &= \cdots = V_p = P, & V_{p+q+1} &= \cdots = V_{p+q+r} = R. \end{aligned}$$

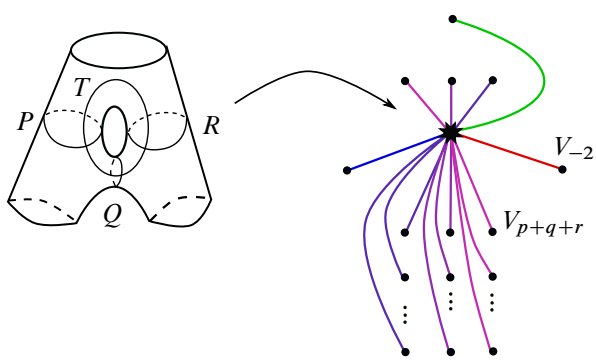


Figure 4: A Lefschetz fibration of $T_{p,q,r}$ whose vanishing cycles are Dehn twists of the four curves T, P, Q and R on the regular fiber.

We now apply the affine dictionary and obtain the Weinstein handlebody diagram shown in Figure 3. First we use the Lefschetz fibration on the regular fiber $\Xi^{-1}(0)$ where the Lagrangian skeleton of consists of Lagrangian spheres T, P, Q and R. We already know how to write the vanishing cycles of Ξ as Dehn twists of these four curves T, P, Q and R, so we are in Step 5 of the affine dictionary and can use [7, Proposition 2.23] to draw the front projection of their Legendrian lifts. Note that the order of the vanishing cycles determines the Reeb height of the Legendrian lifts. □

Theorem 1.1 *For any $p, r \geq 1$ and $q \geq 3$, the Weinstein 4-manifold $T_{p,q,r}$ contains infinitely many Hamiltonian nonisotopic exact Maslov-0 Lagrangian tori that are all smoothly isotopic and primitive in homology.*

Proof We start with the Weinstein handlebody diagram of $T_{p,q,r}$ obtained in Proposition 1.3 and shown in Figure 5. After performing Legendrian isotopy, Legendrian handle slides and handle cancellations, as described in Figures 5, 6 and 7, we find a Weinstein handlebody diagram of $T_{p,q,r}$ that contains a sublink that is Legendrian isotopic to $\Lambda(\beta_{22})$ if $p, r \geq 1$ and $q \geq 3$. It follows from Corollary 3.5 that if $p, r \geq 1$ and $q \geq 3$, then $T_{p,q,r}$ contains infinitely many Hamiltonian nonisotopic Lagrangian tori that are smoothly isotopic and primitive in homology.

We will now give a brief summary of the Legendrian isotopies, handle slides and handle cancellations employed in Figures 5, 6 and 7. For simplicity of notation we do not change the notation for the attaching sphere of a 2-handle after handle slides, although they are distinct as Legendrians. We first simplify the Weinstein handlebody diagram of $T_{p,q,r}$ as shown in Figure 5: A–B perform various handle slides, C cancels out the 1-handles, D–E perform Legendrian Reidemeister moves and F uses the fact that one can handle slide an N-copy of the unknot to a chain of N unknots. As soon as one cancels the 1-handles one can check that $T_{p,q,r}$ has the correct intersection form for $p, q, r \geq 1$. One can also check that $T_{0,0,0}$ is Weinstein homotopy equivalent to D^*T^2, that is that there exists a sequence of handle slides and Legendrian isotopies that takes the Weinstein handlebody diagram of $T_{0,0,0}$ to the standard Weinstein handlebody diagram of D^*T^2. We are now ready to use the Weinstein handlebody diagram of $T_{p,q,r}$ to find closed exact Lagrangian surfaces in $T_{p,q,r}$.

From here on out we do not include the vanishing cycles that are chains of unknots, as they do not play an active role in the computation. Next we simplify the Weinstein handlebody diagram of $T_{p,q,r}$ as shown in Figure 6: A handle slides V_{p+1} under V_0, B–D perform Legendrian Reidemeister moves, E handle slides V_{-1} over V_{p+1}, F performs Legendrian Reidemeister moves, G handle slides V_{p+2} under V_{p+1}, H performs Legendrian Reidemeister moves, I handle slides V_{p+2} over V_0, J performs Legendrian Reidemeister moves, K handle slides V_{p+2} over V_{-1}, L performs Legendrian Reidemeister moves and M handle slides V_{-1} over V_{p+2}. Finally, as described by Figure 7, we perform a series of Legendrian Reidemeister moves on a sublink from the last diagram in Figure 6 to prove that it is Legendrian isotopic to $\Lambda(\beta_{22})$, the (-1) closure of the braid $(\sigma_2\sigma_1\sigma_3\sigma_2)^4\sigma_1^2\sigma_3^2 \in \mathrm{Br}_4^+$. □

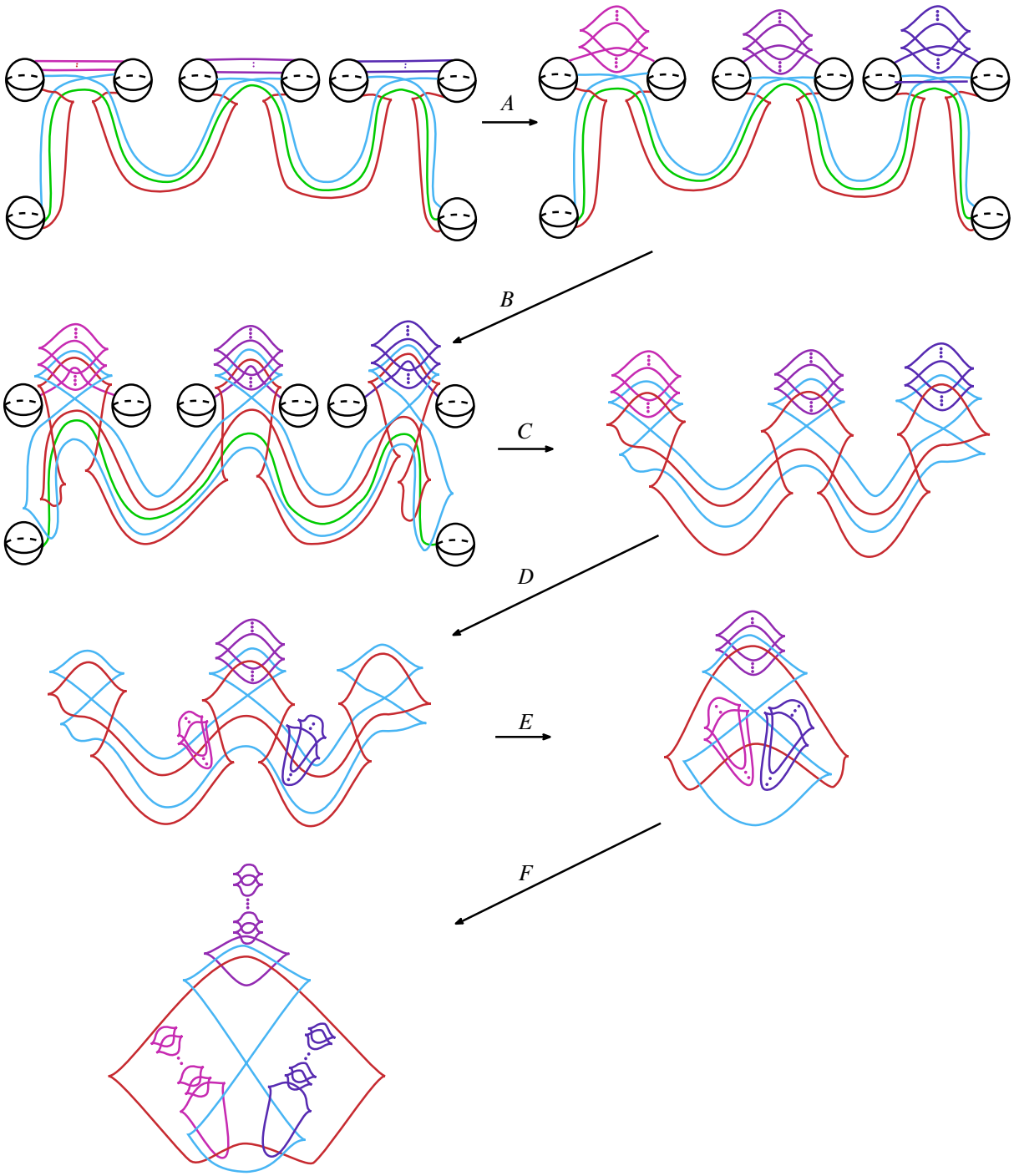


Figure 5: We simplify the Weinstein handlebody diagram of $T_{p,q,r}$ for $p,q,r \geq 1$ as follows: A–B perform various handle slides, C cancels out the 1-handles, D–E perform Legendrian Reidemeister moves and F uses the fact that one can handle slide an N-copy of the unknot to a chain of N unknots.

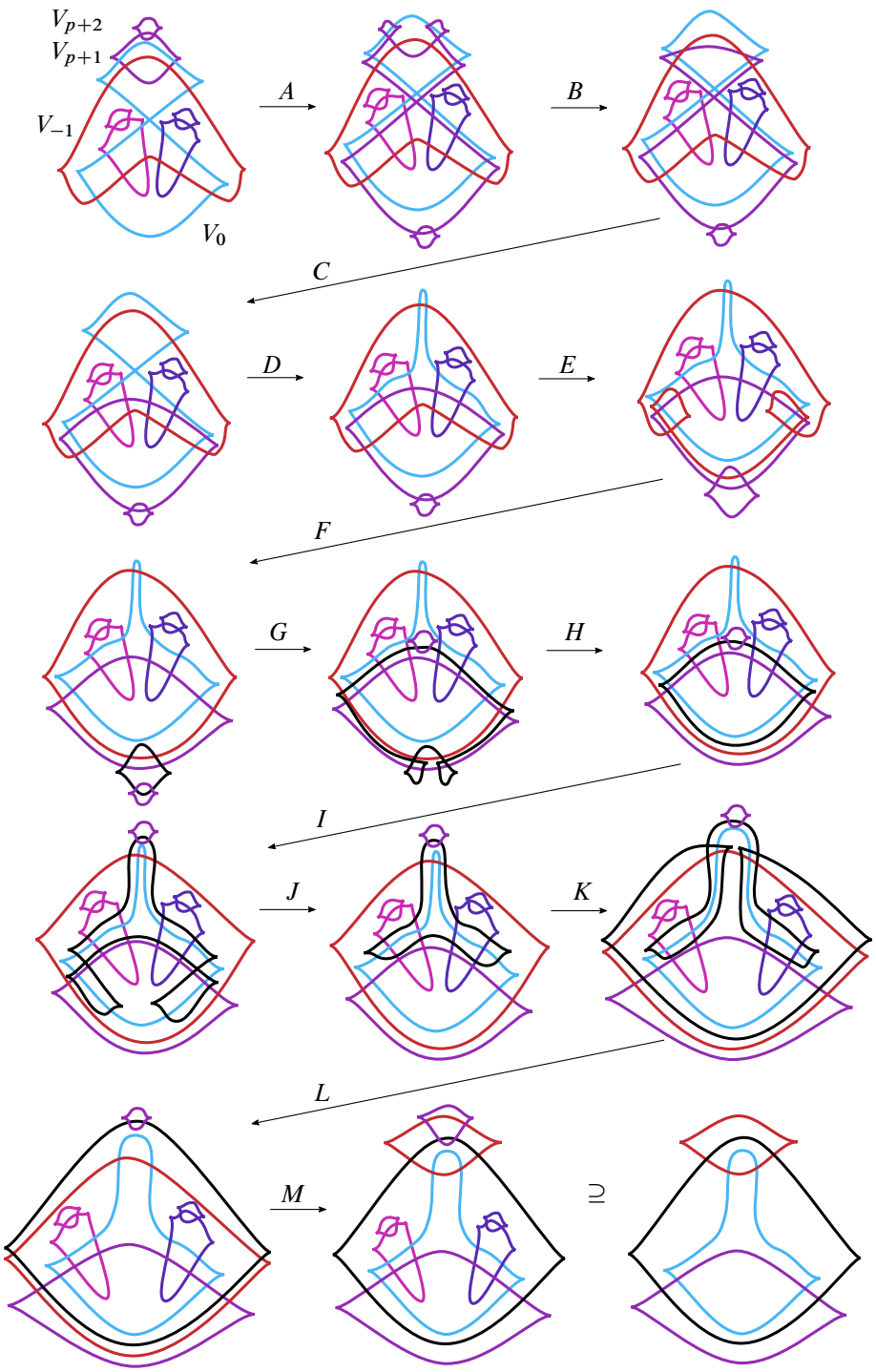


Figure 6: We find a Weinstein handlebody diagram of $T_{p,q,r}$ that contains as a sublink $\Lambda(\beta_{22})$ when $p, r \geq 1$ and $q \geq 3$. The steps A, E, F, I, K and M are handle slides, and the steps B–D, F, H, J and L are Legendrian Reidemeister moves.

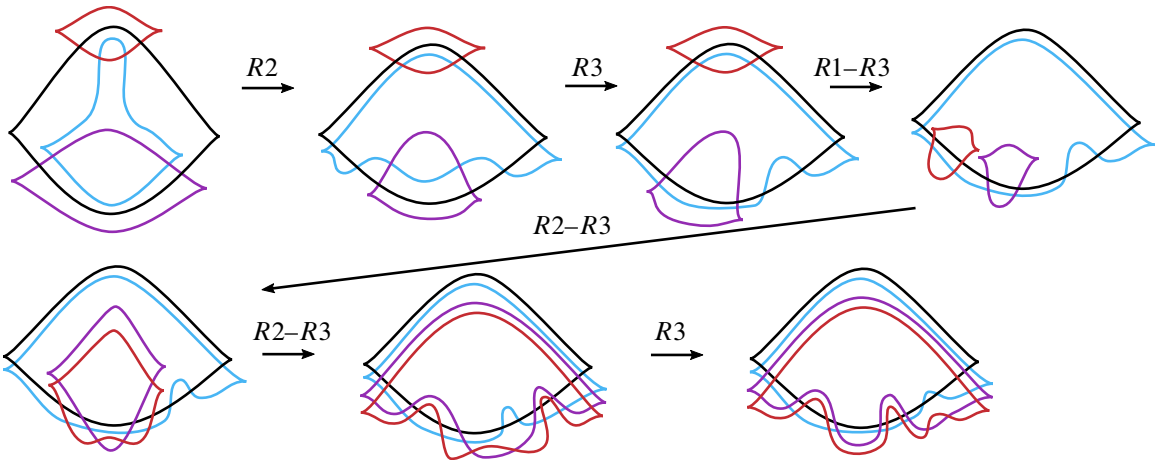


Figure 7: A sequence of Legendrian Reidemeister moves on the Legendrian link $\Lambda(\beta_{22})$, the (-1) closure of $(\sigma_2\sigma_3\sigma_1\sigma_2)^4\sigma_3^2\sigma_1^2 \in \mathrm{Br}_4^+$. We indicate which types of Legendrian Reidemeister moves are used at each intermediate step.

There are also important relations between singularities that one can consider, one of which is whether a singularity is adjacent to another. A singularity $[f]$ is *adjacent* to $[g]$ if there exists an arbitrarily small perturbation p such that $[f+p]=[g]$.

Lemma 4.2 [30, Lemma 9.9] *Suppose $[f]$ and $[g]$ are singularities such that $[f]$ is adjacent to $[g]$. Then there exists an exact symplectic embedding from a noncompleted Milnor fiber of g into a completed Milnor fiber of f.*

Each of Arnold's exceptional singularities is adjacent to one of the three parabolic singularities. For $p'\geq p$, $q'\geq q$ and $r'\geq r$, $T_{p',q',r'}$ is adjacent to $T_{p,q,r}$. More generally, Durfee showed:

Theorem 4.3 [15] *Any positive-modality hypersurface singularity is adjacent to a modality-1 hypersurface singularity.*

Corollary 1.2 *If M_f is the Milnor fiber of a positive-modality isolated hypersurface singularity f, then M_f contains infinitely many Hamiltonian nonisotopic exact Maslov-0 Lagrangian tori that are all smoothly isotopic.*

Proof By Theorem 1.1, the Milnor fiber of any of the three parabolic singularities, that is the $T_{p,q,r}$ singularity for $(p,q,r)\in\{(3,3,3),(2,4,4),(2,3,6)\}$, has infinitely many Hamiltonian nonisotopic exact Maslov-0 Lagrangian tori that are smoothly isotopic. All positive-modality singularities f are adjacent to a unimodular singularity by Theorem 4.3, and all unimodular singularities are adjacent to at least one of the three parabolic singularities. By Lemma 4.2, there then exists an exact symplectic embedding of the noncompleted Milnor fiber of such $T_{p,q,r}$ singularities into M_f. □

5 Criterion for the nonvanishing of the symplectic homology of a Weinstein 4-manifold

Bourgeois, Ekholm and Eliashberg in [4; 16] gave a relation between Legendrian invariants of the critical attaching Legendrian spheres of 2-handles and symplectic invariants of the resulting Weinstein manifold. In particular, they showed that

$$\mathrm{SH}(X_\Lambda) = \mathrm{LH}^{H_0}(\Lambda),$$

where $\mathrm{LH}^{H_0}(\Lambda)$ is the homology of the Hochschild-like complex associated to the Legendrian contact homology differential graded algebra of Λ over $\mathbb{Q}$. Leverson in [34, Corollary 1.5] used this result to show that if Λ has a graded augmentation, then the symplectic homology $\mathrm{SH}(X_\Lambda)$ does not vanish. We provide a generalization of this nonvanishing criterion, where now for Legendrian links Λ that have at least one sublink with a graded augmentation or a graded representation, $\mathrm{SH}(X_\Lambda) \neq 0$. See Figure 8 for three such Legendrians. The first two Legendrian links shown on the left of Figure 8 do not have a graded augmentation but have a sublink (the max-tb unknot link component) that does. The Legendrian knot $m(10_{132})$ whose front is shown on the right of Figure 8 has a graded representation but no graded augmentations or finite graded representations [40, Section 2.2] (the proof is written over $\mathbb{Z}_2$ but also holds with minor modifications over $\mathbb{Z}$).

Theorem 1.4 *Let X_Λ be the Weinstein 4-manifold resulting from attaching 2-handles along a Legendrian link $\Lambda \subset \left(\#^m(S^1 \times S^2), \xi_{\mathrm{std}}\right)$ with n link components. If there is any sublink Λ' with l link components for $l \leq n$ such that its differential graded algebra has a graded representation*

$$\rho\colon (\mathcal{A}(\Lambda'; \mathbb{Z}[t_1^{\pm 1}, \dots, t_l^{\pm 1}]), \partial) \to \mathrm{End}(V),$$

where V is a vector space over $\mathbb{Q}$ and $\rho(t_k) = -\mathrm{Id}$ for $k = 1, \dots, l$, then $\mathrm{SH}(X_\Lambda) \neq 0$.

There is a fair amount of work that goes into defining a chain complex and differential of $\mathrm{LH}^{H_0}(\Lambda)$ in [4]. For the reader's convenience we recall the relevant definitions and notation, with the slight modification that chain complexes are denoted by C instead of CH.

The generators of $\mathrm{LC}^{H_0}(\Lambda)$ Consider the Legendrian contact homology differential graded algebra $(\mathcal{A}(\Lambda; \mathbb{Z}), \partial)$, with the null-cobordant spin structure on each $\Lambda^{(i)}$. For the purposes of this section, we

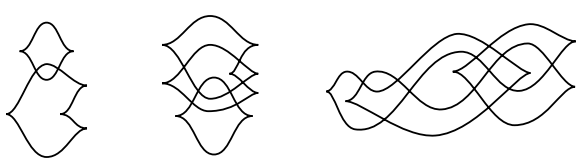


Figure 8: Weinstein handlebody diagrams of Weinstein 4-manifolds with nonvanishing symplectic cohomology. The two Legendrian links on the left have one link component that does not have a graded augmentation and one link component (the max-tb unknot) that does have a graded augmentation. The $m(10_{132})$ Legendrian knot on the right has a graded representation but no graded augmentations.

assume the dg-algebra also had generators e_i, where e_i is an idempotent element associated to the link component $\Lambda^{(i)}$ and which we refer to as an empty Reeb chord. Moreover, we also require the empty Reeb chords to satisfy the relations

$$e_1 + \cdots + e_n = 1, \quad e_i \cdot e_j = \delta_{ij},$$

and

$$e_i \cdot c = \begin{cases} c & \text{if } c \in \mathcal{R}_{ij}(\Lambda) \text{ for some } j, \\ 0 & \text{otherwise,} \end{cases} \qquad c \cdot e_j = \begin{cases} c & \text{if } c \in \mathcal{R}_{ij}(\Lambda) \text{ for some } i, \\ 0 & \text{otherwise.} \end{cases}$$

The subalgebra $\mathrm{LCO}(\Lambda) \subset \mathcal{A}(\Lambda)$ is generated by cyclically composable monomials of Reeb chords, while $\mathrm{LCO}^+(\Lambda) := \mathrm{LCO}(\Lambda)/\langle e_1, \dots, e_n \rangle$ is the subalgebra generated by nontrivial cyclically composable monomials of nonempty Reeb chords. Let $\widetilde{\mathrm{LCO}}^+(\Lambda) := \mathrm{LCO}^+(\Lambda)$ and $\widehat{\mathrm{LCO}}^+(\Lambda) := \mathrm{LCO}^+(\Lambda)[1]$. For any element $w = c_1 \cdots c_j \in \mathrm{LHO}^+(\Lambda)$, write the corresponding elements $\check{w} \in \widetilde{\mathrm{LCO}}^+(\Lambda)$ and $\hat{w} \in \widehat{\mathrm{LCO}}^+(\Lambda)$ as $\check{w} = \check{c}_1 \cdots c_n$ and $\hat{w} = \hat{c}_1 \cdots c_n$. The chain complex is then

$$\mathrm{LC}^{H_0}(\Lambda) = \widetilde{\mathrm{LCO}}^+(\Lambda) \oplus \widehat{\mathrm{LCO}}^+(\Lambda) \oplus \mathbb{Q}\langle \tau_1, \dots, \tau_n \rangle,$$

where each τ_i has grading 0. Each τ_i is in bijective correspondence with each e_i and thus with each Legendrian link component $\Lambda^{(i)}$.

The differential d_{H_0} Given an element $\left(\check{w}, \hat{v}, \sum_{i=1}^n m_i \tau_i\right)$ in $\mathrm{LC}^{H_0}(\Lambda)$, where $\hat{v} = \hat{c}_1 \cdots c_k$ and $m_i \in \mathbb{Q}$, the differential is given by

$$d_{H_0}\left(\check{w}, \hat{v}, \sum_{i=1}^n m_i \tau_i\right) := (\check{d}_{\mathrm{LCO}^+}(\check{w}) + \check{c_1} \cdots c_k - c_1 \cdots \check{c_k}, \hat{d}_{\mathrm{LCO}^+}(\hat{v}), \delta_{H_0}(\check{w} + \hat{v})).$$

The differentials $\check{d}_{\mathrm{LCO}^+}$ and $\hat{d}_{\mathrm{LCO}^+}$ are the differential of $\mathcal{A}(\Lambda)$ restricted to $\widetilde{\mathrm{LCO}}^+(\Lambda)$ and $\widehat{\mathrm{LCO}}^+(\Lambda)$, respectively. The differential δ_{H_0} is defined as follows:

- If c is a nonempty pure Reeb chord beginning and ending on $\Lambda^{(i)}$, then $\delta_{H_0}(\check{c}) = n_{c_i} \tau_i$, where n_{c_i} is the count of the zero-dimensional moduli space of holomorphic disks asymptotic at $+\infty$ to c.
- If w is a nonlinear monomial, then $\delta_{H_0}(\check{w}) = 0$.
- For any element $\hat{v}$, $\delta_{H_0}(\hat{v}) := 0$.

Proof Our goal is to construct a representation $\tilde{\rho}\colon (\mathrm{LC}^{H_0}(\Lambda), d_{H_0}) \to (\mathrm{End}(V^{\oplus n}), 0)$ such that $\tilde{\rho} \circ d_{H_0} = 0$. We use such a representation $\tilde{\rho}$ to show that $d_{H_0}(\tau_k) = 0$ and $\tau_k \notin \mathrm{Im}(d_{H_0})$ for $k = 1, \dots, n$. We can then conclude that $\mathrm{LCH}^{H_0}(\Lambda) = \mathrm{SH}(X_\Lambda)$ is nonzero. We will label the link components in $\Lambda = \Lambda^{(1)} \cup \cdots \cup \Lambda^{(n)}$ so that Λ' is given by the first l link components, $\Lambda^{(1)} \cup \cdots \cup \Lambda^{(l)}$.

First define a graded map $\rho'\colon \mathcal{A}(\Lambda) \oplus \mathbb{Q}\langle \tau_1, \dots, \tau_n \rangle \to \mathrm{End}(V^{\oplus n})$ as follows. For $1 \le k \le l$,

$$\rho'(t_k)_{ij} = \begin{cases} \mathrm{Id} & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases}$$

For $k > l$,

$$\rho'(\tau_k)_{ij} = 0.$$

For $1 \leq k \leq l$,

$$\rho'(\tau_k)_{ij} = \begin{cases} \mathrm{Id} & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases}$$

If $c \in \mathcal{R}_{kr}(\Lambda)$ is a Reeb chord beginning on Λ_k and ending on Λ_r and $1 \leq k, r \leq l$, then

$$\rho'(c)_{ij} = \begin{cases} \rho(c) & \text{if } i = k \text{ and } j = r, \\ 0 & \text{otherwise.} \end{cases}$$

If $c \in \mathcal{R}_{kr}(\Lambda)$ but $l < k$ or $l < r$, then

$$\rho'(c)_{ij} = 0.$$

Thus, for any element $\left(\check{w}, \hat{v}, \sum_{i=1}^{n} m_i \tau_i\right) \in \mathrm{LC}^{H_0}(\Lambda)$, where $\hat{v} = \hat{c}_1 \cdots c_k$, and $m_i \in \mathbb{Q}$, define

$$\tilde{\rho}\left(\check{w}, \hat{v}, \sum_{i=1}^{n} m_i \tau_i\right) = \rho'(w) + \rho'(m_i \tau_i).$$

We need to check that $\tilde{\rho} \circ d_{H_0} = 0$:

$$\begin{aligned} \tilde{\rho}\left(d_{H_0}\left(\check{w}, \hat{v}, \sum_{i=1}^{n} m_i \tau_i\right)\right) &= \tilde{\rho}(dw + \check{c_1} \cdots c_j - c_1 \cdots \check{c_j}, \hat{d}_{\mathrm{LCO}^+}(\hat{v}), \delta_{H_0}(\check{w} + \hat{v})) \\ &= \rho'(\check{d}_{\mathrm{LCO}^+}(\check{w}) + c_1 \cdots c_k - c_1 \cdots c_k) + \rho'(\delta_{H_0}(\check{w} + \hat{v})) \\ &= \rho'(\check{d}_{\mathrm{LCO}^+}(\check{w}) + \delta_{H_0}(\check{w})). \end{aligned}$$

By definition of δ_{H_0} there are two cases to consider. If w is a nonlinear monomial that begins and ends on the link component Λ_k for $1 \leq k \leq n$, then

$$\tilde{\rho}\left(d_{H_0}\left(\check{w}, \hat{v}, \sum_{i=1}^{n} m_i \tau_i\right)\right) = \rho'(\check{d}_{\mathrm{LCO}^+}(\check{w}) + 0),$$

and by definition of ρ'

$$\rho'(\check{d}_{\mathrm{LCO}^+}(\check{w}))_{ij} = \begin{cases} \rho(dw) & \text{if } i = j = k, \\ 0 & \text{otherwise,} \end{cases}$$

since $\check{d}_{\mathrm{LCO}^+} = d|_{\widetilde{\mathrm{LCO}}^+}$. If $w = c$ is a Reeb chord beginning and ending on the link component Λ_k for $1 \leq k \leq n$, and because we can identify τ_i with the empty Reeb word $e_i \in \mathrm{LCO}(\Lambda)$, we can conclude that

$$\tilde{\rho}\left(d_{H_0}\left(\check{w}, \hat{v}, \sum_{i=1}^{n} m_i \tau_i\right)\right) = \rho'(\check{d}_{\mathrm{LCO}^+}(\check{c}) + n_{c_k} \tau_k) = \rho'(d_{\mathrm{LCO}}(c)).$$

Then

$$\rho'(d_{\mathrm{LCO}}(c))_{ij} = \begin{cases} \rho(d(c)) & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Since we started with a representation ρ such that $\rho \circ \partial = 0$, we can conclude that $\tilde{\rho} \circ d_{H_0} = 0$. Finally, we observe that τ_k for $k = 1, \ldots, n$ is a cycle because $d_{H_0}(0, 0, \tau_k) = (0, 0, 0)$. Moreover, $\tilde{\rho}(\tau_k) = \rho(t_k) = -\mathrm{Id} \neq 0$ and $\tilde{\rho} \circ d_{H_0} = 0$, so we can conclude that $\tau_k \notin \mathrm{Im}(d_{H_0})$ for $k = 1, \ldots, n$. □

References

[1] **M Abouzaid**, **I Smith**, *Exact Lagrangians in plumbings*, Geom. Funct. Anal. 22 (2012) 785–831 MR Zbl

[2] **B Acu**, **O Capovilla-Searle**, **A Gadbled**, **A Marinković**, **E Murphy**, **L Starkston**, **A Wu**, *Weinstein handlebodies for complements of smoothed toric divisors*, Mem. Amer. Math. Soc. 1561, Amer. Math. Soc., Providence, RI (2025) MR Zbl

[3] **V I Arnold**, **V A Vasilev**, **V V Goryunov**, **O V Lyashko**, *Dynamical systems*, *VI*, Encycl. Math. Sci. 6, Springer (1993) MR

[4] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR Zbl

[5] **O Capovilla-Searle**, **R Casals**, *On Newton polytopes of Lagrangian augmentations*, Bull. Lond. Math. Soc. 56 (2024) 1263–1290 MR Zbl

[6] **O Capovilla-Searle**, **N Legout**, **M Limouzineau**, **E Murphy**, **Y Pan**, **L Traynor**, *Obstructions to reversing Lagrangian surgery in Lagrangian fillings*, J. Symplectic Geom. 22 (2024) 599–672 Zbl

[7] **R Casals**, **E Murphy**, *Legendrian fronts for affine varieties*, Duke Math. J. 168 (2019) 225–323 MR Zbl

[8] **R Casals**, **L Ng**, *Braid loops with infinite monodromy on the Legendrian contact DGA*, J. Topol. 15 (2022) 1927–2016 MR Zbl

[9] **B Chantraine**, *Lagrangian concordance of Legendrian knots*, Algebr. Geom. Topol. 10 (2010) 63–85 MR Zbl

[10] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Geometric generation of the wrapped Fukaya category of Weinstein manifolds and sectors*, Ann. Sci. École Norm. Sup. 57 (2024) 1–85 MR Zbl

[11] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR Zbl

[12] **K Cieliebak**, *Handle attaching in symplectic homology and the chord conjecture*, J. Eur. Math. Soc. 4 (2002) 115–142 MR Zbl

[13] **K Cieliebak**, **Y Eliashberg**, *From Stein to Weinstein and back: symplectic geometry of affine complex manifolds*, Amer. Math. Soc. Colloq. Publ. 59, Amer. Math. Soc., Providence, RI (2012) MR Zbl

[14] **F Ding**, **H Geiges**, *Handle moves in contact surgery diagrams*, J. Topol. 2 (2009) 105–122 MR Zbl

[15] **A H Durfee**, *Fifteen characterizations of rational double points and simple critical points*, Enseign. Math. 25 (1979) 131–163 MR Zbl

[16] **T Ekholm**, *Holomorphic curves for Legendrian surgery*, preprint (2019) arXiv 1906.07228

[17] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in* $\mathbb{R}^{2n+1}$, J. Differential Geom. 71 (2005) 177–305 MR Zbl

[18] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Orientations in Legendrian contact homology and exact Lagrangian immersions*, Internat. J. Math. 16 (2005) 453–532 MR Zbl

[19] **T Ekholm**, **K Honda**, **T Kálmán**, *Legendrian knots and exact Lagrangian cobordisms*, J. Eur. Math. Soc. 18 (2016) 2627–2689 MR Zbl

[20] **T Ekholm**, **L Ng**, *Legendrian contact homology in the boundary of a subcritical Weinstein 4-manifold*, J. Differential Geom. 101 (2015) 67–157 MR Zbl

[21] **Y Eliashberg**, *Topological characterization of Stein manifolds of dimension* > 2, Internat. J. Math. 1 (1990) 29–46 MR Zbl

[22] **Y Eliashberg**, **A Givental**, **H Hofer**, *Introduction to symplectic field theory*, Geom. Funct. Anal. (2000) 560–673 MR Zbl

[23] **Y Eliashberg**, **L Polterovich**, *Local Lagrangian 2-knots are trivial*, Ann. of Math. 144 (1996) 61–76 MR Zbl

[24] **J B Etnyre**, **L L Ng**, *Legendrian contact homology in* $\mathbb{R}^3$, from "Surveys in 3-manifold topology and geometry", Surv. Differ. Geom. 25, International, Boston, MA (2022) 103–161 MR Zbl

[25] **E Giroux**, **J Pardon**, *Existence of Lefschetz fibrations on Stein and Weinstein domains*, Geom. Topol. 21 (2017) 963–997 MR Zbl

[26] **R E Gompf**, *Handlebody construction of Stein surfaces*, Ann. of Math. 148 (1998) 619–693 MR Zbl

[27] **P Hacking**, **A Keating**, *Homological mirror symmetry for log Calabi–Yau surfaces*, Geom. Topol. 26 (2022) 3747–3833 MR Zbl

[28] **A Ishii**, **K Ueda**, **H Uehara**, *Stability conditions on* A_n*-singularities*, J. Differential Geom. 84 (2010) 87–126 MR Zbl

[29] **C Karlsson**, *A note on coherent orientations for exact Lagrangian cobordisms*, Quantum Topol. 11 (2020) 1–54 MR Zbl

[30] **A M Keating**, *Dehn twists and free subgroups of symplectic mapping class groups*, J. Topol. 7 (2014) 436–474 MR Zbl

[31] **A Keating**, *Lagrangian tori in four-dimensional Milnor fibres*, Geom. Funct. Anal. 25 (2015) 1822–1901 MR Zbl

[32] **A Keating**, *Homological mirror symmetry for hypersurface cusp singularities*, Selecta Math. 24 (2018) 1411–1452 MR Zbl

[33] **A Keating**, *Families of monotone Lagrangians in Brieskorn–Pham hypersurfaces*, Math. Ann. 380 (2021) 975–1035 MR Zbl

[34] **C Leverson**, *Augmentations and rulings of Legendrian links in* $\#^k(S^1 \times S^2)$, Pacific J. Math. 288 (2017) 381–423 MR Zbl

[35] **M McLean**, *Lefschetz fibrations and symplectic homology*, Geom. Topol. 13 (2009) 1877–1944 MR Zbl

[36] **J Milnor**, *Singular points of complex hypersurfaces*, Ann. of Math. Stud. 61, Princeton Univ. Press (1968) MR Zbl

[37] **Y Pan**, **D Rutherford**, *Functorial LCH for immersed Lagrangian cobordisms*, J. Symplectic Geom. 19 (2021) 635–722 MR Zbl

[38] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR Zbl

[39] **V Shende**, **D Treumann**, **H Williams**, *On the combinatorics of exact Lagrangian surfaces*, preprint (2016) arXiv 1603.07449

[40] **S Sivek**, *The contact homology of Legendrian knots with maximal Thurston–Bennequin invariant*, J. Symplectic Geom. 11 (2013) 167–178 MR Zbl

[41] **R Vianna**, *Infinitely many monotone Lagrangian tori in del Pezzo surfaces*, Selecta Math. 23 (2017) 1955–1996 MR Zbl

[42] **A Weinstein**, *Contact surgery and symplectic handlebodies*, Hokkaido Math. J. 20 (1991) 241–251 MR Zbl

Department of Mathematics, University of California Davis
Davis, CA, United States
ocapovillasearle@ucdavis.edu

Received: 11 February 2022 Revised: 5 December 2023

msp

Algebraic & Geometric Topology **25**:6 (2025) 3251–3270
DOI: 10.2140/agt.2025.25.3251
Published: 1 October 2025

A note on embeddings of 3-manifolds in symplectic 4-manifolds

ANUBHAV MUKHERJEE

We show that any closed oriented 3-manifold can be topologically embedded in some simply connected closed symplectic 4-manifold and that it can be made a smooth embedding after one stabilization. As a corollary of the proof, we show that the homology cobordism group is generated by Stein fillable 3-manifolds. We also find obstructions on smooth embeddings: there exist 3-manifolds that cannot smoothly embed in a way that appropriately respects orientations in any symplectic manifold with a weakly convex boundary.

57K43; 53D05

1 Introduction

Embedding 3-manifolds into higher-dimensional spaces presents an intriguing challenge. Whitney's embedding theorem [31] establishes that every closed oriented 3-manifold smoothly embeds in $\mathbb{R}^6$. Hirsch [16] later refined this by demonstrating that any 3-manifold can be smoothly embedded in S^5. Concurrently, Lickorish [19] and Wallace [29] independently showed that every 3-manifold can be smoothly embedded in a 4-manifold. Moreover, a generalization of their approach reveals that any 3-manifold can be smoothly embedded in the connected sum of $S^2 \times S^2$ copies. Freedman's work [12] established that all integer homology 3-spheres can be topologically and locally flatly embedded in S^4. However, the Rokhlin invariant μ and Donaldson's diagonalization theorem [6] demonstrate that certain integer homology spheres cannot be smoothly embedded in S^4. This leads to the question: can a compact 4-manifold exist wherein all 3-manifolds embed? Shiomi [28] provided a negative answer to this inquiry. Aceto, Golla, and Larson [1] delved into embedding 3-manifolds in spin 4-manifolds. Among the diverse class of 4-manifolds, symplectic manifolds hold particular interest. In [10], Etnyre, Min, and the author conjectured the following.

Conjecture 1 *Every closed, oriented smooth 3-manifold smoothly embeds in a symplectic 4-manifold.*

For example, notice that if Y is obtained by doing integer surgery on a knot in S^3 then Y can have a smooth oriented embedding in $\mathbb{CP}^2 \# \overline{\mathbb{CP}}^2$ or $\mathbb{CP}^2 \#_2 \overline{\mathbb{CP}}^2$ depending on whether the surgery coefficient is odd or even. There does not seem to be an analogous result for links.

© 2025 The Author, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

1.1 Embeddings in symplectic manifolds

While we cannot resolve the above conjecture, we can show the existence of topological embeddings and smooth embeddings after stabilization.

Theorem 1.1 *Given a closed, connected, oriented 3-manifold Y, there exists a simply connected symplectic closed 4-manifold X such that Y can be embedded topologically, locally flatly in X. This embedding can be made a smooth embedding after one stabilization, ie Y can smoothly embed in $X \# (S^2 \times S^2)$.*

The above Theorem 1.1 indicates that no obstructions are coming from the topological category; however, when smooth embeddings in symplectic manifolds do exist, they can imply interesting things about topology.

When considering the smooth embeddings of a 3-manifold Y into a smooth 4-manifold X, an ambiguity arises. As an oriented manifold, Y differs significantly from its orientation reversal $-Y$: for instance, the Poincaré homology sphere with a positive orientation bounds a negative-definite 4-manifold, whereas its counterpart with a negative orientation does not. Conversely, if Y smoothly embeds in X, then the boundary of a small neighborhood of Y is $-Y \sqcup Y$, implying that $-Y$ also smoothly embeds in X. However, this ambiguity can be resolved using cobordisms.

Definition 1.2 A smooth embedding of Y in a cobordism W from Y_0 to Y_1 is termed an *oriented cobordism embedding* if Y is either nonseparating or if Y separates W into $W_1 \sqcup W_2$ such that Y serves as an oriented manifold boundary component of W_1, with all other components of ∂W_1 (if any) belonging to Y_0.

Recall that a closed oriented rational homology sphere is called an *L-space* if its Heegaard Floer homology group is "as simple as possible", as specified in Section 2.3. Embeddings of L-spaces in symplectic manifolds are constrained as follows.

Theorem 1.3 *If an L-space Y smoothly embeds in a closed symplectic 4-manifold X, it must be separating. Furthermore, if $X = X_1 \cup_Y X_2$, then one of the X_i must be a negative-definite 4-manifold. Therefore, if an L-space Y does not bound a negative-definite 4-manifold, it cannot possess an oriented cobordism embedding in any symplectic 4-manifold with a weakly convex boundary.*

Remark 1.4 Ozsváth and Szabó [25] have proven the aforementioned result for separating L-spaces in closed symplectic manifolds. Therefore, the key contribution of the theorem above is to demonstrate that L-spaces cannot be embedded as nonseparating hypersurfaces in symplectic manifolds. Our proof of this result was inspired by Agol and Lin's work on hyperbolic 4-manifolds [2].

Remark 1.5 There exist numerous such L-spaces, including the Poincaré homology sphere with negative orientation and r-surgery for $r \in [9, 15)$ on the pretzel knot $P(-2, 3, 7)$ in S^3 [17; 18] (the latter L-spaces

are hyperbolic). Prior research had already established that these manifolds are not Stein fillable [17; 18; 20]. Here, we have demonstrated that beyond not being weakly fillable, they cannot even possess a smooth oriented cobordism embedding in any weak filling of any 3-manifold.

Corollary 1.6 *If Y' admits a weakly fillable contact structure then any L-space Y which does not bound negative-definite 4-manifolds cannot have any smooth oriented cobordism embedding in $Y' \times I$.*

Theorem 1.3 gives rise to a very interesting question.

Question 1 *Does every L-space bound a definite 4-manifold?*

Remark 1.7 Notice that if Conjecture 1 is true then the above question has a positive answer.

In Section 4, following the proof of Theorem 1.3, we will explore a strategy outlined in Remark 4.1 aimed at providing a negative response to Question 1. Such a response would, in turn, furnish a counterexample to the conjecture positing that closed 3-manifolds can be smoothly embedded in symplectic 4-manifolds.

1.2 Cobordisms and symplectic structures

We call a closed oriented 3-manifold *Stein fillable* if it admits a Stein fillable contact structure, the definition of which is postponed until Section 2.2. Not every smooth filling X of a 3-manifold Y can be equipped with a Stein structure. To illustrate this, note that if a $\mathbb{Z}HS^3$ bounds a $\mathbb{Q}HB^4$ that is not a $\mathbb{Z}HB^4$, then every handle decomposition must possess a 3-handle. However, Eliashberg [7] demonstrated that any Stein decomposition cannot include a 3-handle. For instance, Fintushel and Stern [11] established that $\Sigma(2,3,7)$ does not bound a $\mathbb{Z}HB^4$ but does bound a $\mathbb{Q}HB^4$. Therefore, this rational ball cannot serve as a Stein filling of $\Sigma(2,3,7)$ or, indeed, any 3-manifold homology cobordant to $\Sigma(2,3,7)$ cannot possess a rational ball Stein filling. Moreover, certain 3-manifolds do not admit any Stein fillings; see Lisca [20]. Hence, two questions naturally arise:

Question 2 *Is every 3-manifold integral homology cobordant to a Stein fillable 3-manifold?*

Question 3 *If X is an oriented smooth compact 4-manifold with boundary Y, does there exist a 3-manifold Y' which is rational homology cobordant to Y and has a symplectic filling X' with the same algebraic topology as X?*

We will address these two questions in the following theorem. First, let's recall that we refer to an integral homology cobordism from Y_0 to Y_1 as a *$\mathbb{Z}$-ribbon cobordism* if this integer homology cobordism is achieved by attaching handles of index only 1 and 2 to $Y_0 \times [0,1]$ along $Y_0 \times 1$. We denote such a cobordism by saying Y_0 is *ribbon cobordant* to Y_1. It is important to note that this relation is a partial ordering on 3-manifolds and not necessarily a symmetric relation. We also refer to such a cobordism W as an *invertible cobordism* if there is a cobordism W' from Y_1 to Y_0 such that $W \cup W' = Y_0 \times I$. (Similarly, we can define $\mathbb{Q}$-ribbon homology cobordism.)

Theorem 1.8 *Given any closed oriented 3-manifold Y there exists a Stein fillable 3-manifold Y' and a $\mathbb{Z}$-ribbon invertible homology cobordism W from Y to Y' which is obtained from $Y \times [0,1]$ by attaching a single pair of algebraically canceling 1- and 2-handle. Moreover if X is an oriented compact 4-manifold with connected boundary $\partial X = Y$, then:*

(i) *If $b_1(X) = 0$, there exists a Stein 4-manifold X' with boundary $\partial X' = Y'$ such that there is a $\mathbb{Q}$-ribbon homology cobordism from Y to Y' and $b_2(X) = b_2(X')$.*

(ii) *If $\partial X = Y$ is a $\mathbb{Q}HS^3$, there exists a Stein 4-manifold X' with boundary $\partial X' = Y'$ such that the intersection form of X is isomorphic to the intersection form of X' and there is a $\mathbb{Q}$-ribbon homology cobordism from Y to Y'.*

Remark 1.9 Yasui has pointed out to the author that, although they do not specifically discuss cobordisms, the first part of this result, namely constructing W from Y to Y', without the restriction of only needing a single 1- and 2-handle pair, can be proven by assembling various results from [5]. One of the primary techniques for proving this result involved utilizing embedded cork and cork twists, which theoretically can transform non-Stein manifolds into Stein ones. These ideas had been previously employed in the works of Akbulut and Matveyev [4] and Akbulut and Yasui [5]. The nontrivial aspect of the aforementioned result lies in how to handle the 3-handles in such cobordisms.

In low-dimensional topology, the investigation of the integer homology cobordism group $\Theta^3_{\mathbb{Z}}$ and rational homology cobordism group $\Theta^3_{\mathbb{Q}}$ holds special significance. The aforementioned result provides a new generator for these groups.

Corollary 1.10 *The homology cobordism groups $\Theta^3_{\mathbb{Z}}$ and $\Theta^3_{\mathbb{Q}}$ are generated by Stein fillable 3-manifolds.*

Remark 1.11 It remains unknown whether $\Theta^3_{\mathbb{Q}}$ is generated by L-spaces. Nozaki, Sato and Taniguchi [21] demonstrated that $\Sigma(2,3,11) \#_2 (-\Sigma(2,3,5))$ does not bound a definite 4-manifold. If we can discover an L-space Y that is rationally cobordant to $\Sigma(2,3,11)$, then $Y \#_2 (-\Sigma(2,3,5))$ cannot bound a definite 4-manifold. As $Y \#_2 (-\Sigma(2,3,5))$ is itself an L-space, Theorem 1.3 asserts that this manifold cannot be smoothly embedded in any symplectic 4-manifold. Conversely, if all 3-manifolds can be embedded in some symplectic 4-manifold, then $\Theta^3\mathbb{Q}$ is not generated by L-spaces. Thus, we arrive at the following question:

Question 4 *Is $\Sigma(2,3,11)$ rationally cobordant to some L-space?*

For a closed oriented 3-manifold Y, $H_3(Y;\mathbb{Z})$ is canonically isomorphic to $\mathbb{Z}$. Thus, a map $f: Y_0 \to Y_1$ induces a homomorphism on the top-dimensional homology group, $f: \mathbb{Z} \to \mathbb{Z}$. The degree of f is denoted by $f(1) \in \mathbb{Z}$. It has been established, through the work of Ruberman [27], that all 3-manifolds are the target of a degree one map from a hyperbolic 3-manifold. By employing a similar line of reasoning, we can assert the following:

Corollary 1.12 *Given any 3-manifold Y there exists a Stein fillable 3-manifold Y' and a degree one map $f : Y' \to Y$.*

The distinction between smooth and topological embeddings serves as a tool for detecting exotic structures on compact manifolds. If we encounter two homeomorphic 4-manifolds such that a 3-manifold embeds smoothly in one but not the other, then they are not diffeomorphic; they form an exotic pair. Corollary 1.13, which we will present shortly, was initially demonstrated by Akbulut [3] and subsequently proven by many others. However, we will offer an alternative proof stemming from the study of embeddings of 3-manifolds into 4-manifolds.

Corollary 1.13 *There exist compact 4-manifolds with boundary X and X' such that $b_2(X) = b_2(X') = 1$ that are homeomorphic but not diffeomorphic.*

Acknowledgements The author expresses gratitude to Chris Gerig, Robert Gompf, Kyle Hayden, Tye Lidman, Hyunki Min, Maggie Miller, Lisa Piccirillo, Danny Ruberman, Oğuz Şavk, and Bülent Tosun for their insightful discussions and feedback on the first draft. Special thanks are extended to Jen Hom and Tom Mark for their assistance with aspects of Heegaard Floer homology. The author acknowledges Kouchi Yasui for bringing the paper [5] to their attention and for engaging in helpful discussions. Marco Golla's inquiry, which led to Theorem 1.8, is appreciated, along with his thorough review of earlier drafts and constructive comments. Lastly, the author expresses gratitude to their advisor John Etnyre for constant guidance and support, as well as to the referees for their valuable feedback.

2 Background

2.1 Contact geometry

Recall that a (*coorientable*) *contact structure* ξ on an oriented 3-manifold Y is the kernel of 1-form $\alpha \in \Omega^1(Y)$ such that $\alpha \wedge d\alpha$ is nondegenerate. Geometrically a contact structure on a 3-manifold is a distribution of 2-plane fields on the manifold that is not tangent to any embedded surface in the manifold. Darboux's theorem says that every contact 3-manifold (Y, ξ) is locally contactomorphic to $(\mathbb{R}^3, \xi_{\text{std}} = \ker(dz - ydx))$. All orientable 3-manifolds admit contact structures. A knot $L \subset (Y, \xi)$ is called *Legendrian* if at every point of L the tangent line to L lies in the contact plane at that point. A Legendrian knot L in a contact manifold (Y, ξ) has a standard neighborhood N and a framing fr_ξ given by the contact planes. If L is null-homologous then fr_ξ relative to the Seifert framing is the *Thurston–Bennequin* invariant of L. Suppose one does $fr_\xi - 1$ surgery on L by removing N and gluing back a solid torus to effect the desired surgery. In that case, there is a unique way to extend $\xi|_{Y-N}$ over the surgery torus so that it is tight on the surgery torus. The resulting contact manifold is said to be obtained from (Y, ξ) by *Legendrian surgery* on L.

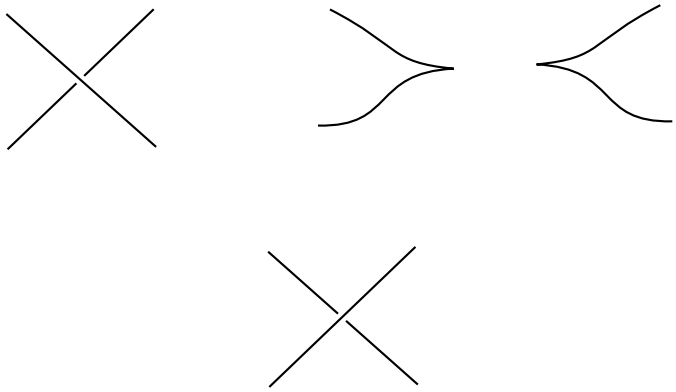


Figure 1: The top row indicates the correct crossing and the cusps in the front projection. The bottom picture crossing will not occur in the front projection diagram.

A Legendrian knot L in $(\mathbb{R}^3, \xi_{\text{std}})$ projects to a closed curve γ in the xz-plane which also known as *front projection* of L. The curve γ uniquely determines the Legendrian knot L which can be reconstructed by setting $y(t)$ as the slope of $\gamma(t)$. Thus, at a crossing of the projection, the most negatively sloped curve always stays at the front. There are two types of cusps singularity possible when $dz/dx = 0$ which are called left cusps and right cusps. See Figure 1.

Looking at an oriented front projection one can compute the Thurston–Bennequin invariant of a Legendrian knot $\text{tb}(K) = \text{writhe}(K) - \#\{\text{left cusps}\}$. For more details, we refer to [13; 14].

A contact 3-manifold (Y, ξ) is called *overtwisted* if there exists a Legendrian unknot with Thurston–Bennequin number 0, otherwise it is called *tight*.

2.2 Symplectic fillings, cobordisms and caps

We recall that a compact symplectic manifold (X, ω) is a *strong symplectic filling* of (Y, ξ) if $\partial X = Y$ and there is a vector field v defined near ∂X such that the Lie derivative of ω satisfies $\mathcal{L}_v \omega = \omega$, v points out of X and $\iota_v \omega$ is a contact form for ξ. Moreover, (X, ω) is a strong symplectic cap for (Y, ξ) if it satisfies all the properties above, except $\partial X = -Y$, and v points into X. We also say (X, ω) is a *weak filling* of (Y, ξ) if $\partial X = Y$ and $\omega|_\xi > 0$ (here all our contact structures are cooriented). Similarly, (X, ω) is a weak cap of (Y, ξ) if $\partial X = -Y$ and $\omega|_\xi > 0$. We shall say that (Y, ξ) is (strongly or weakly) *semifillable* if there is a connected (strong or weak) filling (X, ω) whose boundary is a disjoint union of (Y, ξ) with an arbitrary nonempty contact manifold.

A *symplectic cobordism* from the contact manifold (Y_-, ξ_-) to (Y_+, ξ_+) is a compact symplectic manifold (W, ω) with boundary $-Y_- \cup Y_+$ where Y_- is a *concave* boundary component and Y_+ is *convex*; this means that there is a vector field v near ∂W which points transversally inwards at Y_- and transversally outwards at Y_+, and $\mathcal{L}_v \omega = \omega$. The first result we will need concerns when symplectic cobordisms can be glued together.

Lemma 2.1 [8] *Let (X_i, ω_i) be a symplectic cobordism from (Y_i^-, ξ_i^-) to (Y_i^+, ξ_i^+), for $i = 1, 2$, and (Y_1^+, ξ_1^+) is contactomorphic to (Y_2^-, ξ_2^-). Then we can construct a symplectic cobordism (X, ω) from (Y_1^-, ξ_1^-) to (Y_2^+, ξ_2^+) such that X is diffeomorphic to $X_1 \cup_{Y_1^+} X_2$.*

Now recall that a *Stein domain* is a triple (X, J, ψ) where J is a complex structure on X and $\psi : X \to \mathbb{R}$ is a proper plurisubharmonic function, that is a smooth function such that $\psi^{-1}(-\infty, c]$ is compact for all $c \in \mathbb{R}$ and $\omega_\psi(v, w) = -d(d\psi \circ J)(v, w)$ is a symplectic form. A closed contact manifold (Y, ξ) is called Stein fillable if there exists a Stein manifold (X, J, ψ) such that ψ is bounded from below, M is an inverse image of a regular value of ψ and $\xi = \ker(-d\psi \circ J)$. We have the following characterization of Stein 4-manifolds.

Theorem 2.2 (Elaishberg [7] and Gompf [14]) *A 4-manifold is a Stein domain if and only if it has a handle decomposition with 0-handles, 1-handles and 2-handles, and the 2-handles are attached along Legendrian knots in $\#S^1 \times S^2$ with framing one less than the contact framing.*

Another way to build cobordisms is by Weinstein handle attachment [30]. One may attach a 0-, 1-, or 2-handle to the convex end of a symplectic cobordism to get a new symplectic cobordism with the new convex end described as follows. For a 0-handle attachment, one merely forms the disjoint union with a standard 4-ball and so the new convex boundary will be the old boundary disjoint union with the standard contact structure on S^3. For a 1-handle attachment, the convex boundary undergoes a, possibly internal, connected sum. A 2-handle is attached along a Legendrian knot L with framing one less than the contact framing, and the convex boundary undergoes a Legendrian surgery.

Theorem 2.3 *Given a contact 3-manifold (Y, ξ) let W be a part of its symplectization, that is*

$$(W = Y \times [0, 1], \omega = d(\alpha e^t)).$$

Let L be a Legendrian knot in (Y, ξ) where we think of Y as $Y \times \{1\}$. If W' is obtained from W by attaching a 2-handle along L with framing one less than the contact framing, then the upper boundary (Y', ξ') is still a convex boundary. Moreover, if the 2-handle is attached to a Stein filling (respectively strong, weak filling) of (Y, ξ) then the resultant manifold would be a Stein filling (respectively strong, weak filling) of $(Y'\xi')$.

The theorem for Stein fillings was proven by Eliashberg [7], for strong fillings by Weinstein [30], and was first stated for weak fillings by Etnyre and Honda [9].

Starting with a Stein filling (respectively strong, weak filling) of (Y, ξ) one can construct a symplectic closed manifold by capping it off. Various people have studied concave caps on contact manifolds but for our purpose, we need the result of Etnyre, Min, and the author [10].

Theorem 2.4 *If (W, ω) is weak filling of (Y, ξ) then there exists a closed symplectic 4-manifold (X, ω') in which (W, ω) symplectically embeds such that the complement of W in X is simply connected and has $b_2^+ > 0$.*

2.3 Heegaard Floer homology

Recall that Heegaard Floer homology is an Abelian group associated to a 3-manifold Y, equipped with a Spin^c structure $\mathfrak{t} \in \mathrm{Spin}^c(Y)$. These homology groups are invariant of the pair $(Y, \mathfrak{t})$ and are denoted by $HF^\infty(Y, \mathfrak{t})$, which is a graded $\mathbb{Z}[U, U^{-1}]$ module; $HF^+(Y, \mathfrak{t})$, which is a graded $\mathbb{Z}[U^{-1}]$ module; $HF^-(Y, \mathfrak{t})$, which is a graded $\mathbb{Z}[U]$ module. These invariants fit into a long exact sequence

$$\cdots \longrightarrow HF^-(Y, \mathfrak{t}) \xrightarrow{\iota} HF^\infty(Y, \mathfrak{t}) \xrightarrow{\pi} HF^+(Y, \mathfrak{t}) \xrightarrow{\delta} \cdots .$$

Recall that associated to this long exact sequence there is another 3-manifold invariant

$$HF^+_{\mathrm{red}}(Y, \mathfrak{t}) = \mathrm{Coker}(\pi) \cong \mathrm{Ker}(\iota) = HF^-_{\mathrm{red}}(Y, \mathfrak{t}).$$

The isomorphism in the middle is induced by the coboundary map. Recall that $d(Y, \mathfrak{t})$ is the minimum grading of the torsion-free elements in the image of $\pi\colon HF^\infty(Y, \mathfrak{t}) \to HF^+(Y, \mathfrak{t})$. For more details, readers are referred to [23; 24].

Now recall that an *L-space* Y is a rational homology 3-sphere whose Heegaard Floer homology is as simple as possible, that is $HF^+_{\mathrm{red}}(Y, \mathfrak{t}) = 0$ for all Spin^c structures $\mathfrak{t} \in \mathrm{Spin}^c(Y)$.

A cobordism between two 3-manifolds induces a map on Heegaard Floer homology. More precisely, if W is a cobordism from Y_0 to Y_1 and $\mathfrak{s}$ is a Spin^c structure in W whose restriction on Y_i is denoted by $\mathfrak{s}_i$ for $i = 0, 1$, then there is a map $F^\circ_{W,\mathfrak{s}}\colon HF^\circ(Y_0, \mathfrak{s}_0) \to HF^\circ(Y_1, \mathfrak{s}_2)$, where $\circ = +, -$ or ∞.

Theorem 2.5 (Ozsváth and Szabó [26]) *If W is a cobordism between Y_0 and Y_1, and $\mathfrak{s}$ is a* Spin^c *structure on W whose restriction to Y_i is denoted by $\mathfrak{s}_i$ for $i = 0, 1$ then we have*

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & HF^-(Y_0, \mathfrak{s}_0) & \xrightarrow{\iota_0} & HF^\infty(Y_0, \mathfrak{s}_0) & \xrightarrow{\pi_0} & HF^+(Y_0, \mathfrak{s}_0) & \xrightarrow{\delta_0} & \cdots \\
 & & \downarrow{\scriptstyle F^-_{W,\mathfrak{s}}} & & \downarrow{\scriptstyle F^\infty_{W,\mathfrak{s}}} & & \downarrow{\scriptstyle F^+_{W,\mathfrak{s}}} & & \\
\cdots & \longrightarrow & HF^-(Y_1, \mathfrak{s}_1) & \xrightarrow{\iota_1} & HF^\infty(Y_1, \mathfrak{s}_1) & \xrightarrow{\pi_1} & HF^+(Y_1, \mathfrak{s}_1) & \xrightarrow{\delta_1} & \cdots
\end{array}$$

where the vertical maps are uniquely determined up to an overall sign, and all the squares are commutative.

The *composition law* states that if W_0 is a cobordism from Y_0 to Y_1 and W_1 is a cobordism from Y_1 to Y_2, and if $\mathfrak{s}_i$ is the Spin^c structure on W_i for $i = 0, 1$, then the relationship between composition of $F_{W_0,\mathfrak{s}_0}$ with $F_{W_1,\mathfrak{s}_1}$ and the maps induced by the composite cobordism $W = W_0 \cup_{Y_1} W_1$ is

$$F^\circ_{W_1,\mathfrak{s}_1} \circ F^\circ_{W_0,\mathfrak{s}_0} = \sum_{\{\mathfrak{s} \in \mathrm{Spin}^c(W) \,|\, \mathfrak{s}|W_i = \mathfrak{s}_i, i = 0,1\}} \pm F^\circ_{W,\mathfrak{s}}.$$

2.4 Closed 4-manifold invariants

There is a variant of the cobordism invariant which is defined for cobordism with $b^+_2(W) > 1$. The following lemma was proved by Ozsváth and Szabó [26]

Lemma 2.6 *Let W be a cobordism between Y_0 and Y_1 with $b_2^+(W) > 0$. Then the induced cobordism map $F^{\infty}_{W,\mathfrak{s}}$ vanishes for all* Spin^c *structures on W.*

If we have a cobordism W with $b_2^+(W) > 1$, then we can cut W along a 3-manifold N, which divides W into two cobordisms, W_0 and W_1, both of which have $b_2^+(W_i) > 0$, in such a way that the map induced by the restriction

$$\text{Spin}^c(W) \to \text{Spin}^c(W_0) \times \text{Spin}^c(W_1)$$

is injective. Such a cut N is called an *admissible cut*.

Remark 2.7 Notice that if in a cobordism W with $b_2^+(W) > 1$ we find a separating rational homology 3-sphere N such that both the pieces have $b_2^+ > 0$, then N is an admissible cut.

If $\mathfrak{s}$ is a Spin^c structure on W whose restriction to W_i is $\mathfrak{s}_i$ and the induced Spin^c structures on 3-manifolds Y_0, Y_1 and N is $\mathfrak{t}_0$, $\mathfrak{t}_1$ and $\mathfrak{t}$, then

$$F^-_{W_0,\mathfrak{s}_0}\colon HF^-(Y_0,\mathfrak{t}_0) \to HF^-(N,\mathfrak{t})$$

factors through the inclusion $HF^-_{\text{red}}(N,\mathfrak{t}) \to HF^-(N,\mathfrak{t})$, and

$$F^+_{W_1,\mathfrak{s}_1}\colon HF^+(N,\mathfrak{t}) \to HF^+(Y_1,\mathfrak{t}_1)$$

factors through the projection $HF^+(N,\mathfrak{t}) \to HF^+_{\text{red}}(N,\mathfrak{t})$. And thus by using the identification of $HF^+_{\text{red}}(N,\mathfrak{t}) \cong HF^-_{\text{red}}(N,\mathfrak{t})$ in the middle, we can define the *mixed invariant* as a map

$$F^{\text{mix}}_{W,\mathfrak{s}}\colon HF^-(Y_0,\mathfrak{t}_0) \to HF^+(Y_1,\mathfrak{t}_1).$$

Remark 2.8 It is also proven in [26] that F^{mix} does not depend on the choice of the admissible cut.

From the discussion above one immediately sees the following result.

Lemma 2.9 *If an admissible cut N of W is an L-space then $F^{\text{mix}}_{W,\mathfrak{s}}$ vanishes.*

Theorem 2.10 (Ozsváth and Szabó [25]) *If (X,ω) is a closed, symplectic manifold with $b_2^+(X) > 1$, then for the canonical* Spin^c *structure $\mathfrak{k}$ corresponding to the symplectic form, $F^{\text{mix}}_{X,\mathfrak{k}}$ is nonvanishing. Here we think of X as a cobordism from S^3 to S^3 by taking out two 4-balls.*

Remark 2.11 The above discussion implies that an L-space cannot be an admissible cut for a closed symplectic 4-manifold with $b_2^+ > 1$.

3 Topological embedding of 3-manifolds in symplectic 4-manifolds

Now we will begin the proof of topological embedding of 3-manifolds into symplectic 4-manifolds.

Proof of Theorem 1.1 We will topologically embed a 3-manifold Y into a symplectic manifold X in three steps. In the fourth step, we will show that the embedding is smooth after a single stabilization with $S^2 \times S^2$. We start with a Kirby picture, consisting of only a 0-handle and 2-handles, for a 4-manifold whose boundary is Y.

Step 1 Stein modification of the Kirby picture.

Let $K_1, \dots, K_m$ be the attaching spheres for the 2-handles. We can Legendrian realize the K_i so that each K_i intersects a fixed Darboux ball B in a horizontal arc. We can blow up meridians to each K_i so that the framing on K_i is less than $\mathrm{tb}(K_i) - 2$. All the blown-up unknots can be gathered in the Darboux ball as shown in the top left of Figure 2. Blow up one more unknot as indicated in the upper right of Figure 2 and notice that the resulting link L can be Legendrian realized as in the bottom diagram of Figure 2. Let K be the unknot with framing -1 in the figure. We now stabilize the components of $L - K$ so that the Thurston–Bennequin invariant of each component is one larger than its surgery coefficients. Legendrian surgery on $L - K$ together with -1 surgery on K gives our manifold Y. (Notice that to realize L each K_i might need to be stabilized one extra time as shown in the figure. This is why we arranged the surgery coefficients to be less than $\mathrm{tb}(K_i) - 2$.) So the manifold W_0 obtained by attaching Stein 2-handles to $L - K$ is a Stein manifold and we denote the boundary by Y_0. Now attaching a 2-handle to W_0 along K with framing -1 gives a 4-manifold W with boundary Y.

Step 2 Attach a cork and apply a cork twist.

This similar technique of using cork-twist to produce Stein manifold has been previously used in the work of Akbulut and Yasui [5] and Akbulut and Matveyev [4]. We adopt this method in our setup by constructing a manifold W_2 by modifying the surgery presentation that is adding the 1- and 2-handle shown in Figure 3, where the 2-handle links K as indicated and is otherwise disjoint from L (by abuse of language we will call this operation attaching a Mazur cork). Said another way, we can build a cobordism W_1 by attaching the 1- and 2-handle to $Y \times [0, 1]$ along $Y \times \{1\}$. The manifold W_2 is now simply $W \cup W_1$ with ∂W glued to $-Y \subset W_1$ and W_1 is a cobordism from Y to some manifold Y'.

We can apply a cork-twist by interchanging the 1-handle and the 0-framed 2-handle. The cork twist does not change the boundary 3-manifold Y'. After the cork twists the knot K passes over the 1-handle geometrically once and thus they cancel each other. After this handle cancellation the knot K in the original picture, Figure 2, is replaced by -1 framed knot in the third picture of Figure 3. Notice that this new knot can be realized by a Legendrian knot that has Thurston–Bennequin invariant $+1$ and thus -1 smooth surgery on this knot can be realized as Legendrian surgery on stabilization of the knot. So we get a Stein filling of the boundary.

Step 3 Construct a simply connected closed symplectic 4-manifold.

Also, notice that the W_1 deformation retracts onto Y and W is a 2-handlebody, so in particular W_2' is simply connected. So in Step 2 when we do the cork-twist the new manifold W_2' is still homeomorphic to

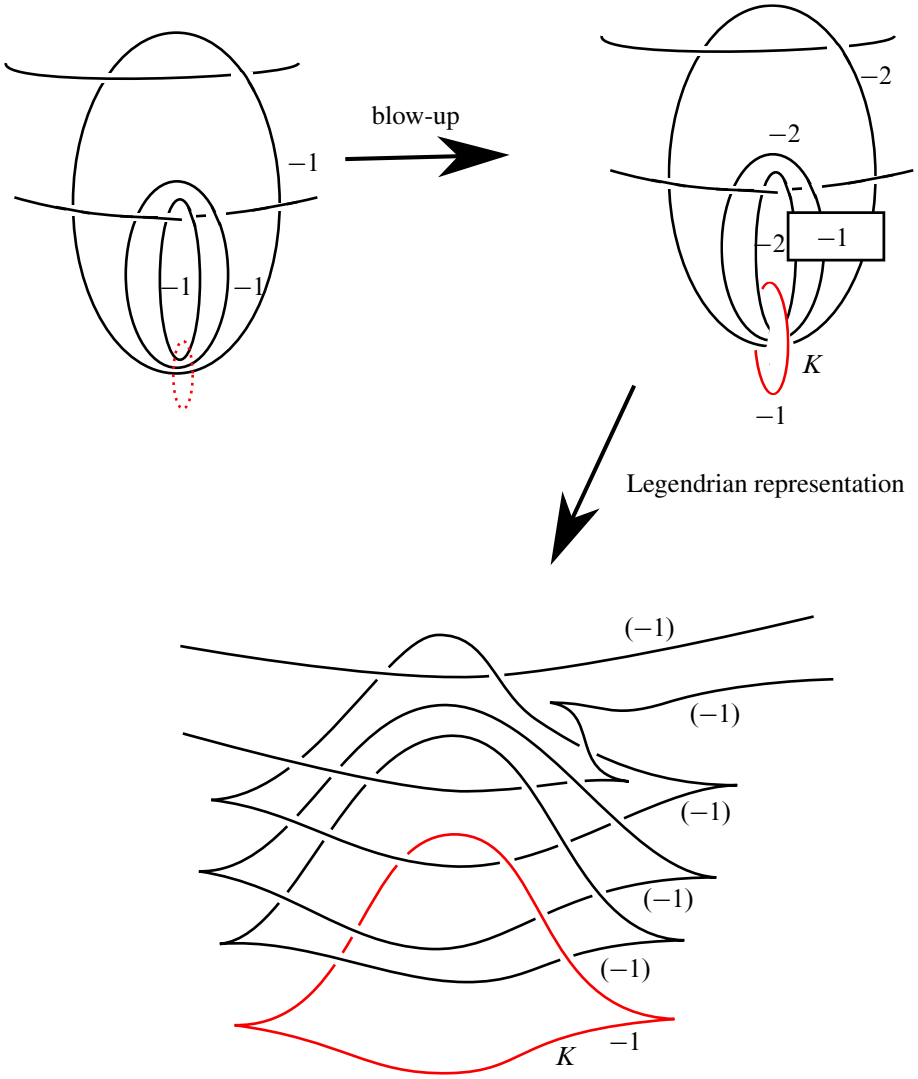


Figure 2: Converting a Kirby picture of the 3-manifold by blowing up such that removing the red knot yields a framed link representing a Stein handle diagram. Here (-1) is measured relative to the contact framing.

the compact manifold W_2 by a result of Freedman [12] and thus the original 3-manifold has a topological, locally flat embedding in W_2'. Now we use the simply connected cap constructed in Theorem 2.4 to cap off the upper boundary W_2' to get a closed simply connected symplectic 4-manifold X into which the 3-manifold Y topologically, locally flatly embeds.

Step 4 Smooth embedding after one stabilization.

We can stabilize X by adding a Hopf link to W_2' and using the same cap (of course this stabilized 4-manifold is no longer symplectic). We can now handle slide one of the components C of the Hopf link

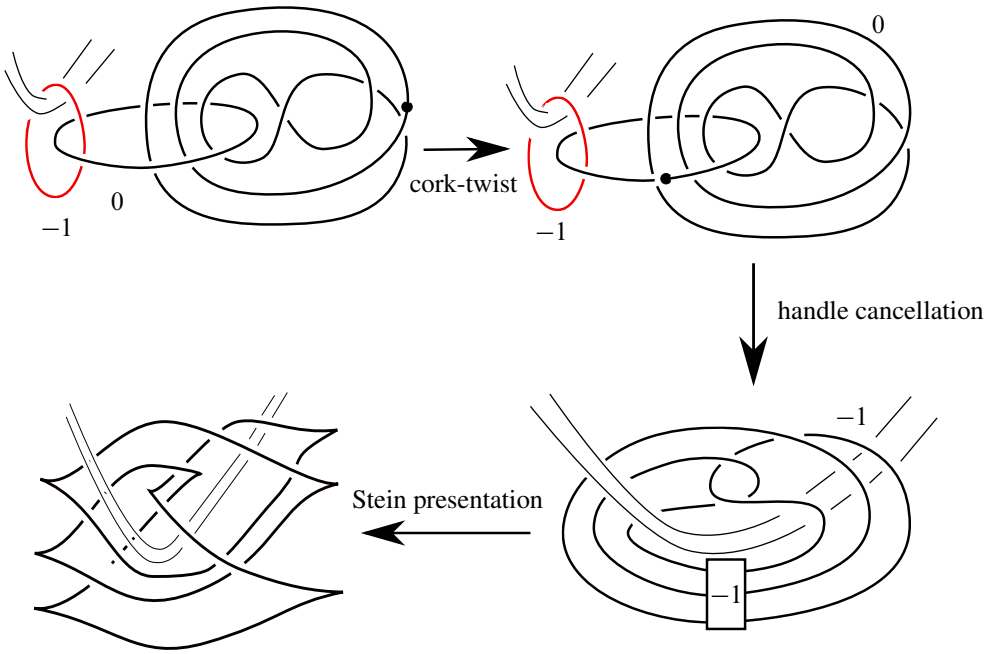


Figure 3: The maximal Thurston–Benequin number of the black knot in the bottom picture is $+1$. So it is a Stein 2-handle attachment.

over the 0-framed knot in the Mazur cork as indicated at the top of Figure 4. Using the 0-framed meridian to C we can untangle C from the 2-handle in the Mazur cork as shown in the middle of Figure 4. We can further slide C over the 0-framed meridian to turn C into a meridian of the 1-handle. Thus the 1-handle can be canceled with C, leaving the bottom picture in Figure 4. Thus we see a smooth embedding of Y into $W_2' \# S^2 \times S^2$ and thus into $X \# S^2 \times S^2$. □

4 Embedding L-spaces in symplectic 4-manifolds

We now prove that smooth embeddings of L-spaces in symplectic 4-manifolds are always separating.

Proof of Theorem 1.3 Suppose Y is an L-space that smoothly embeds in a closed symplectic 4-manifold. We begin by showing it is separating. To this end, we assume it is nonseparating. Let X_1 be the compact manifold obtained from X by cutting along Y. Notice that $\partial X_1 = Y \sqcup -Y$, so we can glue two copies, X_1^1 and X_1^2, of X_1 along their boundaries to get a closed manifold X'. As constructed, X' is a double cover of X so, in particular, we can lift the symplectic form using the covering map, and thus X' is symplectic. Let N be a neighborhood of an arc in X_1^1 connecting its boundary components. Set $X_1' = \overline{X_1^1 - N}$ and $X_2' = X_1^2 \cup N$. Clearly $\partial X_i' = Y \# -Y$ which is an L-space. As X is symplectic and Y is a rational homology sphere, by using the Mayer–Vietoris sequence we can see that $b_2^+(X_i') = b_2^+(X) > 0$ (since X is symplectic, the cohomological element corresponds to the symplectic form produces an element of b_2^+). So $Y \# -Y$ is an admissible cut for a symplectic manifold X' which contradicts Remark 2.11.

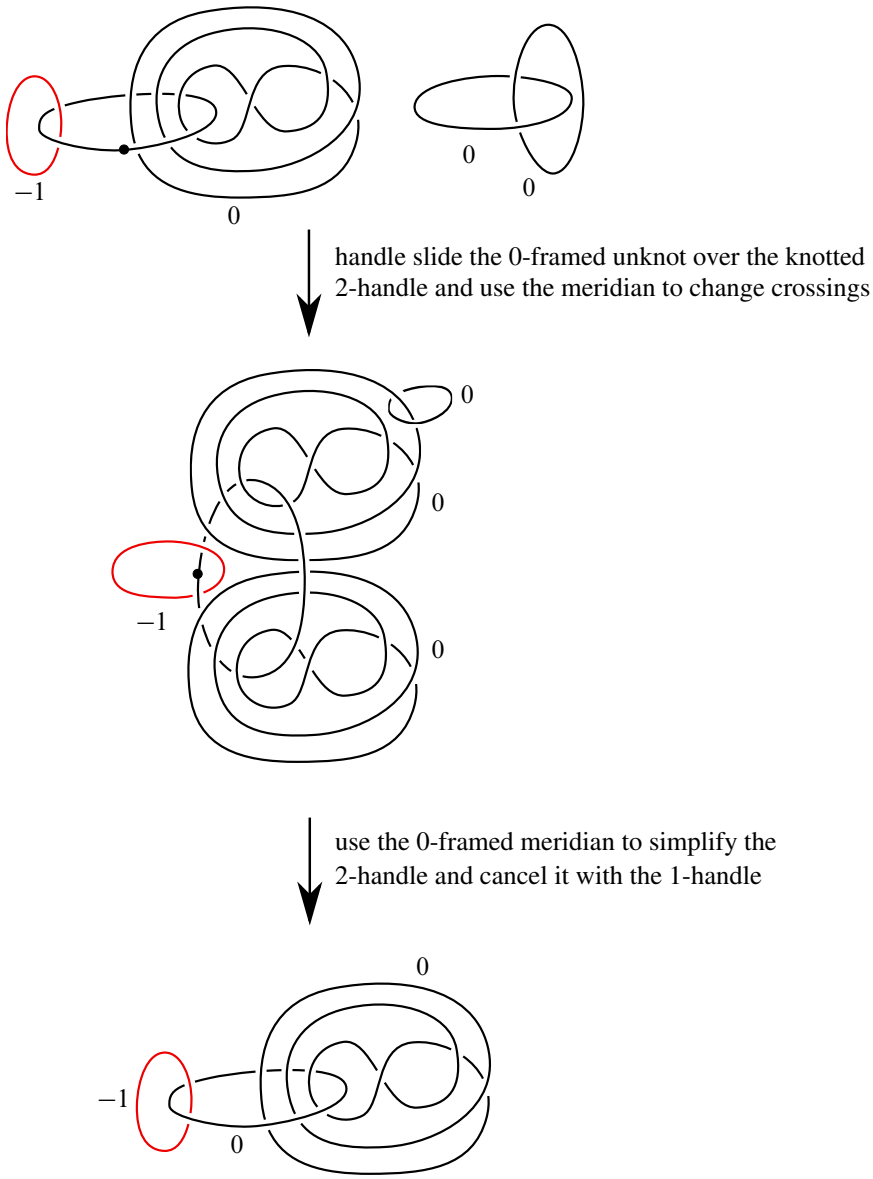


Figure 4: In this picture we are describing the Kirby moves of how connected summing with $S^2 \times S^2$ helps to cancel the 1-handle of the cork.

Now when Y embeds in X in a separating manner then one of the components of $X - Y$ must have $b_2^+ = 0$ or we will get the same contradiction as before.

Now let Y be an L-space that does not bound a negative-definite 4-manifold. If Y is embedded in any symplectic 4-manifold with weakly convex boundary W then it has to be separating since otherwise we can cap off with a concave cap to get a closed symplectic manifold where Y is nonseparating which contradicts the above discussion. So Y has to be separated. When we cap off the upper boundary of

W by a cap with $b_2^+ > 0$, since Y does not bound a negative-definite 4-manifold, both the sides of Y have $b_2^+ > 0$. In particular, Y is an admissible cut for a symplectic 4 manifold with $b_2^+ > 1$ which is a contradiction by Remark 2.11. □

Remark 4.1 We now discuss a strategy to show the negative answer to Question 1 about L-spaces bounding definite 4-manifolds. Before that, notice all lens spaces bound both positive-definite and negative-definite 4-manifolds, because every lens space can be thought of as the boundary of a negative plumbed manifold and $-L(p,q) = L(p, p-q)$. One can obstruct rational homology spheres Y with $H_1(Y,\mathbb{Z}) \neq 0$ from bounding negative-definite manifolds by using the technique developed by Owens and Strle [22] where in Theorem 2 they proved that if the maximum value of the d-invariant of Y is smaller than $1/4$ (with some more algebraic conditions) then Y cannot bound a negative-definite 4 manifold. Now $d(Y,\mathfrak{s}) = -d(-Y,\mathfrak{s})$, so if we find an L-space Y with $H_1(Y,\mathbb{Z}) \neq 0$, for which the absolute differences between d-invariants for different Spin^c structures are very small then that could be used to obstruct it from bounding positive-definite and negative-definite 4-manifolds (as both Y and $-Y$ cannot bound negative-definite 4-manifolds). So we can ask:

Question 5 *For every $n \in \mathbb{N}$ does there exist an L-space which is not a $\mathbb{Z}HS^3$ and whose d-invariant values are in $(-1/n, 1/n)$?*

Proof of Corollary 1.6 Let Y' admit a weakly fillable contact structure and Y be an L-space that does not bound a negative-definite 4-manifold. If Y has an oriented cobordism embedding in $Y' \times [0,1]$, then since Y' is weakly fillable Y has an oriented cobordism embedding in a symplectic 4 manifold with weakly convex boundary, contradicting Theorem 1.3. □

We now show the existence of exotic manifolds with boundary and $b_2 = 1$ using the ideas above.

Proof of Corollary 1.13 Start with B^4 and attach a 2-handle h along the pretzel knot $K = P(-2,3,7)$ to get a 4-manifold W' with $S^3_9(K)$ (which is an L-space as mentioned in Remark 1.5) as its boundary. Attach a cork as in Step 2 of Theorem 1.1 and get W with $b_2^+(W) = 1$. After a cork-twist, we can see that the 2-handle h now passing over the 1-handle of the cork and this will increase the contact framing of h by one as in Figure 7; thus the resulting manifold W' will be Stein by Theorem 2.2. Before the cork-twist, we had a smooth embedding of $S^3_9(K)$ in W. But by Theorem 1.3 $S^3_9(K)$ cannot embed smoothly in W' so they are exotic pairs. □

5 Constructing ribbon cobordism

We now turn to Theorem 1.8 that given any 3-manifold Y there is a simple invertible $\mathbb{Z}$-ribbon homology cobordism to a Stein fillable manifold Y'. Moreover, that says that given a compact 4-manifold with some specific conditions, one can construct a Stein 4-manifold with the same algebraic topology but a different boundary.

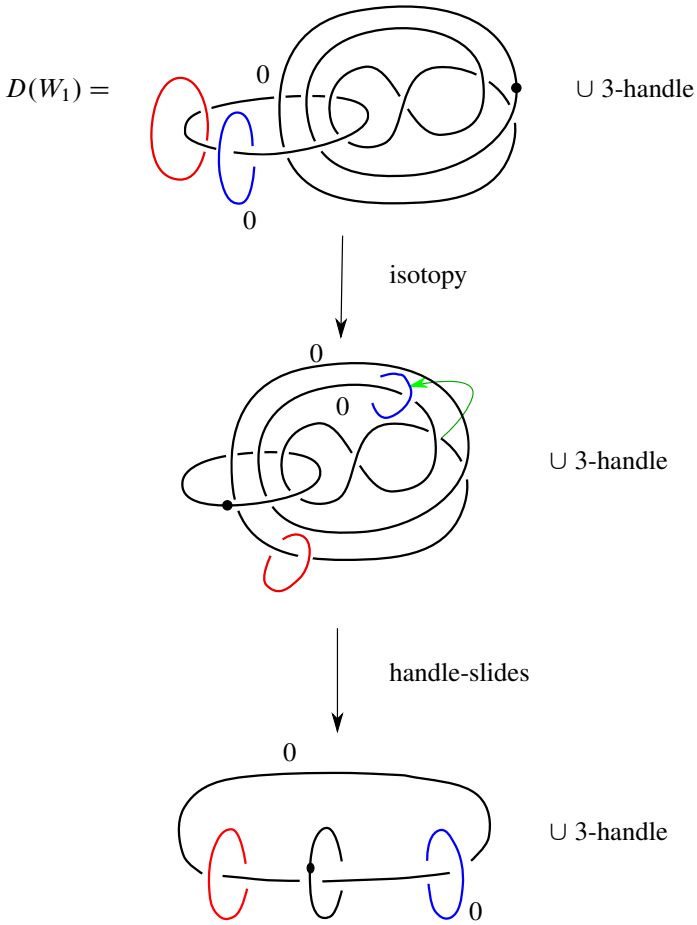


Figure 5: After isotopy, in the second picture, the 0-framed blue 2-handle will help to resolve the crossings of the black 2-handle so that it can cancel the 1-handle. And after that, the 3-handle will cancel the 0-framed unknotted blue 2-handle.

Proof of Theorem 1.8 The cobordism W_1 from Step 2 is the desired ribbon $\mathbb{Z}$-homology invertible cobordism which is attaching a cork along the red knot at the top of Figure 3.

To see that it is invertible, let us consider $D(W_1)$ to be the double of W_1 along Y', that is glue an upside-down copy of W_1 on top of W_1 along the boundary. If h is the 2-handle in W_1, then $D(W_1)$ is formed by attaching a 2- and 3-handle to W_1, with the 2-handle attached to a 0-framed meridian to h [15, Section 4.2]. Check Figure 5. Now by changing crossings on the attaching circle of h using the 0-framed meridian, we can arrange that h passes over the 1-handle geometrically once. Thus they cancel each other. And after the cancellation, the 0-framed meridian h will cancel the 3-handle. And thus the resultant manifold $D(W_1) = Y \times I$.

(i) Let X be a compact oriented 4-manifold with boundary Y and $b_1(X) = 0$. Turning a handle structure on X upside down, we can think of X as a cobordism from $-Y$ to $\varnothing$; this is indicated in Figure 6.

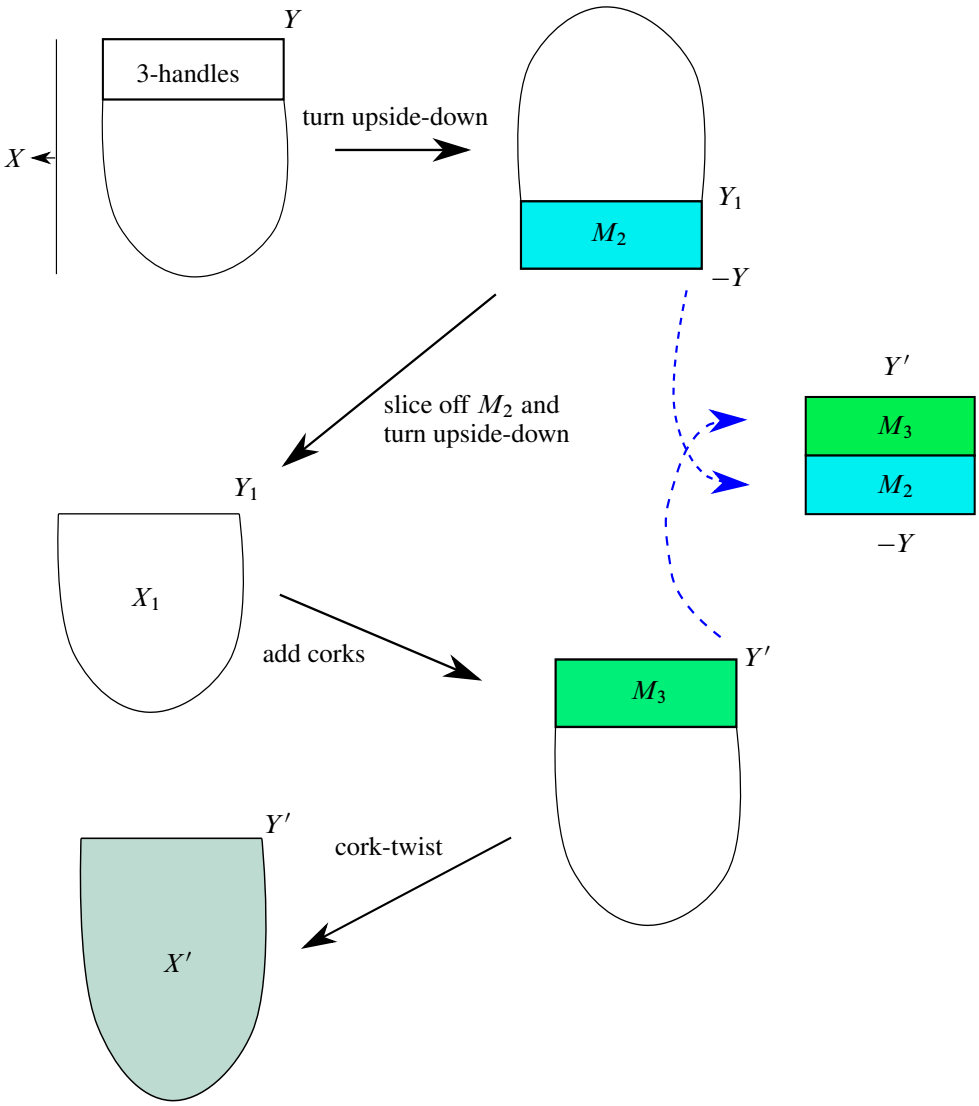


Figure 6: A schematic of the construction of a ribbon cobordism from Y to a Stein fillable Y'.

Notice that in this upside-down cobordism, all the 1-handles of X are converted into 3-handles and all the 3-handles become 1-handles. In the upside-down X, 1-handles are attached onto $-Y \times [0,1]$ along $-Y \times \{1\}$, let us call this cobordism M_1. Notice that $b_1(X) = 0$ so the homology long exact sequence of the pair (X, Y) implies that there exists a minimal set of 2-handles such that if we attach those on top of M_1, and let us call it M_2, then $H_1(M_2, Y; \mathbb{Q})) = 0$. (Here by minimum we mean that if we take any 2-handle out from the set then $H_1(M_2, Y; \mathbb{Q}) \neq 0$.) Since we consider a minimal set of 2-handles for this construction, we have $H_2(M_2, Y; \mathbb{Q}) = 0$ as well because in this case the number of 1-handles of M_2 is the same as the number of 2-handles. Thus M_2 is a rational ribbon cobordism from Y to say Y_1 which is the top boundary of M_2; see the top right of Figure 6. Consider X_1 to be the handlebody

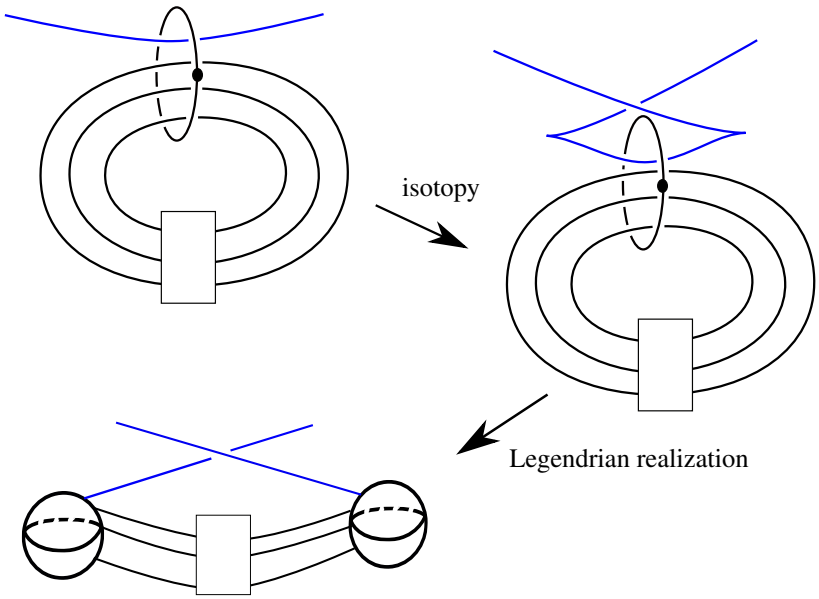


Figure 7: The contact framing of the blue knot increased by $+1$ after a cork-twist.

obtained from X by taking out M_2 and turning what remains upside-down; this is indicated in the third picture Figure 6. Thus X_1 only has 1- and 2-handles with boundary Y_1. If this is Stein then we are done. If not then that implies it has some 2-handles whose smooth framing is bigger than that of the contact framing minus 1 of the attaching circle in $\#S^1 \times S^2$ (in this case we can think of the top boundary Y_1 is obtained after attaching 2-handles on the boundary of 1-handlebody which is connected sum of $S^1 \times S^2$). To fix this framing issue, we repeatedly apply the Step 2 of the proof of Theorem 1.1. That is we attach a cork as in Figure 3 (where the red curve there is the handle that needs its Thurston–Bennequin invariant increased). We then do a cork twist that exchanges the 1-and 2-handles. We claim this increases the contact framing of the original attaching sphere of the 2-handle by 1. To see this, notice that if a knot passes over 1-handle then in the front projection diagram of a knot we are actually deleting two consecutive right and left cusps by connecting them through a 1-handle, and thus we are increasing the contact framing. See Figure 7. However, this process does not change the smooth surgery coefficient. Let us consider the cobordism X_2 obtained by attaching a suitable number of corks to X_1 so that the manifold X_2' obtained by applying the cork twists is Stein. The manifolds X_2 and X_2' are homeomorphic as the cork-twist homeomorphism can always be extended as homeomorphism on the 4-manifold by the result of Freedman [12]. Observe that $b_2(X_2') = b_2(X_1) = b_2(X)$. Let Y' be the top boundary of X_2'. Then there is a homology ribbon cobordism M_3 from Y_1 to Y' which is given by attaching the above corks to the top of X_1; see the fourth picture in Figure 6. Glue this cobordism on top of M_2 to get our desired ribbon rational homology cobordism $M = M_2 \cup M_3$ from Y to Y' with Y' Stein fillable.

(ii) Let X be a compact manifold with connected boundary Y which is a $\mathbb{Q}HS^3$; then we consider a handle decomposition of $X = X_0 \cup X_1 \cup X_2 \cup X_3$ where X_i contains handles of index i. Consider the

minimum set of 1-handles which generate the free part of $(H_1(X;\mathbb{Q}))$. Let $\overline{X}$ be the manifold obtained from X by doing surgery on those 1-handles. (In Kirby calculus this is equivalent to replacing those dotted 1-handles with 0-framed unknotted 2-handles.) We will now show that this surgery operation does not change the b_2 (or more precisely the intersection form). As Y is a $\mathbb{Q}HS^3$, $H_1(\overline{X};\mathbb{Q}) = 0 = H_3(\overline{X};\mathbb{Q})$. But we are not doing anything with the 3-handles of X, so the only way the third homology of $\overline{X}$ vanishes with $\mathbb{Q}$ coefficients is if the 3-handles cancel the 2-handles in homology. And thus from cellular homology, we can see that $b_2(X) = b_2(\overline{X})$. Also, notice that the above surgery does not change the nontorsion elements of $H^2(X;\mathbb{Z})$ so they have the same intersection form. Now apply the proof of Theorem 1.8 on $\overline{X}$ and we get the desired Stein manifold X' with boundary Y'. □

We now prove that the 3-dimensional homology cobordism group is generated by Stein fillable manifolds

Proof of Corollary 1.10 Theorem 1.8 provides a homology cobordism from any manifold to a Stein manifold. Thus the homology cobordism groups are generated by Stein manifolds. □

We now prove the existence of a degree 1 map from a Stein fillable 3-manifold to a given 3-manifold.

Proof of Corollary 1.12 As we noticed previously Y' smoothly embeds in $Y \times I$ in a separating way. So, restriction to Y' of the projection map from $Y \times I$ onto Y will induce a degree 1 map from $Y' \to Y$. □

References

[1] **P Aceto**, **M Golla**, **K Larson**, *Embedding 3-manifolds in spin 4-manifolds*, J. Topol. 10 (2017) 301–323 MR Zbl

[2] **I Agol**, **F Lin**, *Hyperbolic four-manifolds with vanishing Seiberg–Witten invariants*, from "Characters in low-dimensional topology" (O Collin, S Friedl, C Gordon, S Tillmann, L Watson, editors), Contemp. Math. 760, Amer. Math. Soc., Providence, RI (2020) 1–8 MR Zbl

[3] **S Akbulut**, *An exotic 4-manifold*, J. Differential Geom. 33 (1991) 357–361 MR Zbl

[4] **S Akbulut**, **R Matveyev**, *A convex decomposition theorem for 4-manifolds*, Internat. Math. Res. Notices (1998) 371–381 MR Zbl

[5] **S Akbulut**, **K Yasui**, *Cork twisting exotic Stein 4-manifolds*, J. Differential Geom. 93 (2013) 1–36 MR Zbl

[6] **S K Donaldson**, *An application of gauge theory to four-dimensional topology*, J. Differential Geom. 18 (1983) 279–315 MR Zbl

[7] **Y Eliashberg**, *Topological characterization of Stein manifolds of dimension > 2*, Internat. J. Math. 1 (1990) 29–46 MR Zbl

[8] **J B Etnyre**, *Symplectic convexity in low-dimensional topology*, Topology Appl. 88 (1998) 3–25 MR Zbl

[9] **J B Etnyre**, **K Honda**, *Tight contact structures with no symplectic fillings*, Invent. Math. 148 (2002) 609–626 MR Zbl

[10] **J B Etnyre**, **H Min**, **A Mukherjee**, *On 3-manifolds that are boundaries of exotic 4-manifolds*, Trans. Amer. Math. Soc. 375 (2022) 4307–4332 MR Zbl

[11] **R Fintushel**, **R J Stern**, *A μ-invariant one homology 3-sphere that bounds an orientable rational ball*, from "Four-manifold theory" (C Gordon, R Kirby, editors), Contemp. Math. 35, Amer. Math. Soc., Providence, RI (1984) 265–268 MR Zbl

[12] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geometry 17 (1982) 357–453 MR Zbl

[13] **H Geiges**, *An introduction to contact topology*, Cambridge Studies in Advanced Mathematics 109, Cambridge Univ. Press (2008) MR Zbl

[14] **R E Gompf**, *Handlebody construction of Stein surfaces*, Ann. of Math. 148 (1998) 619–693 MR Zbl

[15] **R E Gompf**, **A I Stipsicz**, *4-manifolds and Kirby calculus*, Graduate Studies in Mathematics 20, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[16] **M W Hirsch**, *On imbedding differentiable manifolds in euclidean space*, Ann. of Math. 73 (1961) 566–571 MR Zbl

[17] **A Kaloti**, **B Tosun**, *Hyperbolic rational homology spheres not admitting fillable contact structures*, Math. Res. Lett. 24 (2017) 1693–1705 MR Zbl

[18] **Y Li**, **Y Liu**, *Hyperbolic 3-manifolds admitting no fillable contact structures*, Proc. Amer. Math. Soc. 147 (2019) 351–360 MR Zbl

[19] **W B R Lickorish**, *A representation of orientable combinatorial 3-manifolds*, Ann. of Math. 76 (1962) 531–540 MR Zbl

[20] **P Lisca**, *Symplectic fillings and positive scalar curvature*, Geom. Topol. 2 (1998) 103–116 MR Zbl

[21] **Y Nozaki**, **K Sato**, **M Taniguchi**, *Filtered instanton Floer homology and the homology cobordism group*, J. Eur. Math. Soc. (JEMS) 26 (2024) 4699–4761 MR Zbl

[22] **B Owens**, **S Strle**, *A characterization of the $\mathbb{Z}^n \oplus \mathbb{Z}(\delta)$ lattice and definite nonunimodular intersection forms*, Amer. J. Math. 134 (2012) 891–913 MR Zbl

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

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

[25] **P Ozsváth**, **Z Szabó**, *Holomorphic triangle invariants and the topology of symplectic four-manifolds*, Duke Math. J. 121 (2004) 1–34 MR Zbl

[26] **P Ozsváth**, **Z Szabó**, *Holomorphic triangles and invariants for smooth four-manifolds*, Adv. Math. 202 (2006) 326–400 MR Zbl

[27] **D Ruberman**, *Seifert surfaces of knots in S^4*, Pacific J. Math. 145 (1990) 97–116 MR Zbl

[28] **T Shiomi**, *On imbedding 3-manifolds into 4-manifolds*, Osaka J. Math. 28 (1991) 649–661 MR Zbl

[29] **A H Wallace**, *Modifications and cobounding manifolds, II*, J. Math. Mech. 10 (1961) 773–809 MR Zbl

[30] **A Weinstein**, *Contact surgery and symplectic handlebodies*, Hokkaido Math. J. 20 (1991) 241–251 MR Zbl

[31] **H Whitney**, *The self-intersections of a smooth n-manifold in 2n-space*, Ann. of Math. 45 (1944) 220–246 MR Zbl

Department of Mathematics, Princeton University
Princeton, NJ, United States

anubhavmaths@princeton.edu

Received: 12 March 2022 Revised: 15 March 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3271–3286
DOI: 10.2140/agt.2025.25.3271
Published: 1 October 2025

A note on knot Floer homology of satellite knots with (1, 1)-patterns

WEIZHE SHEN

We prove that if P is a $(1,1)$-pattern, then the two inequalities $\dim \widehat{\mathrm{HFK}}(P(K)) \geqslant \dim \widehat{\mathrm{HFK}}(P(U))$ and $\dim \widehat{\mathrm{HFK}}(P(K)) \geqslant \dim \widehat{\mathrm{HFK}}(K)$ hold for the unknot $U \subset S^3$ and any companion $K \subset S^3$.

57K18

1 Introduction

Knot Floer homology, introduced independently by Ozsváth and Szabó [18] and J Rasmussen [20], is a powerful invariant of knots in the three-sphere. For example, it captures several geometric properties of knots such as genus (see Ozsváth and Szabó [17]) and fiberedness (see Ghiggini [4] and Ni [15]). The theory has several different variants; here we assume the reader is familiar with the hat version, which takes the form of a bigraded finitely generated vector space over the field $\mathbb{F} := \mathbb{Z}/2\mathbb{Z}$:

$$\widehat{\mathrm{HFK}}(K) = \bigoplus_{m,a\in\mathbb{Z}} \widehat{\mathrm{HFK}}_m(K,a).$$

Here K is a knot in S^3, m is the *Maslov* (or *homological*) grading, and a is the *Alexander* grading.

To three-manifolds with parametrized boundary, Lipshitz, Ozsváth, and Thurston [14] associated bordered Heegaard Floer invariants. Moreover, their pairing theorems are well-adapted to the study of the result of gluing two manifolds with torus boundary. Recall that given a knot P (namely, a pattern) embedded in a standard solid torus $S^1 \times D^2 =: V$ and a knot K (namely, a companion) in S^3, the satellite knot $P(K)$ is obtained from P by gluing V to the complement $X_K := \overline{S^3 - \nu(K)}$ — where $\nu(K)$ is a tubular neighborhood of K — in such a way that the meridian of V is identified with the meridian of K, and the longitude of V is identified with the Seifert longitude of K. Therefore, satellite knots can be studied using bordered Heegaard Floer homology. Some early work in this approach includes that of Hom [10], Levine [13], and Petkova [19]. By definition, a pattern is called (1,1) if it admits a genus-one doubly pointed bordered Heegaard diagram. This note concerns $(1,1)$-patterns and satellite knots with such patterns (which hereinafter may be referred to as $(1,1)$-satellites for short). Moreover, we do not consider patterns that lie in a three-ball in V; that is, we assume that any pattern $P \subset V$ is geometrically essential. We call a pattern *unknotted* if it is isotopic to the core of the solid torus containing it. We do not distinguish unknotted patterns and will denote such a pattern by $\mathfrak{U}$. We reserve U for the unknot in S^3.

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

For three-manifolds with a single toroidal boundary component, Hanselman, Rasmussen, and Watson [5; 6] interpreted the relevant bordered Heegaard Floer invariants geometrically as decorated immersed curves in the once-punctured torus. Later, a formula for the behavior of these immersed curves under cabling was given by Hanselman and Watson [7]. More recently, Chen [2] studied the computation of knot Floer chain complexes of $(1,1)$-satellites by using immersed curves.

Does a nonzero-degree map give a rank inequality on Heegaard Floer homology? More specifically, Hanselman, Rasmussen, and Watson [6, Question 12] asked, if there is a degree-one map $Y_1 \to Y_2$ between closed connected orientable three-manifolds, is it the case that $\dim \widehat{\mathrm{HF}}(Y_1) \geq \dim \widehat{\mathrm{HF}}(Y_2)$? For integer homology spheres, Karakurt and Lidman [12, Conjecture 9.4] proposed that if there is a nonzero-degree map $Y_1 \to Y_2$ between them, then $\operatorname{rank} \mathrm{HF}_{\mathrm{red}}(Y_1) \geq \operatorname{rank} \mathrm{HF}_{\mathrm{red}}(Y_2)$ and $\operatorname{rank} \widehat{\mathrm{HF}}(Y_1) \geq \operatorname{rank} \widehat{\mathrm{HF}}(Y_2)$. Karakurt and Lidman [12, Theorem 1.9] also studied maps between Seifert homology spheres. It is natural to ask similar questions about knot complements in S^3 and rank inequalities on knot Floer homology. Given a degree-one map $\varphi\colon X_K \to X_U$ that preserves peripheral structure,[1] the induced map $\tilde{\varphi}\colon X_{P(K)} \to X_{P(U)}$ is well-defined and further induces an epimorphism $\tilde{\varphi}_*\colon \pi_1(X_{P(K)}) \to \pi_1(X_{P(U)})$ that also preserves peripheral structure. This is a special case of [11, Question 1.9], in which Juhász and Marengon asked, for knots K_1 and K_2 in S^3 such that there is an epimorphism $\pi_1(X_{K_1}) \to \pi_1(X_{K_2})$ preserving peripheral structure, is it true that $\dim \widehat{\mathrm{HFK}}(K_1) \geq \dim \widehat{\mathrm{HFK}}(K_2)$? We state the version corresponding to the special case $\pi_1(X_{P(K)}) \to \pi_1(X_{P(U)})$ in the following conjecture:

Conjecture 1.1 *Given any pattern P in $S^1 \times D^2$ and any companion K in S^3, there is an inequality $\dim \widehat{\mathrm{HFK}}(P(K)) \geq \dim \widehat{\mathrm{HFK}}(P(U))$, where U denotes the unknot in S^3.*

Another closely related conjecture is the following:

Conjecture 1.2 *Given any pattern P in $S^1 \times D^2$ and any companion K in S^3, there is an inequality $\dim \widehat{\mathrm{HFK}}(P(K)) \geq \dim \widehat{\mathrm{HFK}}(K)$.*

Our purpose here is to prove Conjectures 1.1 and 1.2 when P is a $(1,1)$-pattern, by using (geometrically interpreted) bordered Heegaard Floer invariants.

Theorem 1.3 *Given a $(1,1)$-pattern $P \subset S^1 \times D^2$, the inequality*

$$\dim \widehat{\mathrm{HFK}}(P(K)) \geq \dim \widehat{\mathrm{HFK}}(P(U)) \tag{1}$$

holds for any companion $K \subset S^3$, where U denotes the unknot in S^3.

Theorem 1.4 *Given a $(1,1)$-pattern $P \subset S^1 \times D^2$, the inequality*

$$\dim \widehat{\mathrm{HFK}}(P(K)) \geq \dim \widehat{\mathrm{HFK}}(K) \tag{2}$$

holds for any companion $K \subset S^3$. Moreover, the equality holds if and only if the pattern P is unknotted in $S^1 \times D^2$, or both the companion K and the satellite $P(U)$ are unknotted in S^3.

[1] We refer the reader to Boileau, Boyer, Rolfsen, and Wang [1, Proposition 1] for a proof of the existence of such a map.

Two natural questions (which were originally pointed out by Tye Lidman) to ask are the following:

Question 1.5 *Is it possible to characterize the equality condition for Theorem 1.3?*

Question 1.6 *Does either of the above theorems have a refinement for Maslov gradings?*

Although these two questions are not fully resolved here, we discuss them in detail in Section 4. We will also see that there is no refinement for Alexander gradings.

We conclude this section by briefly explaining two major parts in proving the above theorems. To begin with, the main result of Chen [2] allows us to obtain the dimension of $\widehat{\mathrm{HFK}}(P(K))$ by counting the minimum intersections of the curves $\beta(P)$ and $\alpha(K)$ in its corresponding pairing diagram (which is defined in Section 2):

Theorem 1.7 [2, Theorem 1.2] *Given a* (1,1)*-pattern* $P \subset S^1 \times D^2$ *and a companion* $K \subset S^3$*, let* $\widehat{\mathrm{HF}}(X_K) \subset \partial X_K \setminus \{w'\}$ *be the immersed curves of the knot complement* X_K*, and let* $(\beta, \mu, \lambda, w, z) \subset \partial(S^1 \times D^2)$ *be a five-tuple corresponding to a genus-one doubly pointed bordered Heegaard diagram for* P*. Let* $h\colon \partial X_K \to \partial(S^1 \times D^2)$ *be an orientation-preserving homeomorphism such that*

(i) *h identifies the meridian and Seifert longitude of K with μ and λ, respectively,*

(ii) $h(w') = w$,

(iii) *there is a regular neighborhood $U \subset \partial(S^1 \times D^2)$ of w such that $z \in U$, $U \cap (\lambda \cup \mu) = \varnothing$, and $U \cap h(\widehat{\mathrm{HF}}(X_K)) = \varnothing$.*

Let $\alpha = h(\widehat{\mathrm{HF}}(X_K))$*. Then there is a chain homotopy equivalence*

$$\widehat{\mathrm{CFK}}(\alpha, \beta, w, z) \cong \widehat{\mathrm{CFK}}(S^3, P(K)).$$

Moreover, if α is connected, this chain homotopy equivalence preserves the Maslov grading and Alexander filtration.

Besides Chen's theorem, the proof of Hanselman, Rasmussen, and Watson [6, Theorem 52] is another inspiration for an essential part of our proof — applying a sequence of moves to curves without increasing intersection numbers. Moreover, Theorem 1.4 is close to [6, Theorem 11], which is a special case of [6, Theorem 52].

Acknowledgements

I would like to thank my advisor, Jennifer Hom, for suggesting this problem, and I cannot thank her enough for her continued support, guidance, and patience. I am grateful to Wenzhao Chen and Tye Lidman for constructive comments on an earlier draft and to Steven Sivek for informative email correspondence. I would also like to thank the referee and Matthew Hedden for their insightful suggestions.

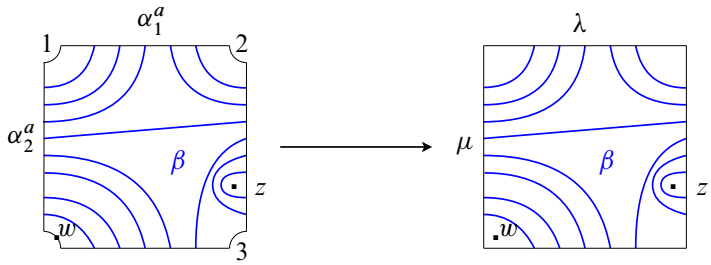


Figure 1: The data contained in a genus-one doubly pointed bordered Heegaard diagram (on the left) of the Mazur pattern can be equivalently understood as a five-tuple (on the right). This convention comes from [2, Sections 1 and 5].

2 Preliminaries

In this section, we primarily recapitulate some conventions and results in [2; 3] and then set up several notations, in preparation for proving Theorems 1.3 and 1.4 in Section 3.

We begin with a more thorough discussion about Theorem 1.7. The five-tuple in Theorem 1.7 is obtained from a genus-one doubly pointed bordered Heegaard diagram, $(\overline{\Sigma}, \{\alpha_1^a, \alpha_2^a\}, \beta, w, z)$, of P. This is done by viewing β, w, and z as embedded in $\partial(S^1 \times D^2)$ and identifying the pair of arcs $\{\alpha_1^a, \alpha_2^a\}$ with the longitude–meridian pair of $\partial(S^1 \times D^2)$. See Figure 1 for an example of the Mazur pattern.

In practice, Theorem 1.7 shows that, after identifying the torus $\partial(S^1 \times D^2)$ with the quotient space $[0,1]\times[0,1]/\sim$ in the standard way and dividing the unit square evenly into four quadrants, we can fit $\alpha = \widehat{\mathrm{HF}}(X_K)$ into the first quadrant $\left[\frac{1}{2}, 1\right]\times\left[\frac{1}{2}, 1\right]$, fit (β, w, z) into the third quadrant $\left[0, \frac{1}{2}\right]\times\left[0, \frac{1}{2}\right]$, and extend them both horizontally and vertically to obtain a diagram that yields a chain complex isomorphic to $\widehat{\mathrm{CFK}}(P(K))$. We call such a diagram a *pairing diagram* for $P(K)$, and we denote by $\alpha(K)$ and $\beta(P)$ the curves obtained from α and β by extension, respectively. Figure 2 displays four examples, in which M denotes the Mazur pattern, $T_{p,q}$ denotes the (p,q)-torus knot, and $K_{p,q}$ denotes the (p,q)-cable of a knot K.

The proof of Theorem 1.3 needs some caution, as we will be moving α-curves, which are immersed in general. The following lemma, which is widely known as the *Whitney–Graustein theorem*, allows us to get rid of self-intersections of immersed curves (after certain modifications, which will be explained in Section 3).

Lemma 2.1 ([21, Theorem 1]; see also [3, Theorem 1]) *Regular homotopy classes of regular closed curves $\bar{\gamma}: S^1 \to \mathbb{R}^2$ are in one-to-one correspondence with the integers, the correspondence being given by*

$$[\bar{\gamma}] \mapsto \mathrm{rot}(\bar{\gamma}),$$

where $\mathrm{rot}(\bar{\gamma})$ *is the degree*[2] *of the map* $S^1 \to \mathbb{R}^2\backslash\{\mathbf{0}\}$, $s \mapsto \bar{\gamma}'(s)$.

[2] By [3], the integer $\mathrm{rot}(\bar{\gamma})$ is called the *rotation number* of $\bar{\gamma}$, and it is a signed count of the number of complete turns of the velocity vector $\bar{\gamma}'$ as we traverse $\bar{\gamma}$ in a prefixed orientation.

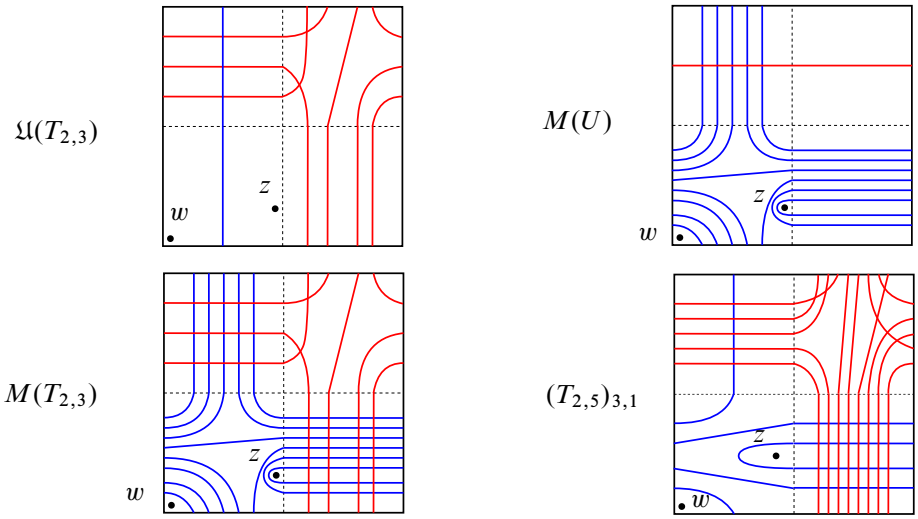


Figure 2: Examples of paring diagrams. The curves $\beta(P)$ and $\alpha(K)$ are drawn in blue and red, respectively.

The last thing we recall here is how to obtain α-curves from immersed curves in a (punctured) infinite cylinder, and vice versa. Given immersed curves in $\left(\mathbb{R}/\left(\frac{1}{2}+\mathbb{Z}\right)\right)\times\mathbb{R}$, we place a grid system consisting of two vertical columns of unit squares, with the middle vertices identified with the punctured points of the cylinder. Then we follow the curve and replicate its segment in a square every time we meet an edge of a grid square. In this way, we build its corresponding α-curves. Likewise, if we start with a torus with α-curves, we can trace the curves and recover its immersed curves in an infinite cylinder, as illustrated in Figure 3.

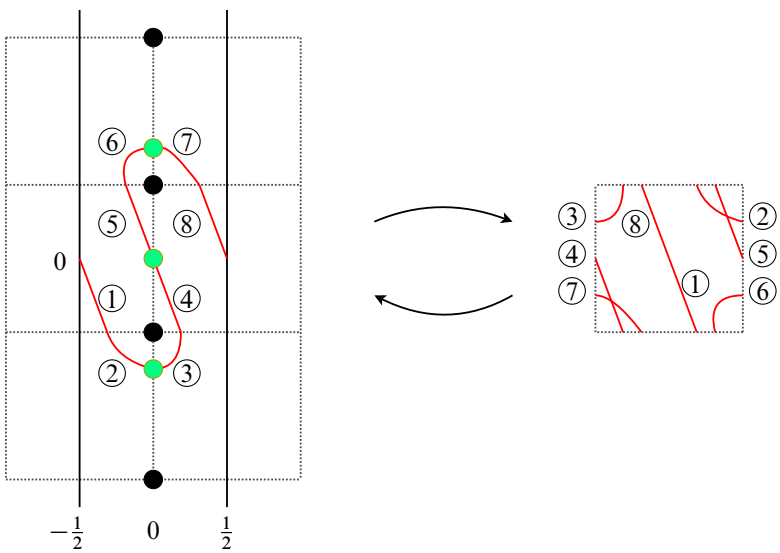


Figure 3: The immersed curve for $-T_{2,3}$ in an infinite cylinder (left) and the α-curves for $-T_{2,3}$ in a torus (right), where the circled numbers can be used to follow the construction.

3 Proof of theorems

Proof of Theorem 1.3 Given a $(1,1)$-pattern $P \subset S^1 \times D^2$ and a companion knot $K \subset S^3$, let $(T^2, \alpha(K), \beta(P), w, z)$ be a pairing diagram for $P(K)$, where $T^2 \cong S^1 \times S^1$. Lift the diagram to $\mathbb{R}^2$ by the covering map $\pi := p \times p : \mathbb{R}^2 \to T^2$, where $p : \mathbb{R} \to S^1$ is given by $x \mapsto (\cos(2\pi x), \sin(2\pi x))$. Let β_0 be a lift of $\pi^{-1}(\beta(P))$. By the construction of $\beta(P)$, the lift β_0 is connected. (See Figure 4 for an example of the Mazur pattern, where the lifts of extended α-curves are omitted.) Let α_0 be a lift of $\alpha(K)$. Notice that α_0 may not be connected, for the immersed curves α may consist of multiple components. (See Figure 5 for an example of the right-handed trefoil, where the lifts of extended β-curves are omitted.)

By [7, page 3], any immersed multicurve in an infinite cylinder has a unique component wrapping around the cylinder. Then the construction at the end of Section 2 implies that there is at least one horizontal line segment in the second quadrant of $(T^2, \alpha(K), \beta(P), w, z)$. Thus, the lift α_0 contains horizontal line segments.

Ignore any closed component that α_0 may contain. Then α_0 becomes a connected piece going to the left- and right-infinity on $\mathbb{R}^2$. (Note that we may lose some data by doing so; nevertheless, we shall see by the end of the proof that (1) still holds.) Without loss of generality, give α_0 an overall rightward orientation. Then there exist rightward-oriented horizontal line segments in α_0; indeed, otherwise, α_0 would not extend to the right infinity. Tracing α_0 from left to right, denote by μ_1 the first such segment

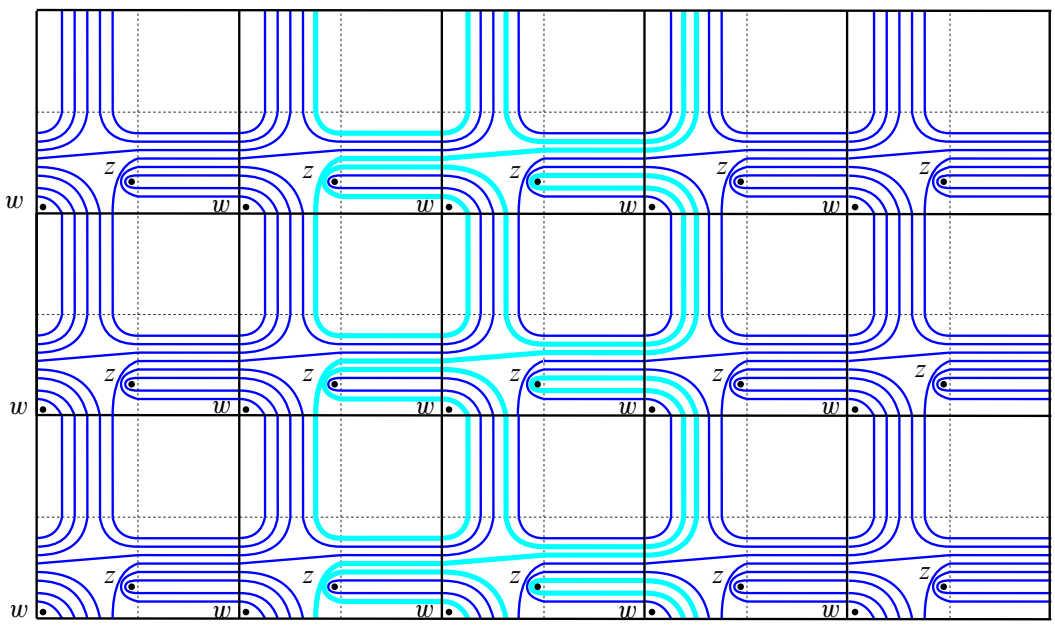


Figure 4: A choice of β_0 (highlighted in cyan) that corresponds to the Mazur pattern. (With a slight abuse of notation, we use the same symbols, w and z, for the lifts of the two basepoints, respectively.)

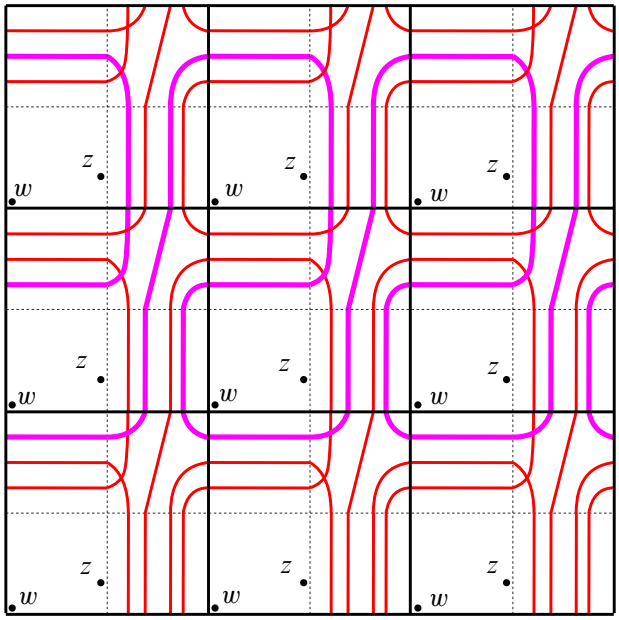


Figure 5: A choice of α_0 (highlighted in magenta) that corresponds to $T_{2,3}$.

that intersects β_0. Without loss of generality, suppose that $\mu_1 \subset \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right]$. Consider the set $\mathcal{M} := \pi^{-1}(\pi(\mu_1)) \cap ([1,\infty) \times [0,1])$, which consists of all lifts of $\pi(\mu_1)$ that lie in $[1,\infty) \times [0,1]$. For each element in $\mathcal{M}$, denote it by μ_j if it lies in $[j-1, j] \times [0,1]$. Let $k := \max\{j \mid \mu_j \cap \beta_0 \neq \varnothing\}$, and let α_0^* denote the connected portion of α_0 between μ_1 and μ_{k+1}, with μ_1 included and μ_{k+1} excluded. (See Figure 6 for an illustration, where the segments μ_j are highlighted in green.)

Notice that α_0 is periodic, exhibiting horizontal translational symmetry, so α_0^* is periodic as well. Moreover, α_0^* consists of k periods, with the j^{th} period of α_0^* starting from μ_j and terminating at the left endpoint of μ_{j+1}, for $j = 1, 2, \ldots, k$, if we traverse α_0^* from left to right.

As we mentioned in Section 2, the curve α_0^* may contain self-intersections; we claim that all of them can be resolved by regular homotopies (which are allowed to cross any lift of basepoints). Indeed, we can complete α_0^* into an immersed closed curve by attaching the top endpoint of a left semicircle of radius R to the left endpoint of μ_1, attaching the top endpoint of a right semicircle of radius R to the left endpoint of μ_{k+1}, and then connecting the two bottom endpoints of these two semicircles by a line segment, where R is sufficiently large so that the newly added three segments do not intersect α_0^*. By the 180° symmetry of immersed curves for knot complements (up to regular homotopy), the turning number induced from each self-intersection will be canceled by that of its symmetric counterpart, so overall, the rotation number of the closed curve we just created is ± 1, depending on its orientation. By Lemma 2.1, this rotation number is preserved under regular homotopies, so this closed curve is in the class of circles. Therefore, we can resolve all possible self-intersections of α_0^*.

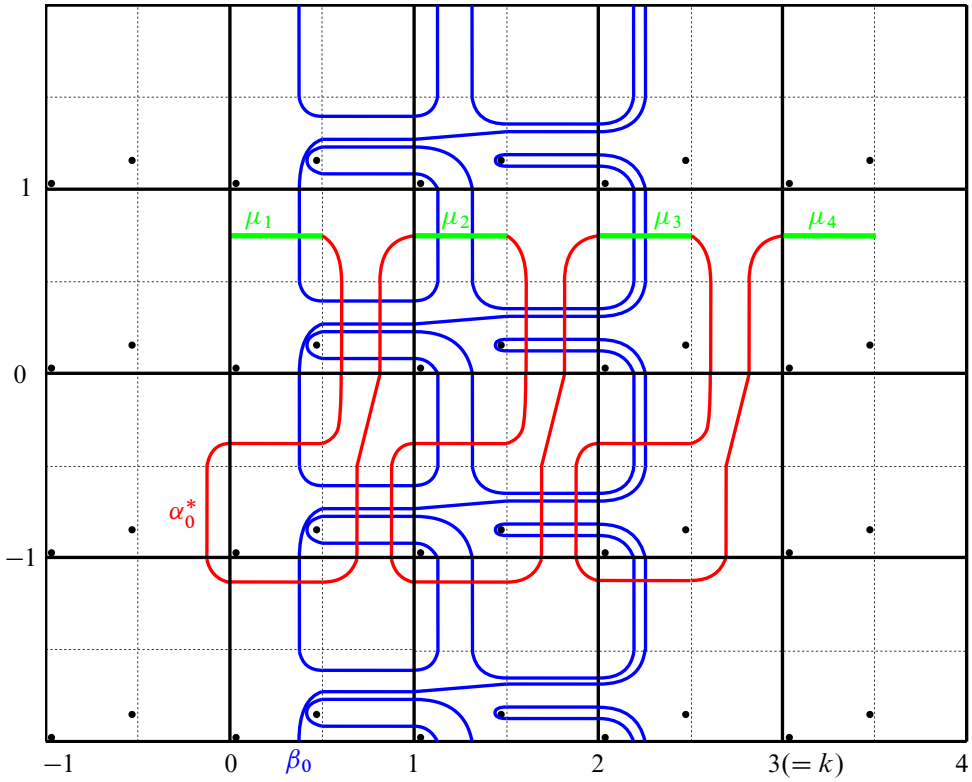


Figure 6: An example of $M(T_{2,3})$.

For each $j \in \{1, 2, \dots, k\}$, consider (the closure of) the complement of μ_j in the j^{th} period of α_0^*. Allowing passing lifts of basepoints, regularly homotope this complement with its two endpoints fixed until the curve is within $\left[j-\frac{1}{2}, j\right] \times \left[\frac{1}{2}, 1\right]$. By the claim above, we may assume all self-intersections have been resolved, so we can further regularly homotope α_0^* until it becomes a horizontal line segment in $[0,k] \times \left[\frac{1}{2}, 1\right]$. We call the resulting curve $\alpha_0^{*'}$.

Since $\alpha(U)$ is a horizontal line segment, the α_0-curve that corresponds to the unknot U, denoted by $\alpha_0(U)$, is a horizontal line in $\mathbb{R} \times \left[\frac{1}{2}, 1\right]$. Moreover, $\alpha_0(U)$ first intersects β_0 in the square $[0,1] \times [0,1]$ and lastly in $[k-1, k] \times [0,1]$, by the definitions of μ_1 and k. Therefore, to get the dimension of $\widehat{\mathrm{HFK}}(P(U))$, it suffices to consider β_0 and the portion of $\alpha_0(U)$ lying in $[0,k] \times \left[\frac{1}{2}, 1\right]$. We denote this portion by $\alpha_0^*(U)$, which is exactly the α_0^*-curve that corresponds to U, and moreover, it can be identified with $\alpha_0^{*'}$.

Consider the set $\alpha_0^{*'} \cap \beta_0$ of intersection points. For each pair $x, y \in \alpha_0^{*'} \cap \beta_0$ that form the two vertices of an empty bigon (ie a bigon that has no basepoint inside) between $\alpha_0^{*'}$ and β_0, we denote the bigon by $B_{x,y}$ and take a sufficiently small open neighborhood $U_{x,y} \supset B_{x,y}$ such that

(i) no basepoint is inside, and

(ii) after we regularly homotope β_0 inside $U_{x,y}$ to eliminate the empty bigon, no new intersection point is generated.

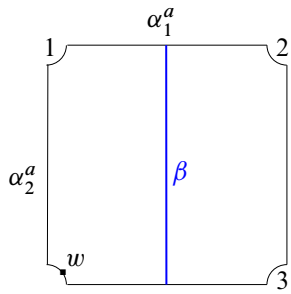


Figure 7: The genus-one bordered Heegaard diagram of $S^1 \times D^2$, up to isotopy.

Condition (ii) can be achieved since we are considering finitely many segments in $\mathbb{R}^2$. If multiple bigons are nested, then we start with the innermost one, and in this order, (ii) can still be achieved. By the definition of $U_{x,y}$, each move of β_0 does not cross basepoints. Therefore, after all such empty bigons are eliminated, we obtain a minimum intersection diagram between $\alpha_0^{*'}$ and β_0, and the final intersection number equals $\dim \widehat{\mathrm{HFK}}(P(U))$.

Observe that the sequence of moves described above eliminates all empty bigons generated by α_0^* and β_0 and does not increase the number of intersections with β_0, the number of intersections between α_0^* and β_0 is at most the number of intersections between the original α_0 and β_0, and the minimum intersection number (obtained by eliminating empty bigons via regular homotopies without passing basepoints) between the original α_0 and β_0 gives $\dim \widehat{\mathrm{HFK}}(P(K))$. These observations, together with the result in the above paragraph, imply that $\dim \widehat{\mathrm{HFK}}(P(K)) \geq \dim \widehat{\mathrm{HFK}}(P(U))$. □

Remark 3.1 In the proof above, we applied regular homotopies in the covering space $\mathbb{R}^2$ of T^2. Recall that the covering map π is defined as $p \times p$, where $p\colon \mathbb{R} \to S^1$ is given by $x \mapsto (\cos(2\pi x), \sin(2\pi x))$. Composing those regular homotopies with π, we shall get regular homotopies in the base space T^2.

Proof of Theorem 1.4 Here we continue with the curves α_0 and β_0 that have been set up in the first paragraph of the proof of Theorem 1.3.

Recall that the five-tuple $(\beta, \mu, \lambda, w, z) \subset \partial(S^1 \times D^2)$ is constructed from a genus-one doubly pointed bordered Heegaard diagram, say $(\overline{\Sigma}, \{\alpha_1^a, \alpha_2^a\}, \beta, w, z)$, for $(S^1 \times D^2, P)$. Forgetting the z-basepoint, we obtain a genus-one bordered Heegaard diagram for the solid torus $S^1 \times D^2$ with the standard parametrization of $\partial(S^1 \times D^2)$. Therefore, up to isotopy (not passing the w-basepoint), the diagram $(\overline{\Sigma}, \{\alpha_1^a, \alpha_2^a\}, \beta, w)$ is in the form shown in Figure 7.

It then follows from the construction of the curve β_0 that, if we forget the basepoint z, we can isotope β_0 without passing any lift of the basepoint w until it becomes a vertical straight line; we denote the line by β_0'. See Figure 8 for an illustration.

Since the β-curve associated with the unknotted pattern $\mathfrak{U}$ is a vertical line segment, the corresponding β_0-curve, denoted by $\beta_0(\mathfrak{U})$, is a vertical line in $\mathbb{R}^2$, which can be identified with β_0'. The number

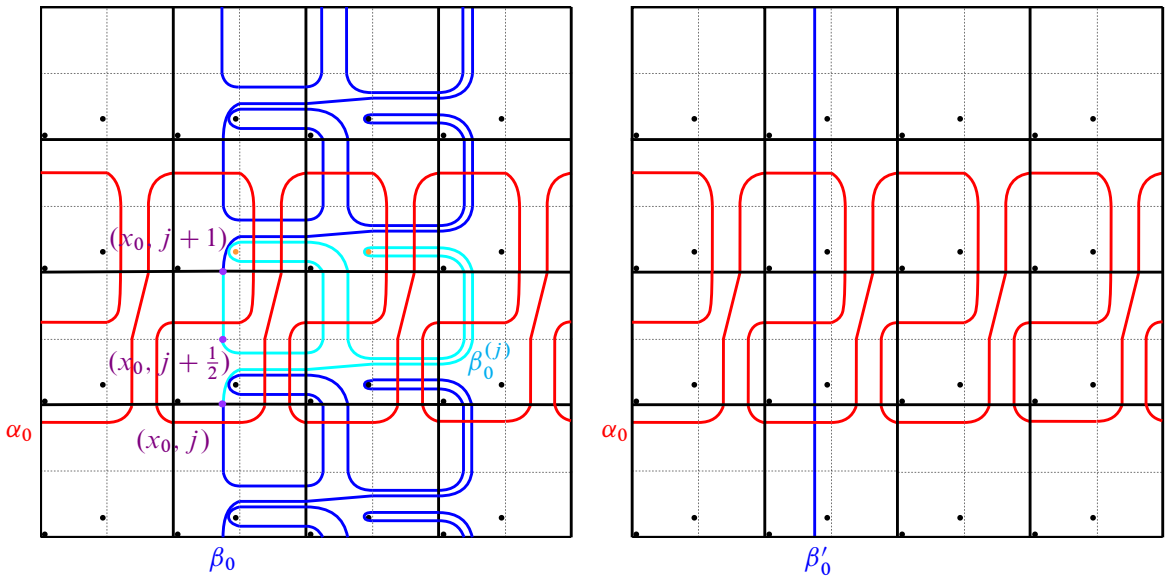


Figure 8: The α_0-, β_0-, and β_0'-curves for $M(T_{2,3})$, in which a β_0-period is highlighted in cyan for some $j \in \mathbb{Z}$.

of intersections between $\beta_0(\mathfrak{U})$ and α_0 gives $\dim \widehat{\mathrm{HFK}}(\mathfrak{U}(K))$; therefore, the number of intersections between β_0' and α_0 is precisely $\dim \widehat{\mathrm{HFK}}(K)$.

Now we compare the pair of curves (α_0, β_0') with the original (α_0, β_0). Observe that the isotopies described above eliminate all empty bigons generated by β_0 and α_0 and do not increase the number of intersections with α_0, and the minimum intersection number (obtained by eliminating empty bigons via isotopies without passing basepoints) between β_0 and α_0 gives $\dim \widehat{\mathrm{HFK}}(P(K))$. These observations, together with the result in the above paragraph, imply that $\dim \widehat{\mathrm{HFK}}(P(K)) \geqslant \dim \widehat{\mathrm{HFK}}(K)$.

In the following, we justify the equality conditions. One direction is clear: If the pattern P is unknotted (ie $P = \mathfrak{U}$), then $\dim \widehat{\mathrm{HFK}}(P(K)) = \dim \widehat{\mathrm{HFK}}(\mathfrak{U}(K)) = \dim \widehat{\mathrm{HFK}}(K)$ for any companion $K \subset S^3$. If both the companion K and the satellite $P(U)$ are unknotted in S^3 (ie $K = U = P(U)$), then $\dim \widehat{\mathrm{HFK}}(P(K)) = \dim \widehat{\mathrm{HFK}}(P(U)) = \dim \widehat{\mathrm{HFK}}(U) = \dim \widehat{\mathrm{HFK}}(K)$.

Next we prove the opposite direction. Assuming that $P \neq \mathfrak{U}$, and that $K \neq U$ or $P(U) \neq U$, we show that we actually have a strict inequality in Theorem 1.4:

$$\dim \widehat{\mathrm{HFK}}(P(K)) > \dim \widehat{\mathrm{HFK}}(K). \tag{3}$$

If $K = U$, then by the assumption we just made $P(U) \neq U$; since knot Floer homology detects the unknot [17, Theorem 1.2], it follows that $\dim \widehat{\mathrm{HFK}}(P(U)) > \dim \widehat{\mathrm{HFK}}(U)$. We may now assume that $K \neq U$ and $P \neq \mathfrak{U}$. Then in terms of the α- and β-curves in the universal cover $\mathbb{R}^2$, α_0 is not a horizontal line and β_0 is not a vertical line. In particular, $y_{\alpha_0}^{\mathrm{top}} - y_{\alpha_0}^{\mathrm{bottom}} > 0$, where $y_{\alpha_0}^{\mathrm{top}}$ and $y_{\alpha_0}^{\mathrm{bottom}}$ are defined as $y_{\alpha_0}^{\mathrm{top}} = \max\{y \in \mathbb{R} \mid (x, y) \in \alpha_0\}$ and $y_{\alpha_0}^{\mathrm{bottom}} = \min\{y \in \mathbb{R} \mid (x, y) \in \alpha_0\}$, respectively, assuming that

a Cartesian coordinate system has been fixed for $\mathbb{R}^2$ beforehand. As for the β_0-curve, we first need to discuss more of its properties in the following paragraph.

We use the notation $\beta_0+(x,y)$, where $x, y\in\mathbb{R}$, to denote the translation of β_0 that is obtained from first sliding it $|x|$ units to the right if $x\geqslant 0$ and to the left if $x<0$, and then sliding it $|y|$ units upward if $y\geqslant 0$ and downward if $y<0$. The fact that the $\beta(P)$-curve is embedded implies that $\beta_0=\beta_0+(0,1)$ and $\beta_0\cap[\beta_0+(\pm 1,0)]=\varnothing$. In words, it means that the β_0-curve is invariant under vertical translations of integral units, and if we consider all the lifts of $\beta(P)$ to $\mathbb{R}^2$, any two of them (in particular, adjacent pairs) are disjoint. Give β_0 an overall downward orientation. Recall that we can isotope β_0, allowing it to pass through the lifts of the z-basepoint but not those of the w-basepoint, to obtain the vertical line β_0'. We may assume that during that process, a vertical line segment $\{x_0\}\times\left[j+\frac{1}{2}, j+1\right]$, where $x_0\in\left(i, i+\frac{1}{2}\right)$ for some $i\in\mathbb{Z}$, is fixed for each $j\in\mathbb{Z}$, and as a result, the vertical line β_0' is $\{x_0\}\times\mathbb{R}$. For each $j\in\mathbb{Z}$, we call the connected part of β_0 that connects the two points $(x_0, j+1)$ and (x_0, j) the j^{th} *period* of β_0 and denote it by $\beta_0^{(j)}$. By our assumption, the vertical line segment $\{x_0\}\times\left[j+\frac{1}{2}, j+1\right]$ is part of $\beta_0^{(j)}$, for each $j\in\mathbb{Z}$. Notice also that these β_0-periods satisfy[3] that $\beta_0^{(j+1)}=\beta_0^{(j)}+(0,1)$, for each $j\in\mathbb{Z}$. Figure 8, left, gives an illustration, in which one β_0-period is highlighted in cyan for some $j\in\mathbb{Z}$, and the points (x_0, j), $\left(x_0, j+\frac{1}{2}\right)$, and $(x_0, j+1)$ are in violet.

We will complete the proof by dividing the possible β_0-curves into three cases and then proving the result in each case, but before introducing them, we need to justify the following claim: For each $j\in\mathbb{Z}$, denoting the infinite horizontal region $\mathbb{R}\times[j, j+1]$ by $\mathcal{H}_j$, if $\beta_0^{(j)}\not\subset\mathcal{H}_j$, then either $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j+1})=\varnothing$ or $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j-1})=\varnothing$, up to isotopy respecting all basepoints. (Roughly speaking, if a β_0-period $\beta_0^{(j)}$ is not contained in the horizontal region $\mathcal{H}_j$, then it either goes up or down, but not in both directions.) We prove the claim by showing that, if $\beta_0^{(j)}\not\subset\mathcal{H}_j$ and $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j+1})\neq\varnothing$, then $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j-1})=\varnothing$; the other case is similar. Indeed, if $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j-1})\neq\varnothing$, then since β_0 can be isotoped to β_0' without passing through w points, there must be some z point (denoted by z_{j-p}) in some $\mathcal{H}_{j-p}$, where $p\in\mathbb{Z}$ and $p\geqslant 1$, such that $\beta_0^{(j)}$ winds around z_{j-p} (or in other words, the point z_{j-p} blocks $\beta_0^{(j)}$ from isotopies respecting all basepoints). For example, the two orange z points in Figure 8, left, block $\beta_0^{(j)}$ in $\mathcal{H}_{j+1}$. Likewise, since $\beta_0^{(j)}\cap\mathcal{H}_{j+1}\neq\varnothing$, there is some z point (denoted by z_{j+q}) in some $\mathcal{H}_{j+q}$, where $q\in\mathbb{Z}$ and $q\geqslant 1$, such that $\beta_0^{(j)}$ winds around z_{j+q}. Consider the $(j+q+p)^{\text{th}}$ β_0-period. We can deduce from the translational properties of β_0 (which have been discussed in the above paragraph) that some horizontal translation of $\beta_0^{(j+q+p)}$ by an integral unit (which might be 0) winds around the point z_{j+q} as well as $\beta_0^{(j)}$. This, however, results in intersections between that curve and $\beta_0^{(j)}$, contradicting the properties that the lifts of $\beta(P)$ are embedded in $\mathbb{R}^2$ and any two of them are disjoint.

Thanks to the claim above, we can divide the β_0-curves into three cases: $\beta_0^{(j)}\subset\mathcal{H}_j$, $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j+1})\neq\varnothing$, and $\beta_0^{(j)}\cap\operatorname{int}(\mathcal{H}_{j-1})\neq\varnothing$, for any $j\in\mathbb{Z}$. We show in the following that in each case the strict inequality (3) holds:

[3] Here we extend the application of the translation notation that was originally defined for β_0.

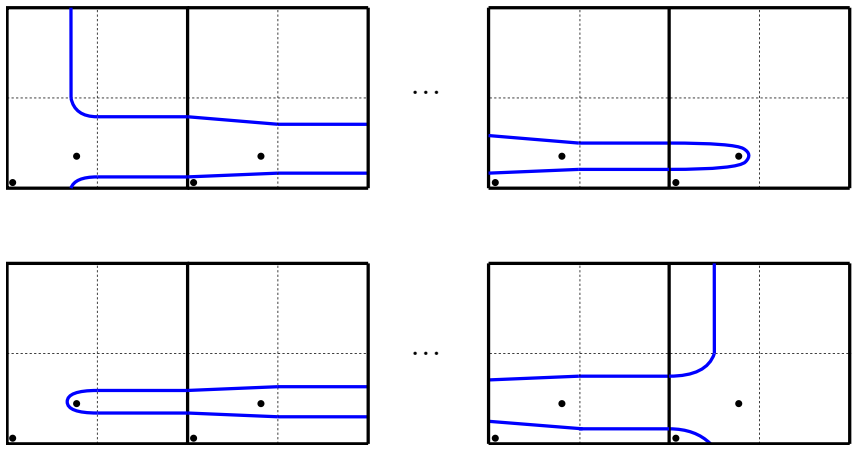


Figure 9: The two cases of a β_0-period when it is contained in a single H_j region.

(I) When $\beta_0^{(j)} \subset \mathcal{H}_j$, each period $\beta_0^{(j)}$ is as depicted in one of the two diagrams in Figure 9, up to isotopy respecting all basepoints. Indeed, if $\beta_0^{(j)}$ appears on both sides of $\beta_0' = \{x_0\} \times \mathbb{R}$, then there exist two integers $n_1, n_2 \in \mathbb{Z}$ with $n_1 + 1 < x_0 < n_2$ such that $\beta_0^{(j)}$ winds around the left side of the z-basepoint in the square $[n_1, n_1 + 1] \times [j, j + 1]$ and winds around the right side of the z-basepoint in the square $[n_2, n_2 + 1] \times [j, j + 1]$. Consider the translation $\beta_0^{(j)} + (n_1 - n_2, 0)$: it winds around the right side of the z-basepoint in the square $[n_1, n_1 + 1] \times [j, j + 1]$, and therefore it intersects with $\beta_0^{(j)}$, which is a contradiction. Comparing Figure 9 with [10, Figure 6], it is not hard to see that the patterns in this case are precisely the $(p, 1)$-cabling patterns, where $p \in \mathbb{Z}$ and $|p| \geqslant 2$. By the formula for the behavior of immersed curves under cabling from [7], if we denote the lift of $\alpha(K)$ to the punctured infinite cylinder $\big(\mathbb{R}/\big(\frac{1}{2} + \mathbb{Z}\big)\big) \times \mathbb{R}$ by $\hat{\alpha}(K)$, then p copies of $\hat{\alpha}(K)$ will be introduced and eventually aligned along the line of punctures. Thus, the number of intersections between the immersed multicurve and the line connecting the punctures strictly increases, which gives us the strict inequality as desired.

(II) If $\beta_0^{(j)} \cap \operatorname{int}(\mathcal{H}_{j+1}) \neq \varnothing$, then $\beta_0^{(j)} \cap \operatorname{int}(\mathcal{H}_{j-1}) = \varnothing$. Let k be the ceiling of $y_{\alpha_0}^{\text{bottom}}$, ie $k \in \mathbb{Z}$ and $y_{\alpha_0}^{\text{bottom}} < k < y_{\alpha_0}^{\text{bottom}} + 1$. The isotopy that takes β_0 to β_0' can be considered as the composition of isotopies, each of which takes one β_0-period into a vertical line segment lying in β_0'. Now in this case, we only partially isotope β_0: we do such isotopies for all the $\beta_0^{(j)}$ where $j \geqslant k$ but leave the remaining β_0-periods as they are. See Figure 10 for an illustration. Note that, by construction, the ray $\beta_0' \cap \big(\{x_0\} \times \big[k - \frac{1}{2}, \infty\big)\big)$ is still part of the β_0-curve after isotopies, and the intersection number between it and α_0 gives us $\dim \widehat{\mathrm{HFK}}(K)$. Now further isotope $\big(\bigcup_{j \leqslant k-1} \beta_0^{(j)}\big) - \{x_0\} \times \big[k - \frac{1}{2}, k\big]$ to eliminate all empty bigons. Since $\beta_0^{(k-1)} \cap \operatorname{int}(\mathcal{H}_k) \neq \varnothing$, it is blocked by at least one z-point above $\mathcal{H}_{k-1}$. For β_0 to reach such a z-point, it is impossible for it to avoid entering the second and fourth quadrants of all unit squares with integral-coordinates vertices. Thus, there must be additional intersections between β_0 and α_0, so the intersection number must exceed $\dim \widehat{\mathrm{HFK}}(K)$. Hence, the inequality (3) holds.

(III) If $\beta_0^{(j)} \cap \operatorname{int}(\mathcal{H}_{j-1}) \neq \varnothing$, then $\beta_0^{(j)} \cap \operatorname{int}(\mathcal{H}_{j+1}) = \varnothing$. The proof of this case is similar to (II). □

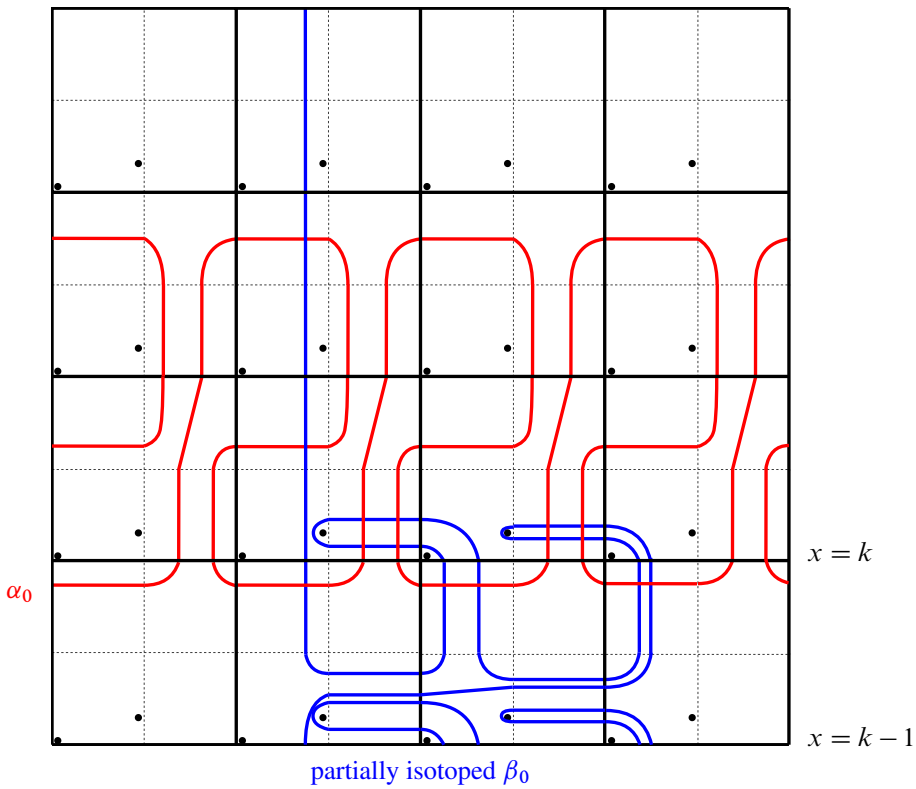


Figure 10: An illustration of how the β_0 curve is partially isotoped as needed in case (II) of the proof of the strict inequality (3). Here k is the ceiling of $y_{\alpha_0}^{\text{bottom}}$.

4 Further remarks

In this section, we make several remarks, further discussing the two inequalities we have proved. In particular, we give more details about the two questions, which were suggested by Jennifer Hom and Lidman, from Section 1.

We begin with an easy observation. A major property of knot Floer homology is that it categorifies the Alexander polynomial $\Delta_K(t)$ of knots $K \subset S^3$ [18]:

$$\Delta_K(t) = \sum_{a\in\mathbb{Z}} \Big[\sum_{m\in\mathbb{Z}} (-1)^m \dim \widehat{\mathrm{HFK}}_m(K,a) \Big] t^a.$$

Also, there is a symmetry [18, Proposition 3.10]:

$$\widehat{\mathrm{HFK}}_m(K,a) \cong \widehat{\mathrm{HFK}}_{m-2a}(K,-a).$$

Since one of the characterizing conditions for Alexander polynomials is that

$$\Delta_K(1) = 1,$$

it follows that the parity of $\dim \widehat{\mathrm{HFK}}(K)$ is odd.

Remark 4.1 Ozsváth and Szabó [17] proved that if $\dim \widehat{\mathrm{HFK}}(K) = 1$ then K is the unknot. Hedden and Watson [9, Corollary 8] showed that if $\dim \widehat{\mathrm{HFK}}(K) = 3$ then K is a (left- or right-handed) trefoil; see also [4, Corollary 1.5]. From these two knot-detecting results, the example depicted in Figure 11 (which shows that there exists a $(1,1)$-satellite K with $\dim \widehat{\mathrm{HFK}}(K) = 5$), and the above parity result, we deduce that if $\dim \widehat{\mathrm{HFK}}(K) < 5$ then K is not a $(1,1)$-satellite.

Next, we discuss some special cases when we have a strict inequality:

Remark 4.2 In the proof of Theorem 1.3, we mentioned that any immersed multicurve in an infinite cylinder has a unique component wrapping around the cylinder. When the set of immersed curves corresponds to a knot complement, this component is an invariant of the knot, and furthermore, an invariant of the concordance class of the knot [7, Proposition 2]. It follows that all slice knots have a trivial such component as the unknot does; in terms of the notation we used in the proof above, with all closed components removed, α_0 is a straight horizontal line. Moreover, for nontrivial slice knots, there are some additional closed components. Therefore, for a nontrivial slice companion knot K, inequality (2) in Theorem 1.4 is strict:

$$\dim \widehat{\mathrm{HFK}}(P(K)) > \dim \widehat{\mathrm{HFK}}(K).$$

In addition to that case, Petkova [19, Lemma 7] showed that the complex $\mathrm{CFK}^-(K)$ for Floer homologically thin knots K splits into exactly one staircase summand and possibly multiple square summands. Geometrically, a square summand is represented by a closed component in α_0, so for Floer homologically thin knots K containing a square summand in $\mathrm{CFK}^-(K)$, the above strict inequality holds as well.

Our last remark is related to gradings:

Remark 4.3 In the proof of Theorem 1.4, we managed to isotope the curve β_0 in a desired way, passing only the z-lifts. Then by Theorem 1.7, if α is connected, and if we only consider Maslov gradings, what we have proved is also true. That is, if α is connected, the inequality

$$\sum_{a\in\mathbb{Z}} \dim \widehat{\mathrm{HFK}}_m(P(K), a) \geqslant \sum_{a\in\mathbb{Z}} \dim \widehat{\mathrm{HFK}}_m(K, a)$$

holds for any Maslov grading $m \in \mathbb{Z}$.

On the other hand, the regular homotopies in the proof of Theorem 1.3 cannot achieve that property in general; for example, Figure 6 shows that we cannot tighten the curve α_0 corresponding to $T_{2,3}$ to a horizontal line without passing any w-lift.

If we just consider Alexander gradings, we will not arrive at rank inequalities that work for all $(1,1)$-satellites. Indeed, consider the example of $(T_{2,3})_{2,3}$, the $(2,3)$-cable of the right-handed trefoil. Since $T_{2,3}$ is alternating, its knot Floer homology is determined by the Alexander polynomial and the knot signature [16], and it is depicted in Figure 11, left. By Hedden's work [8], the knot Floer homology

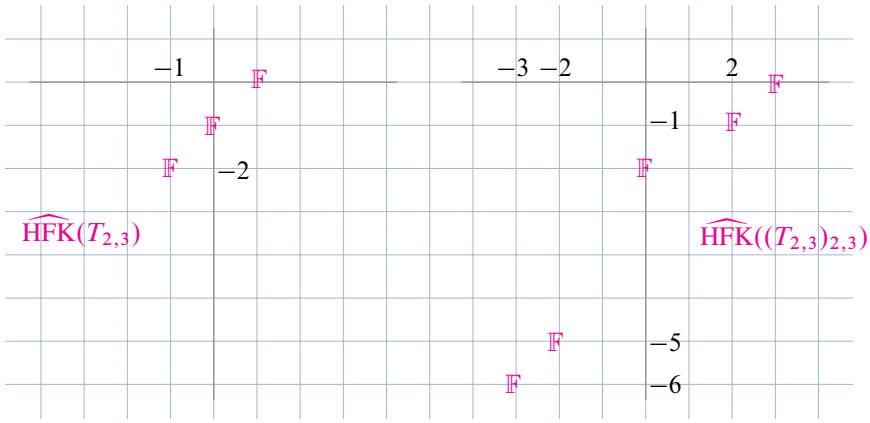


Figure 11: Knot Floer homologies of $T_{2,3}$ and $(T_{2,3})_{2,3}$, plotted on an (a,m)-coordinate system (where the horizontal axis is for a).

of $(T_{2,3})_{2,3}$ can also be determined from its symmetrized Alexander polynomial, and it is depicted in Figure 11, right. This example was originally constructed by Hom; see also [11, Example 1.8]. Now we can make the following observations:

a	$\sum_{m\in\mathbb{Z}} \dim \widehat{\mathrm{HFK}}_m(T_{2,3},a)$	$\sum_{m\in\mathbb{Z}} \dim \widehat{\mathrm{HFK}}_m((T_{2,3})_{2,3},a)$	observations
-2	0	1	$0<1$
-1	1	0	$1>0$
0	1	1	$1=1$

The last column in the above table shows that Theorems 1.3 and 1.4 do not have refinements for Alexander gradings.

References

[1] **M Boileau**, **S Boyer**, **D Rolfsen**, **S C Wang**, *One-domination of knots*, Illinois J. Math. 60 (2016) 117–139 MR Zbl

[2] **W Chen**, *Knot Floer homology of satellite knots with* $(1,1)$ *patterns*, Selecta Math. 29 (2023) art. id. 53 MR Zbl

[3] **H Geiges**, *A contact geometric proof of the Whitney–Graustein theorem*, Enseign. Math. 55 (2009) 93–102 MR Zbl

[4] **P Ghiggini**, *Knot Floer homology detects genus-one fibred knots*, Amer. J. Math. 130 (2008) 1151–1169 MR Zbl

[5] **J Hanselman**, **J Rasmussen**, **L Watson**, *Heegaard Floer homology for manifolds with torus boundary: properties and examples*, Proc. Lond. Math. Soc. 125 (2022) 879–967 MR Zbl

[6] **J Hanselman**, **J Rasmussen**, **L Watson**, *Bordered Floer homology for manifolds with torus boundary via immersed curves*, J. Amer. Math. Soc. 37 (2024) 391–498 MR Zbl

[7] **J Hanselman**, **L Watson**, *Cabling in terms of immersed curves*, Geom. Topol. 27 (2023) 925–952 MR Zbl

[8] **M Hedden**, *On knot Floer homology and cabling*, Algebr. Geom. Topol. 5 (2005) 1197–1222 MR Zbl

[9] **M Hedden**, **L Watson**, *On the geography and botany of knot Floer homology*, Selecta Math. 24 (2018) 997–1037 MR Zbl

[10] **J Hom**, *Bordered Heegaard Floer homology and the tau-invariant of cable knots*, J. Topol. 7 (2014) 287–326 MR Zbl

[11] **A Juhász**, **M Marengon**, *Concordance maps in knot Floer homology*, Geom. Topol. 20 (2016) 3623–3673 MR Zbl

[12] **Ç Karakurt**, **T Lidman**, *Rank inequalities for the Heegaard Floer homology of Seifert homology spheres*, Trans. Amer. Math. Soc. 367 (2015) 7291–7322 MR Zbl

[13] **A S Levine**, *Knot doubling operators and bordered Heegaard Floer homology*, J. Topol. 5 (2012) 651–712 MR Zbl

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

[15] **Y Ni**, *Knot Floer homology detects fibred knots*, Invent. Math. 170 (2007) 577–608 MR Zbl

[16] **P Ozsváth**, **Z Szabó**, *Heegaard Floer homology and alternating knots*, Geom. Topol. 7 (2003) 225–254 MR Zbl

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

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

[19] **I Petkova**, *Cables of thin knots and bordered Heegaard Floer homology*, Quantum Topol. 4 (2013) 377–409 MR Zbl

[20] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR arXiv math/0306378

[21] **H Whitney**, *On regular closed curves in the plane*, Compositio Math. 4 (1937) 276–284 MR Zbl

School of Mathematics, Georgia Institute of Technology
Atlanta, GA, United States

wshen41@gatech.edu

Received: 22 September 2022 Revised: 29 October 2023

msp

Algebraic & Geometric Topology **25**:6 (2025) 3287–3314
DOI: 10.2140/agt.2025.25.3287
Published: 1 October 2025

A K-theory spectrum for cobordism cut and paste groups

RENEE S HOEKZEMA
CARMEN ROVI
JULIA SEMIKINA

Cobordism groups and cut and paste groups of manifolds arise from imposing two different relations on the monoid of manifolds under disjoint union. By imposing both relations simultaneously, a cobordism cut and paste group $\overline{\mathrm{SK}}_n$ is defined. We extend this definition to manifolds with boundary, obtaining the group $\overline{\mathrm{SK}}_n^\partial$, and study the relationship of the group to an appropriately defined cobordism group of manifolds with boundary. The main results are the construction of a spectrum that recovers on π_0 the cobordism cut and paste groups of manifolds with boundary, $\overline{\mathrm{SK}}_n^\partial$, and a map of spectra that lifts the canonical quotient map $\mathrm{SK}_n^\partial \to \overline{\mathrm{SK}}_n^\partial$.

18D99, 18F25, 18F30, 19D99, 57N70; 55P15, 57N40, 57R75, 57R90

1. Introduction 3287
2. Cobordism cut and paste groups of manifolds with boundary 3291
3. Cubes of manifolds with boundary 3296
4. Recovering cobordism cut and paste groups as $K_0^{\square}$ 3300
5. $K^{\square}$ is a spectrum 3309
6. Maps from $B\overline{\mathrm{Cob}}_{n+1}^{\partial_{\mathrm{triv}}}$ and $K^{\square}(\overline{\mathrm{Mfd}}_n^\partial)$ 3312
References 3313

1 Introduction

The classical notion of cut and paste equivalence for closed manifolds was first introduced by Karras, Kreck, Neumann, and Ossa [8]. The cut and paste operation on closed n-dimensional manifolds is described by the following procedure. Given a manifold, one cuts along a separating codimension-1 submanifold with trivial normal bundle and glues the resulting two pieces back together along an orientation-preserving diffeomorphism to obtain a new manifold. Manifolds M and N are called SK-equivalent (SK stands for "schneiden und kleben", German for "cut and paste") if N can be obtained from M by a sequence of

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

finitely many cut and paste operations. Equivalence classes of n-manifolds up to SK-equivalence form the groups SK_n of cut and paste invariants of manifolds, with the operation given by disjoint union.

In [7] the notions of SK-equivalence and SK-groups were generalized to the case of manifolds with boundary. Allowing boundary made it possible to formulate a suitable notion of a pointed category with squares Mfd_n^∂ that fits into the framework of K-theory with squares due to [3]. Their construction is a modified version of the Thomason construction of the K-theory spectrum for Waldhausen's category, but it allows more flexibility for the category. The distinguished squares for the category Mfd_n^∂ correspond to pushout squares that glue manifolds along a common codimension-0 submanifold, and these can be shown to encode the SK^∂-relations. The K-theory spectrum $K^\square(\mathrm{Mfd}_n^\partial)$, constructed in [7], recovers the SK_n^∂ as its zeroth homotopy group.

A different notion of equivalence of manifolds is cobordism. Two n-dimensional manifolds M and N are cobordant if their disjoint union forms the boundary of an $(n+1)$-dimensional manifold. Cobordism classes of n-manifolds again form a group Ω_n under disjoint union. The cobordism cut and paste group $\overline{\mathrm{SK}}_n$ is given by quotienting n-manifolds by both cobordism and SK-equivalence. For oriented manifolds it was shown in [8] that the only cobordism cut and paste invariant is the signature.

Here we generalize the definition of the $\overline{\mathrm{SK}}$-groups to the case of manifolds with boundary, and denote the corresponding group by $\overline{\mathrm{SK}}_n^\partial$. In order to define this group we need to quotient by an appropriate notion of cobordism for manifolds with boundary. The usual notion of cobordism for manifolds with boundary allows for cobordisms with corners and free boundary components, but in this setting any two manifolds of the same dimension are cobordant. To fix this issue we restrict to cobordisms that are trivial on the boundary, meaning that when restricted to the boundary of the in- and outgoing manifolds, the cobordism is diffeomorphic to a cylinder. We show that the groups $\overline{\mathrm{SK}}_n^\partial$ are related to the classical $\overline{\mathrm{SK}}_n$ groups via an exact sequence analogous to the one found for SK_n^∂ in [7]:

$$0 \to \overline{\mathrm{SK}}_n \xrightarrow[[M]\mapsto[M]]{\alpha} \overline{\mathrm{SK}}_n^\partial \xrightarrow[[N]\mapsto[\partial N]]{\beta} C_{n-1} \to 0.$$

Here C_{n-1} is the group completion of the monoid of orientation-preserving diffeomorphism classes of bounding $(n-1)$-manifolds under disjoint union. This is a free abelian group.

Next we construct a K-theory spectrum that recovers the $\overline{\mathrm{SK}}_n^\partial$ as its zeroth homotopy group. For this we take inspiration from the K-theory with squares construction of [3] and generalize it in several ways. The input for the $K^\square$-construction is a category with squares $\mathcal{C}$, which is a pointed category with two subcategories of morphisms referred to as horizontal (denoted by $\hookrightarrow$) and vertical maps (denoted by $\rightarrowtail$), and distinguished squares

$$\begin{array}{ccc} A & \hookrightarrow & B \\ \downarrow & \square & \downarrow \\ C & \hookrightarrow & D \end{array}$$

satisfying certain conditions; see [3; 7].

Given a category with squares $\mathcal{C}$, one associates to it a bisimplicial set $N_\bullet \mathcal{C}^\bullet$, where the set of (k,l)-simplices $N_k \mathcal{C}^{(l)}$ is given by the set of all $k \times l$ diagrams

$$\begin{array}{ccccccc}
C_{00} & \hookrightarrow & C_{01} & \hookrightarrow & \cdots & \hookrightarrow & C_{0l} \\
\downarrow & \square & \downarrow & \square & & \square & \downarrow \\
C_{10} & \hookrightarrow & C_{11} & \hookrightarrow & \cdots & \hookrightarrow & C_{1l} \\
\downarrow & \square & \downarrow & \square & & \square & \downarrow \\
\vdots & & \vdots & & \ddots & & \vdots \\
\downarrow & \square & \downarrow & \square & & \square & \downarrow \\
C_{k0} & \hookrightarrow & C_{k1} & \hookrightarrow & \cdots & \hookrightarrow & C_{kl}
\end{array}$$

with face maps given by deleting a corresponding row or column and degeneracy maps given by inserting the copy of the corresponding row or column. The squares K-theory space of $\mathcal{C}$ is

$$K^{\square}(\mathcal{C}) \simeq \Omega_O |N_\bullet \mathcal{C}^\bullet|,$$

where Ω_O is the based loop space, based at the distinguished object $O \in N_0 \mathcal{C}^{(0)}$. For our purposes we generalize the construction as follows.

(i) Instead of having two simplicial directions and distinguished squares we will have three directions ($\hookrightarrow$, $\rightarrowtail$ and $\rightsquigarrow$) and distinguished cubes as part of the data. It is easy to see that the construction of [3] has a straightforward generalization for the categories with distinguished m-cubes for any $m \geq 2$.

(ii) The substantial modification is that we do not require morphisms corresponding to different simplicial directions to come from the same category. In our application, objects are n-dimensional manifolds with boundary, but the three classes of morphisms are of a different nature: two simplicial directions, horizontal ($\hookrightarrow$) and vertical ($\rightarrowtail$), come from a particular type of embeddings of manifolds that we refer to as SK-embeddings; the third (depth) simplicial direction ($\rightsquigarrow$) is given by cobordisms between manifolds that are cylindrical on the boundary. We moreover have SK-embeddings between the cobordisms as part of the data defining a cube.

(iii) We take the topology into account by means of a fourth simplicial direction, in the spirit of [11]. As a result of this modification, we obtain a topological version of the cut and paste spectrum of manifolds (for the discrete one see [7]), which also has SK_n^∂ as its π_0. We denote it by $K^{\square}(\overline{\mathrm{Mfd}}_n^\partial)$.

We construct a certain quadrisimplicial set $X^*_{\bullet,\bullet,\bullet}$ with (r,k,l,m)-simplices given by $k \times l \times m$ diagrams consisting of distinguished cubes fibered over Δ^r and show that its geometric realization is an infinite loop space. Since the resulting space is connected, its π_0 does not say anything interesting, and we shift π_0 by looking at the space of loops instead. We denote the resulting spectrum by $K^{\square\!\!\square}(\overline{\mathrm{Mfd}}_n^\partial)$, where the $\square\!\!\square$ notation reflects the three simplicial directions given by maps of manifolds that are used in the

construction, Mfd_n^∂ refers to the nature of objects, and the bar over it is present to emphasize the topology captured in the fourth simplicial direction. Let us denote by $K_i^{\overline{\square}}(\overline{\mathrm{Mfd}}_n^\partial)$ the π_i of the spectrum.

Theorem A *There is an isomorphism* $K_0^{\overline{\square}}(\overline{\mathrm{Mfd}}_n^\partial) \cong \overline{\mathrm{SK}}_n^\partial$.

From our definition of the quadrisimplicial set $X^*_{\bullet,\bullet,\bullet}$ we automatically deduce a topological analogue of the squares cut and paste spectrum, whose infinite loop space is obtained from a trisimplicial space $X^*_{\bullet,\bullet,0}$ by taking the loop space of its geometric realization. We show that the inclusion of the trisimplicial into the quadrisimplicial set provides a categorification of the quotient map $\mathrm{SK}_n^\partial \to \overline{\mathrm{SK}}_n^\partial$.

Theorem B *The map of spectra*

$$K^{\square}(\overline{\mathrm{Mfd}}_n^\partial) \to K^{\overline{\square}}(\overline{\mathrm{Mfd}}_n^\partial)$$

coming from the inclusion $X^*_{\bullet,\bullet,0} \to X^*_{\bullet,\bullet,\bullet}$ *induces the canonical quotient map*

$$\mathrm{SK}_n^\partial \to \overline{\mathrm{SK}}_n^\partial$$

on the zeroth homotopy groups.

The paper is organized as follows. In Section 2 we introduce an appropriate trivial boundary cobordism relation and corresponding cobordism groups Ω_n^∂ for smooth compact oriented n-manifolds with boundary. Then we define $\overline{\mathrm{SK}}^\partial$-groups ("cut and paste cobordism groups") for manifolds with boundary. We relate these new groups to the classical ones for closed manifolds via exact sequences. In Section 3 we construct a quadrisimplicial set $X^*_{\bullet,\bullet,\bullet}$ in the spirit of the construction of [3], which gives rise to a spectrum $K^{\overline{\square}}(\overline{\mathrm{Mfd}}_n^\partial)$. In Section 5 we verify that $|X^*_{\bullet,\bullet,\bullet}|$ is indeed an infinite loop space. In Section 4 we prove Theorem A, and finally in Section 6 we prove Theorem B.

Conventions All manifolds in this paper will be smooth, oriented and compact. All maps between manifolds/cobordisms will be smooth and orientation preserving.

Acknowledgements We would like to thank Jonathan Campbell, Jim Davis, Johannes Ebert, Fabian Hebestreit, Manuel Krannich, Achim Krause, Wolfgang Lück, Mona Merling, Thomas Nikolaus, Oscar Randal-Williams, George Raptis, Jens Reinhold, Jan Steinebrunner, Wolfgang Steimle and Inna Zakharevich for helpful discussions. We would also like to thank the referee for very helpful comments We thank the Mathematics and Statistics Department at Loyola Chicago for hosting us. This research was supported through the program "Research in Pairs" by the Mathematisches Forschungsinstitut Oberwolfach in 2020 and 2022, and through the program "Summer Research in Mathematics" by the Mathematical Sciences Research Institute. Hoekzema was supported by the Emerson Collective and by the Dutch Research Council (NWO) through the grant VI.Veni.212.170. Rovi was supported by a summer research grant from Loyola University Chicago and by MPIM Bonn. Semikina was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics–Geometry–Structure.

2 Cobordism cut and paste groups of manifolds with boundary

2.1 Classical $\overline{\mathrm{SK}}$ groups

The cut and paste group SK_n and the cobordism group Ω_n are both formed by quotienting the monoid of manifolds under disjoint union by the cut and paste relation in the first case, and by the cobordism relation in the second. There are no natural maps from one group to the other in either direction. They can be compared by quotienting the monoid of manifolds by both relations, forming the cobordism cut and paste group $\overline{\mathrm{SK}}_n$. The following pair of short exact sequences from [8] describe the kernels of the two quotients:

$$F_n \to \Omega_n \to \overline{\mathrm{SK}}_n, \tag{1}$$

$$I_n \to \mathrm{SK}_n \to \overline{\mathrm{SK}}_n. \tag{2}$$

In even dimensions I_n is the subgroup of SK_n generated by spheres S^n. In odd dimensions I_n is zero. The group $F_n \subseteq \Omega_n$ is the subgroup of bordism classes of closed manifolds that fiber over S^1. It was shown in [8] that the only cut and paste invariants of oriented manifolds are the Euler characteristic χ and the signature σ. Noting that σ and χ have the same parity, we have

$$\mathrm{SK}_n \cong \begin{cases} 0 & \text{for } n \text{ odd}, \\ \mathbb{Z} & \text{for } n \equiv 2 \pmod 4 \text{ given by } \frac{1}{2}\chi, \\ \mathbb{Z}^2 & \text{for } n \equiv 0 \pmod 4 \text{ given by } \left(\frac{1}{2}(\chi-\sigma), \sigma\right). \end{cases}$$

In Ω_n, two manifolds are in the same class if and only if their Stiefel–Whitney and Pontrjagin numbers agree. In particular the signature, which is expressed as a linear combination of Pontrjagin numbers by the Hirzebruch signature theorem, is a cobordism invariant, as well as the Euler characteristic modulo 2 given by the top-dimensional Stiefel–Whitney number. In the cobordism cut and paste group only the signature remains:

$$\overline{\mathrm{SK}}_n \cong \begin{cases} 0 & \text{for } n \not\equiv 0 \pmod 4, \\ \mathbb{Z} & \text{for } n \equiv 0 \pmod 4 \text{ given by } \sigma. \end{cases}$$

2.2 Cobordism with trivial boundary

Definition 2.2.1 Let $\mathcal{M}_n^\partial$ be the monoid of diffeomorphism classes of smooth compact oriented n-dimensional manifolds with boundary under disjoint union. We define the *trivial boundary cobordism monoid* $\widetilde{\Omega}_n^\partial$ to be $\mathcal{M}_n^\partial$ quotiented by the following relation. Two classes of manifolds $[M]$ and $[M']$ with boundary are equivalent if there exists a cobordism W with boundary and corners, where the boundary consists of three parts: one that is diffeomorphic to M, one that is diffeomorphic to $\overline{M}'$ and one that is diffeomorphic to $\partial M \times I \cong \partial\overline{M}' \times I$. Recall that $\overline{M}'$ stands for manifold M' with an opposite orientation.

Definition 2.2.2 The *trivial boundary cobordism group* Ω_n^∂ is defined to be the group completion of $\widetilde{\Omega}_n^\partial$.

Remark 2.2.3 The trivial boundary cobordism monoid $\widetilde{\Omega}_n^\partial$ is not cancellative.

A trivial boundary cobordism corresponds to the trace of a sequence of surgery operations applied to the interior of the ingoing manifold, leaving its boundary unchanged. On closed manifolds of dimension $4k$, a surgery operation leaves the signature invariant. The fact that this is also true for a surgery on the interior of a manifold with boundary follows from applying Wall's nonadditivity theorem [9; 14] to the gluing of manifolds with boundary along entire connected components of the boundary. Wall's nonadditivity of the signature features a correction term involving codimension-2 data that is empty in this situation. The theorem therefore reduces to a version of Novikov additivity [1, Proposition 7.1] for manifolds with boundary, provided that the inclusion maps of the two pieces take boundary components entirely to the boundary or to the interior (the maps should be a cut and paste embeddings as defined below in Definition 3.2.1). Hence the signature for manifolds with boundary is an invariant of trivial boundary cobordism.

Lemma 2.2.4 *For every $n \geq 1$, Ω_n^∂ fits in the following exact sequence of abelian groups*:

$$0 \to F_n \xrightarrow{i} \Omega_n \xrightarrow[[M]\mapsto[M]]{\alpha} \Omega_n^\partial \xrightarrow[[N]\mapsto[\partial N]]{\beta} C_{n-1} \to 0.$$

Here C_{n-1} is the group completion of the monoid of orientation-preserving diffeomorphism classes of bounding $(n-1)$-manifolds under disjoint union, which is a free abelian group, and F_n is the subgroup of Ω_n generated by mapping tori.

Proof Two closed manifolds that are cobordant in Ω_n stay cobordant in Ω_n^∂, and hence the map α taking a cobordism class of closed manifolds to the class containing the same manifolds in Ω_n^∂ is well defined. The equivalence relation in Ω_n^∂ preserves the diffeomorphism class of the boundary, and therefore the map β taking a class of manifolds to the diffeomorphism class of the boundary is well defined.

We show exactness at Ω_n. A manifold that can be written as a twisted double $M \cup_\phi \overline{M}$ is zero in Ω_n^∂ as we have the relation

$$[M \cup_\phi \overline{M}] + [M] = [M],$$

as illustrated in Figure 1, if we set N to equal M. It was shown in [15] that any twisted double $M \cup_\phi \overline{M}$ is canonically cobordant to the mapping torus of ϕ; hence F_n is in the kernel of α. Consider $[M]$ an element in $\ker\alpha$, where M is a closed manifold. The signature for manifolds with boundary restricts to the ordinary signature on closed manifolds, and therefore $[M] \in \mathrm{Gr}(\mathcal{M}_n)$ has signature zero. It follows that $[M]$ lies in the image of F_n by [8, Theorem 1.3].

Every class in C_{n-1} is the boundary of an n-dimensional manifold; hence $\tilde{\beta}\colon \widetilde{\Omega}_n^\partial \to \widetilde{C}_{n-1}$ is surjective, and therefore so is the corresponding map on group completions, which is β. We show exactness at Ω_n^∂. It is clear that $\mathrm{Im}\,\alpha \subseteq \ker\beta$. Let $x \in \ker\beta$. We can write $x = [M]-[N]$, where M and N are compact smooth oriented n-manifolds with boundary. As $\beta(x) = 0$, we have that ∂M and ∂N are diffeomorphic. Choose a diffeomorphism $\phi\colon \partial M \to \partial N$. Let us view a cylinder on $M \cup_\phi \overline{N}$ as a cobordism from $M \cup_\phi \overline{N}$ to itself.

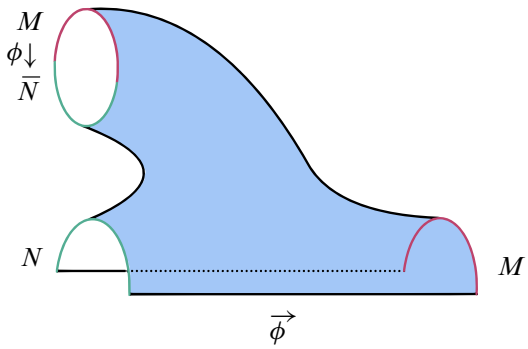


Figure 1: Relation in Ω^{∂}.

Consider the outgoing copy of $M \cup_{\phi} \overline{N}$ and choose a one-sided collar in M of the glued boundary ∂M. This gives the decomposition of the outgoing boundary of the cylinder as $M \cup (\partial M \times I) \cup_{\phi} \overline{N}$. Finally, this can be viewed as a trivial boundary cobordism from $(M \cup_{\phi} \overline{N}) \sqcup N$ to M, where the collar is used to produce the boundary piece of cobordism diffeomorphic to $\partial M \times I$. For illustration see again Figure 1. The following relation holds in $\widetilde{\Omega}_n^{\partial}$:

$$[M] = [M \cup_{\phi} \overline{N}] + [N].$$

Therefore in Ω_n^{∂} we have

$$x = [M] - [N] = [M \cup_{\phi} \overline{N}] \in \operatorname{Im} \alpha. \qquad \square$$

2.3 $\overline{\mathrm{SK}}$ groups of manifolds with boundary

We first recall the definition of SK^{∂}-equivalence from [7]. A cut and paste operation for manifolds with boundary is defined as follows: we cut an n-dimensional manifold M along a codimension-1 smooth separating submanifold Σ with trivial normal bundle that is disjoint from the boundary of M. We paste back the two pieces together along an orientation-preserving diffeomorphism of Σ. In particular, all the cutting boundaries are pasted back together, and the existing boundary of M is left unchanged. We call two manifolds with boundary SK^{∂}-equivalent if they can be obtained from one another through a finite sequence of cut and paste operations.

In other words, given compact oriented manifolds M_1 and M_2, closed submanifolds $\Sigma \subseteq \partial M_1$ and $\Sigma' \subseteq \partial M_2$, and orientation-preserving diffeomorphisms $\phi, \psi : \Sigma \to \Sigma'$, we set

$$M_1 \cup_{\phi} \overline{M}_2 \sim_{\mathrm{SK}^{\partial}} M_1 \cup_{\psi} \overline{M}_2.$$

Definition 2.3.1 The *cobordism cut and paste group of manifolds with boundary* $\overline{\mathrm{SK}}_n^{\partial}$ is defined as the Grothendieck group of diffeomorphism classes of n-dimensional manifolds with boundary where addition is given by disjoint union, $\mathrm{Gr}(\mathcal{M}_n^{\partial})$, quotiented by the relation of cobordism with trivial boundary as well as the relation of cutting and pasting of manifolds with boundary.

Lemma 2.3.2 *The cobordism cut and paste group of manifolds with boundary $\overline{\mathrm{SK}}_n^\partial$ is isomorphic to the trivial boundary cobordism group Ω_n^∂.*

Proof Consider the diagram

$$\begin{array}{ccccc} \Omega_n/F_n & \xhookrightarrow{\overline{\alpha}} & \Omega_n^\partial & \xrightarrow{\beta} & C_{n-1} \\ \cong\downarrow\gamma & & \downarrow & & \downarrow\mathrm{id} \\ \overline{\mathrm{SK}}_n & \xrightarrow{\alpha'} & \overline{\mathrm{SK}}_n^\partial & \xrightarrow{\beta'} & C_{n-1} \end{array}$$

where α' sends a cobordism cut and paste class of closed manifolds to their class as manifolds with boundary, and β' takes a class of manifolds to their boundary, which is preserved up to diffeomorphism by both cut and paste and cobordism relations. The map γ is the canonical isomorphism coming from (1) and the map from Ω_n^∂ to $\overline{\mathrm{SK}}_n^\partial$ is the canonical surjection. Both squares clearly commute.

The top line is a short exact sequence by Lemma 2.2.4. As the cut and paste relation for manifolds with boundary preserves signature of manifolds with boundary, the signature map is a splitting for α', as it is for $\overline{\alpha}$. Hence α' is injective. Both the cut and paste and trivial boundary cobordism relations preserve the boundary up to diffeomorphism; hence β' is surjective. Let x be an element in the kernel of β', and $\bar{x}$ a preimage of x in Ω_n^∂. Then $\bar{x}$ is in the kernel of β; hence we can find an element y in Ω_n/F_n that maps to it. Hence $x=\alpha'(\gamma(y))$, so the lower sequence is exact. By the five lemma, the middle groups are isomorphic. □

2.4 Fibrations of cobordism categories

We can also consider trivial boundary cobordism from the point of view of a fiber sequence of cobordism categories that includes into Genauer's fiber sequence for the cobordism category of manifolds with boundary [5; 6]. We will make this precise in the proposition below. Let $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}}$ be the weakly unital topological cobordism category with boundary, with objects n-dimensional manifolds with boundary and morphisms given by $(n+1)$-dimensional cobordisms with boundary, as defined in [13]. As a subcategory of $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}}$, we consider the weakly unital category $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}}$ of n-dimensional manifolds with boundary, with morphisms given by $(n+1)$-dimensional cobordisms that are trivial on the boundary of the objects, meaning that in the cobordism direction, their boundary is diffeomorphic to $\partial M\times[0,1]$ if M is the source (or equivalently target) object. The boundary map from $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}}$ to Cob_n in [13] restricted to $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}}$ then has image given by the subcategory $\mathrm{Triv}_n^{\varnothing}$ defined below.

Definition 2.4.1 Let ∂M_0 and ∂M_1 be nullbordant $(n-1)$-dimensional manifolds. We call a trivial cobordism (a cobordism that is diffeomorphic to a cylinder) from ∂M_0 to ∂M_1 *extendable* if there are nullbordisms M_0 and M_1 of ∂M_0 and ∂M_1, respectively, and an $(n+1)$-dimensional trivial boundary cobordism W with $M_0=\partial W_{\mathrm{in}}$ and $M_1=\partial W_{\mathrm{out}}$, where ∂W_{in} is the incoming boundary and $\partial W_{\mathrm{out}}$ is the outgoing boundary of W (see Figure 2).

The category $\mathrm{Triv}_n^{\varnothing}$ has objects $(n-1)$-dimensional nullbordant manifolds and morphisms n-dimensional extendable trivial cobordisms. This is a subcategory of $\mathrm{Cob}_n^{\varnothing}$, the full subcategory on bounding manifolds inside Cob_n.

Proposition 2.4.2 *The following diagram is a pullback diagram of weakly unital locally fibrant topological categories and weakly unital continuous functors*:

$$\begin{array}{ccc} \mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}} & \xrightarrow{\partial} & \mathrm{Triv}_n^{\varnothing} \\ \downarrow & & \downarrow \\ \mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}} & \xrightarrow{\partial} & \mathrm{Cob}_n^{\varnothing} \end{array}$$

Proof First note that $\mathrm{Triv}_n^{\varnothing}$ is weakly unital in the sense of [13, Section 2], and the inclusion into $\mathrm{Cob}_n^{\varnothing}$ is weakly unital continuous. The morphism space of the category $\mathrm{Triv}_n^{\varnothing}$ consists of a union of connected components of the morphism space of $\mathrm{Cob}_n^{\varnothing}$, and hence it is locally fibrant in the sense of [13] as Cob_n is.

On the level of objects, we have the following diagram, which is clearly a pullback:

$$\begin{array}{ccc} \{n\text{-dimensional manifolds with boundary}\} & \xrightarrow{\partial} & \{(n-1)\text{-dimensional closed nullbordant manifolds}\} \\ \downarrow{\scriptstyle \mathrm{id}} & & \downarrow{\scriptstyle \mathrm{id}} \\ \{n\text{-dimensional manifolds with boundary}\} & \xrightarrow{\partial} & \{(n-1)\text{-dimensional closed nullbordant manifolds}\} \end{array}$$

The morphism spaces of the above diagram of categories also form a pullback diagram. It is enough to check the pullback property for the morphisms spaces between any two $(n+1)$-dimensional manifolds M and M'. In this case, the diagram is

$$\begin{array}{ccc} \text{Space of cobordisms between } M \text{ and } M' \text{ that} \\ \text{have boundary diffeomorphic to } \partial M \times [0,1] & \xrightarrow{\partial} & \text{Space of extendable trivial cobordisms} \\ & & \text{from } \partial M \text{ to } \partial M' \\ \downarrow{\scriptstyle \mathrm{incl}} & & \downarrow{\scriptstyle \mathrm{incl}} \\ \text{Space of cobordisms with boundary (in} \\ \text{Genauer's sense) between } M \text{ and } M' & \xrightarrow{\partial} & \text{Space of cobordisms} \\ & & \text{from } \partial M \text{ to } \partial M' \end{array}$$

This is indeed a pullback diagram in the category of topological spaces. □

We apply [13, Theorem 2.3] to the pullback square of categories in Proposition 2.4.2. By [13, Lemma 4.4], the boundary map from $\mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}}$ to Cob_n is a level cartesian and cocartesian fibration. We hence obtain a homotopy pullback square of classifying spaces

$$\begin{array}{ccc} B\mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}} & \xrightarrow{B\partial} & B\mathrm{Triv}_n^{\varnothing} \\ \downarrow & & \downarrow \\ B\mathrm{Cob}_{n+1}^{\partial_{\mathrm{all}}} & \xrightarrow{B\partial} & B\mathrm{Cob}_n^{\varnothing} \end{array}$$

The lower line of this diagram is the well-known Genauer fibration with fiber $B\mathrm{Cob}_{n+1}$. Therefore the upper line is also a fibration with homotopy equivalent fiber:

$$B\mathrm{Cob}_{n+1} \to B\mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}} \to B\mathrm{Triv}_n^{\varnothing}.$$

By construction this fibration is surjective.

The fibration above gives rise to the long exact sequence in homotopy

$$\pi_1 B\mathrm{Cob}_{n+1} \to \pi_1 B\mathrm{Cob}_{n+1}^{\partial_{\mathrm{triv}}} \to \pi_1 B\mathrm{Triv}_n^{\varnothing} \to \Omega_n \to \widetilde{\Omega}_n^{\partial} \to \widetilde{D}_{n-1} \to 0,$$

where $\widetilde{D}_{n-1}$ is the monoid of nullbordant oriented $(n-1)$-dimensional manifolds up to extendable trivial cobordism and $\widetilde{\Omega}_n^{\partial}$ is the trivial boundary cobordism monoid. On π_0 we have based the sets at $\varnothing$ to define the exact sequence, and note that the maps are also monoid maps. The group $\pi_1 B\mathrm{Cob}_{n+1}$ is also known as the controlled cut and paste group SKK_{n+1}; see for example [8]. It is clear that the map of monoids $\Omega_n \to \widetilde{\Omega}_n^{\partial}$ is injective; hence we have the last three terms as a short exact sequence of monoids.

3 Cubes of manifolds with boundary

In what follows, we will define a trisimplicial set $X^0_{\bullet,\bullet,\bullet}$ of cubes of manifolds. Two of three simplicial directions are defined the same way as in the bisimplicial construction used in the K-theory of manifolds from [7]. These two directions encode the cut and paste data. The third direction corresponds to cobordisms that fit into our description of trivial boundary cobordism (See Section 2.2 and Figure 2).

On the faces in the cobordism direction we have a cut and paste square of the cobordisms themselves that extends the cut and paste squares of manifolds on the front and the back of the cube. The cobordisms with boundary will have collars both in the composition direction (of width $\varepsilon_c \to 0$) as well as in the trivial cobordism boundary direction (of width $\varepsilon_\partial \to 0$).

Instead of directly giving this trisimplicial set a topology, we add a fourth simplicial direction encoding topology and view the initial trisimplicial set as a value at [0] of a simplicial object in the category of trisimplicial sets. This approach was introduced in [11] to define parametrized cobordism categories and is also applied in [13, Section 4]. In the new simplicial direction in level r, we will have $(0,0,0)$-simplices

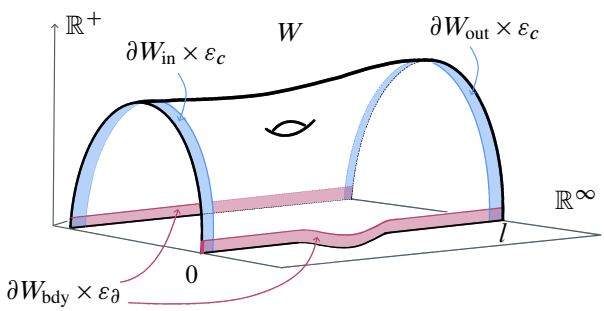


Figure 2: The ε_c and ε_∂ collars on a cobordism with trivial boundary.

given by fiber bundles $E \to \Delta^r$, with fibers being manifolds with boundary and generic trisimplicial datum described further.

3.1 The simplicial set $X^{*}_{0,0,0}$ of manifolds with boundary

Definition 3.1.1 Let M be a smooth oriented submanifold of $\mathbb{R}_+ \times \mathbb{R}^\infty$, with $\partial M = M \cap (\{0\} \times \mathbb{R}^\infty)$. Then we say that M is *neat* (in the sense of [6; 13]) if it is cylindrical near the boundary, ie there exists an $\varepsilon_\partial > 0$ such that

$$M \cap ([0, \varepsilon_\partial) \times \mathbb{R}^\infty) = [0, \varepsilon_\partial) \times \partial M.$$

We say an *embedding* e of ε_∂-neat manifold M into $\mathbb{R}_+ \times \mathbb{R}^\infty$ is ε_∂-*neat* if the following diagram commutes:

$$\begin{array}{ccc} [0,\varepsilon_\partial) \times \partial M & \xhookrightarrow{\text{incl}} & M \\ \downarrow{\scriptstyle \mathrm{id} \times e|_{\partial M}} & & \downarrow{\scriptstyle e} \\ [0,\varepsilon_\partial) \times \mathbb{R}^\infty & \hookrightarrow & \mathbb{R}^\infty \times \mathbb{R}_+ \end{array}$$

We now define the simplicial set $X^{*}_{0,0,0}$ to be, on level $* = r$, the set of fiber bundles $E \mapsto \Delta^r$ with fibers being smooth manifolds with boundary, which is fiberwise ε_∂-neatly embedded into $\Delta^r \times \mathbb{R}_+ \times \mathbb{R}^\infty$ for some $\varepsilon_\partial > 0$.

3.2 The simplicial sets $X^{*}_{1,0,0}$ and $X^{*}_{0,1,0}$ of SK-embeddings

Definition 3.2.1 Let M and N be neat n-dimensional manifolds. We define a *cut and paste embedding* or SK-*embedding* from M to N to be a smooth orientation-preserving embedding of manifolds (not necessarily neat) that satisfies the following conditions. Each boundary component of M maps either entirely to ∂N, or entirely to the interior of N as a submanifold with trivial normal bundle.

For $E, E' \in X^r_{0,0,0}$ manifold bundles over Δ^r, let SK-Emb(E, E') denote the set of bundle maps from E to E' which are fiberwise SK-embeddings, which we refer to as *bundle* SK-*embeddings*. We define the sets $X^r_{1,0,0}$ and $X^r_{0,1,0}$ to be all bundle SK-embeddings between the corresponding source and target, ie

$$X^r_{1,0,0} = X^r_{0,1,0} := \bigsqcup_{X^r_{0,0,0} \times X^r_{0,0,0}} \text{SK-Emb}(E, E').$$

3.3 The simplicial set $X^{*}_{0,0,1}$ of cobordisms with trivial boundary

In the third simplicial direction, (0,0,1)-simplices are given by trivial boundary cobordisms. Given two ε_∂-neat manifolds $M, M' \in X_{0,0,0}$, a $(0,0,1)$-simplex between them (ie a simplex with a source M and target M') is given by a neat cobordism defined as follows. For $l > 0$ and $\varepsilon_c, \varepsilon_\partial > 0$, an $(\varepsilon_c, \varepsilon_\partial)$-*neat* cobordism (l, W) from M to M' is an $(n+1)$-dimensional submanifold (with corners) of $[0, l] \times \mathbb{R}_+ \times \mathbb{R}^\infty$, with boundary consisting of three pieces $\partial W_{\text{in}} = M$, $\partial W_{\text{out}} = M'$ and ∂W_{bdy}, that are the intersections of W with $\{0\} \times \mathbb{R}_+ \times \mathbb{R}^\infty$, $\{l\} \times \mathbb{R}_+ \times \mathbb{R}^\infty$ and $[0, l] \times \{0\} \times \mathbb{R}^\infty$, respectively, adhering to the following

conditions. We have that W is ε_c-cylindrical (c for composition) near the in- and out-components of the boundary, ie

$$W \cap ([0,\varepsilon_c)\times\mathbb{R}_+\times\mathbb{R}^\infty) = [0,\varepsilon_c)\times M, \quad W \cap ((l-\varepsilon_c, l]\times\mathbb{R}_+\times\mathbb{R}^\infty) = (l-\varepsilon_c, l]\times M'.$$

Moreover W is ε_∂-cylindrical near the ∂W_{bdy} component

$$\tau_{(12)}(W \cap ([0,l]\times[0,\varepsilon_\partial]\times\mathbb{R}^\infty)) = [0,\varepsilon_\partial]\times\partial W_{\text{bdy}},$$

where $\tau_{(12)}$ swaps the first two coordinates. Lastly, the cobordism is required to be diffeomorphic to a cylinder on ∂W_{bdy}, ie

$$\partial W_{\text{bdy}} = W|_{[0,l]\times\{0\}\times\mathbb{R}^\infty} \cong [0,l]\times\partial M.$$

We now define the set $\overline{X}^*_{0,0,1}$ to be, on level $*=r$, a map $l\colon\Delta^r\to(0,\infty)$ and a fiber bundle over Δ^r with fibers compact, smooth $n+1$ trivial boundary cobordisms fiberwise $(\varepsilon_c,\varepsilon_\partial)$-neatly embedded into $\Delta^r\times[0,l]\times\mathbb{R}_+\times\mathbb{R}^\infty$ for some $\varepsilon_c>0$ and $\varepsilon_\partial>0$.

We then define

$$X^r_{0,0,1} = X^r_{0,0,0} \sqcup \overline{X}^*_{0,0,1}.$$

The first term can be thought of as cobordisms of length zero, which are the image of the degeneracy map from $X_{0,0,0}$ to $X_{0,0,1}$. The second summand encodes cobordism bundles of nonzero length.

3.4 The simplicial set $X^*_{1,1,0}$ of cut and paste squares

This set is given by the set of cut and paste squares that are distinguished squares of the category with squares Mfd^∂_n from [7]. Recall that those squares are given by the commutative diagrams

$$\begin{array}{ccc} M_0 & \hookrightarrow & M_1 \\ \downarrow & & \downarrow \\ M_2 & \hookrightarrow & M_1\cup_{M_0} M_2 \end{array}$$

where maps are SK-embeddings and the diagram is a pushout square in the category of spaces, ie the pushout $M_1\cup_{M_0}M_2$ is a smooth manifold.

On the simplicial level $*=r$, the set $X^r_{1,1,0}$ is given by the set of squares of manifold bundles over Δ^r, ie the set of diagrams of total spaces of bundles

$$\begin{array}{ccc} E_0 & \hookrightarrow & E_1 \\ \downarrow & & \downarrow \\ E_2 & \hookrightarrow & E_3 \end{array}$$

which over every point of Δ^r restrict to the distinguished squares of manifolds as above.

3.5 The simplicial sets $X^*_{1,0,1}$ and $X^*_{0,1,1}$ of SK-embeddings of cobordisms

Analogously to how we defined SK-embeddings of manifolds with boundary, we now define SK-embeddings of neat cobordisms.

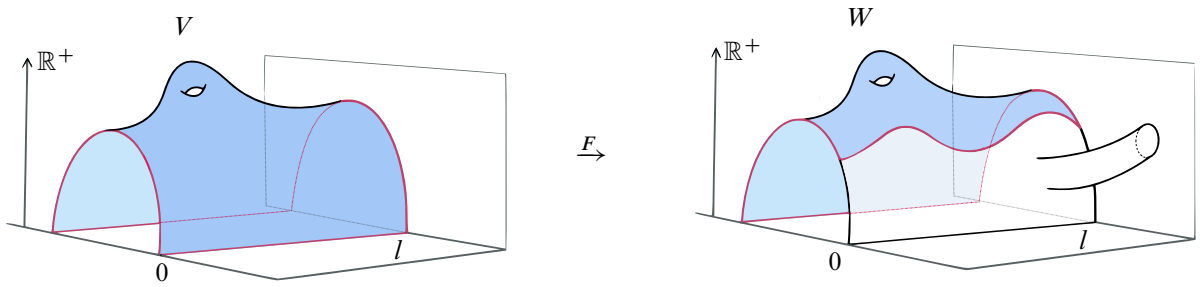


Figure 3: SK embedding of cobordisms.

Definition 3.5.1 Let V and W be neat $(n+1)$-cobordisms in $[0,l]\times\mathbb{R}_+\times\mathbb{R}^\infty$. An SK-embedding F from V to W is a smooth orientation-preserving embedding such that the following holds.

(i) F sends ∂V_{in} and ∂V_{out} to ∂W_{in} and ∂W_{out}, correspondingly; the restriction maps $F_{|\partial V_{\text{in}}}$ are $F_{|\partial V_{\text{out}}}$ are SK-embeddings of neat manifolds with boundary.

(ii) F preserves the collars at in- and outgoing boundary components, ie F is cylindrical near ∂V_{in} and ∂V_{out}.

(iii) F sends every component of ∂V_{bdy} either entirely to ∂W_{bdy} or entirely (apart from the corner points) to the interior of W, with the image being a submanifold of W with trivial normal bundle.

Figure 3 shows an example of an SK-embedding of cobordisms.

For $D, D'\in X^r_{0,0,1}$ trivial cobordism bundles over Δ^r, let SK-Emb(D,D') denote the set of bundle maps from D to D' which are fiberwise SK-embeddings of trivial boundary cobordisms, preserving the length l fiberwise. We define the sets $X^r_{1,0,1}$ and $X^r_{0,1,1}$ to be all bundle SK-embeddings between the corresponding source and target trivial cobordism bundles, ie

$$X^r_{0,1,1}=X^r_{1,0,1}:=\coprod_{E,E'\in X^r_{0,0,0}}\text{SK-Emb}(E,E')\sqcup\coprod_{D,D'\in X^r_{0,0,1}}\text{SK-Emb}(D,D').$$

3.6 The simplicial set $X^*_{1,1,1}$ of cubes of manifolds

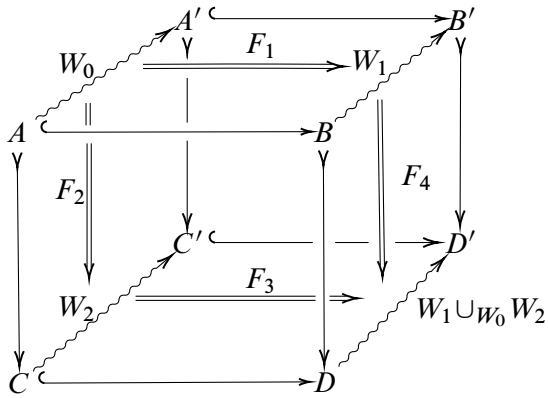
We define a cut and paste square of trivial boundary cobordism as a quadruple (F_1,F_2,F_3,F_4), where F_i are SK-embeddings of cobordisms of the same length that fit into a commutative square, which is a pushout square on the level of topological spaces

$$\begin{array}{ccc} W_0 & \xrightarrow{F_1} & W_1 \\ \downarrow{\scriptstyle F_2} & & \downarrow{\scriptstyle F_4} \\ W_2 & \xrightarrow{F_3} & W_1\cup_{W_0}W_2 \end{array}$$

On the simplicial level $* = r$, the set $X^r_{1,1,1}$ is then given by the set of squares of cobordism bundles over Δ^r, ie the set of diagrams

$$\begin{array}{ccc} D_0 & \hookrightarrow & D_1 \\ \downarrow & & \downarrow \\ D_2 & \hookrightarrow & D_3 \end{array}$$

such that over each point of Δ^r this restricts to a cut and paste square of trivial boundary cobordisms as above. Fiberwise we have a cube of manifolds, trivial boundary cobordisms and SK-embeddings of cobordisms extending the SK-embeddings on manifolds at the in- and outgoing boundary, as illustrated below:



Note that the cubes compose in each of the three directions. The horizontal and vertical stacking of cubes is given by composing the SK-embeddings of cobordisms fiberwise. In the cobordism direction, the fiberwise composition of two cobordisms bundles with corresponding manifold bundles on the boundary is defined to be a cobordism bundle over Δ^r where the maps l and l' are added fiberwise and the cobordisms are fiberwise composed along the common boundary over a point in Δ^r.

3.7 General construction of $X^{*}_{\bullet,\bullet,\bullet}$

Now we use elements of $X^r_{a,b,c}$ with $a, b, c \in \{0, 1\}$ as building blocks to define the simplices of arbitrary size. Any admissible composition of the building blocks over Δ^r that form a diagram of size $k \times l \times m$ with $k, l, m \in \mathbb{Z}_{\geq 0}$ represents an element in $X^r_{k,l,m}$. Note that for $k, l, m \geq 1$, the set $X^r_{k,l,m}$ of all (k, l, m)-simplices over Δ^r is given by a set of all "cubical" diagrams of size $k \times l \times m$, consisting of klm copies of size-1 cubes fibered over Δ^r.

4 Recovering cobordism cut and paste groups as $K^{\square}_0$

In this section we build a spectrum out of the quadrisimplicial set $X^{*}_{\bullet,\bullet,\bullet}$ and show it recovers the cobordism cut and paste group with boundary $\overline{\mathrm{SK}}^{\partial}_n$ on π_0. The proof of the latter (Theorem 4.3) is a generalization of [3, Theorem 3.1], but requires more complicated simplicial combinatorics to deal with the additional

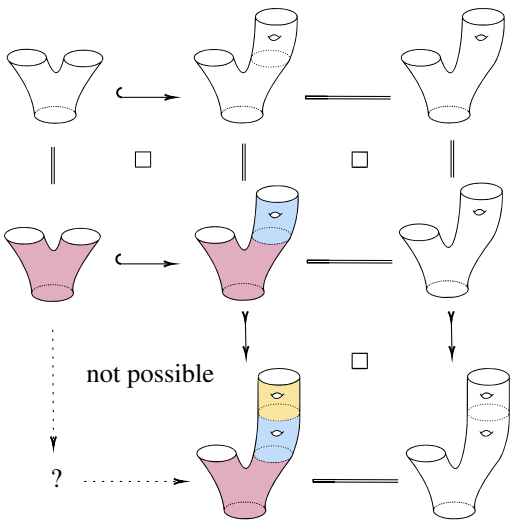


Figure 4: Composing squares (or cubes) that coincide in only one vertex is not always possible

simplicial directions. We define $X_\bullet$ to be the diagonal of the quadrisimplicial set. Note that for a quadrisimplicial set, the realization of the diagonal simplicial set is homotopy equivalent to the realization of the quadrisimplicial set itself

$$|X_\bullet| \simeq |X^*_{\bullet,\bullet,\bullet}|.$$

Remark 4.1 Note that $X_\bullet$ cannot be interpreted as the nerve of a topological category as there is no natural way to define a composition map

$$X_1 \times_{X_0} X_1 \to X_2.$$

Given two cubes over Δ^1 that agree on the appropriate vertex, it is not always possible to find a $2\times2\times2$ cube over Δ^2 incorporating the two. This also holds on the level of squares of SK-embeddings in $X^*_{\bullet,\bullet,0}$. For an example, see Figure 4.

Definition 4.2 We define

$$K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n) := \Omega_\varnothing |X^*_{\bullet,\bullet,\bullet}| \simeq \Omega_\varnothing |X_\bullet|.$$

In the proofs below we will use a bracket to denote 1-simplices (cubes over Δ^1) and 2-simplices ($2\times2\times2$ cubes over Δ^2) that are generated from the diagram inside brackets by extending with degeneracies in all other directions. For example

$$[A \hookrightarrow B] := \begin{array}{ccc} A & \hookrightarrow & B \\ \| & & \| \\ A & \hookrightarrow & B \end{array} \qquad \text{and} \qquad \left[\begin{array}{ccc} & & A' \\ A & & \downarrow \\ & & C' \\ \downarrow & & \\ C & & \end{array}\right] := \begin{array}{ccc} A' & = & A' \\ \uparrow & & \uparrow \\ A & = & A \\ \downarrow & & \downarrow \\ C' & = & C' \\ \uparrow & & \uparrow \\ C & = & C \end{array}$$

where we think of vertices of the cubes on the right as being the total spaces of fiber bundles over Δ^1, which fiberwise has the shape of the prescribed cube. Note that the diagonal face maps ensure that the source and target of a 1-simplex are single manifolds with boundary, fibered over $\Delta^0 = *$.

We now prove our main theorem, which recovers the cobordism cut and paste groups with boundary as π_0 of the cubes K-theory of manifolds spectrum. The strategy for this proof will start by contracting all vertices to a single vertex in the diagonal of the quadrisimiplicial set $X^*_{\bullet,\bullet,\bullet}$, and subsequently compute π_1 (corresponding to π_0 of the spectrum) using direct combinatorial arguments.

Theorem 4.3 *There is an isomorphism*

$$\pi_0 K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n) \cong \overline{\mathrm{SK}}^{\partial}_n.$$

Proof Let $Y := |X_\bullet|$. We have

$$\pi_0 K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n) = \pi_1 |X^*_{\bullet,\bullet,\bullet}| \cong \pi_1 Y.$$

First we will contract all vertices in Y to a single vertex. Consider the set of 1-simplices of the form $[A \hookrightarrow B]$. For each cube of this shape, there is a canonical 2-simplex of the shape

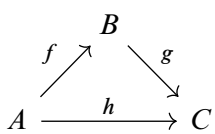
$$\begin{bmatrix} \varnothing & \hookrightarrow & A & \hookrightarrow & B \\ \| & & \| & & \| \\ \varnothing & \hookrightarrow & A & \hookrightarrow & B \\ \| & & \| & & \| \\ \varnothing & \hookrightarrow & A & \hookrightarrow & B \end{bmatrix} \tag{3}$$

where we extend twice in the cobordism direction using identity cobordisms and the induced SK-embeddings on the cobordisms. This 2-simplex has boundary

$$[\varnothing \hookrightarrow A] + [A \hookrightarrow B] - [\varnothing \hookrightarrow B].$$

The subset generated by all 0-simplices together with the 1- and 2-simplices above form a subsimplicial set $X'_\bullet$ of $X_\bullet$, of which the realization Y' is homeomorphic to a cone over all nondegenerate 1-simplices of the shape $[A \hookrightarrow B]$ for $A \neq \varnothing$, with cone point $\varnothing$. As a cone is contractible and Y and Y' form a CW pair, we can perform a topological quotient in Y leaving the homotopy type unchanged. The realization of a quotient of simplicial sets is the quotient of the realizations, so $Y \simeq Y/Y' \cong |X_\bullet / X'_\bullet|$. Compare also with the proof of [3, Theorem 3.1].

We will now compute $\pi_1 Y \cong \pi_1 |X_\bullet / X'_\bullet|$. Note that the fundamental group of a simplicial set with a single vertex has a presentation with generators given by all 1-simplices and relations given by 2-simplices, where a 2-simplex



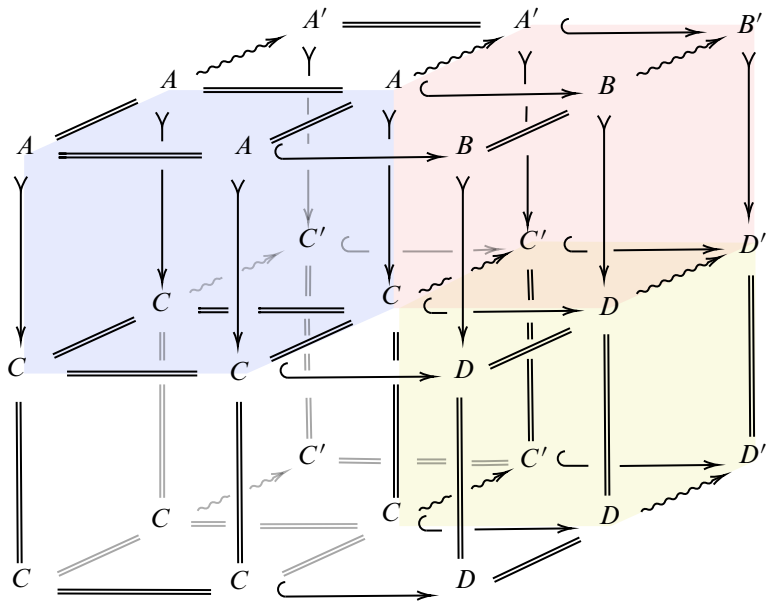


Figure 5

induces the relation $f+g-h=0$. In practice this means that a 2-simplex represented by a $(2\times2\times2)$-cube over Δ^2 induces the relation that the initial plus the final cubes as Δ^1-bundles equals the composite cube over Δ^1 viewed as 1-simplices.

First we will reduce the generators of $\pi_1 Y$ using 2-simplices.

Consider a generic 1-simplex (pictured in red). Extending via identities, we obtain the 2-simplex in Figure 5.

This implies the relation

$$\begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix} + \begin{bmatrix} & & C' & \hookrightarrow & D' \\ & W_2 \nearrow & & & \nearrow W_3 \\ C & & \hookrightarrow & & D \end{bmatrix} = [\text{generic 1-simplex}]. \tag{4}$$

Analogously, we can construct the 2-simplex in Figure 6 and deduce the relation

$$\begin{bmatrix} & & A' & \hookrightarrow & B' \\ & W_0 \nearrow & & & \nearrow W_1 \\ A & & \hookrightarrow & & B \end{bmatrix} + \begin{bmatrix} B' \\ \downarrow \\ D' \end{bmatrix} = [\text{generic 1-simplex}]. \tag{5}$$

Next we consider the 2-simplex in Figure 7, which is obtained by extending the (nongeneric) cube in green. This establishes the relation

$$\begin{bmatrix} & & A' \\ & W_0 \nearrow & \\ A & & \end{bmatrix} + [A' \hookrightarrow B'] = \begin{bmatrix} & & A' & \hookrightarrow & B' \\ & W_0 \nearrow & & & \nearrow W_1 \\ A & & \hookrightarrow & & B \end{bmatrix}$$

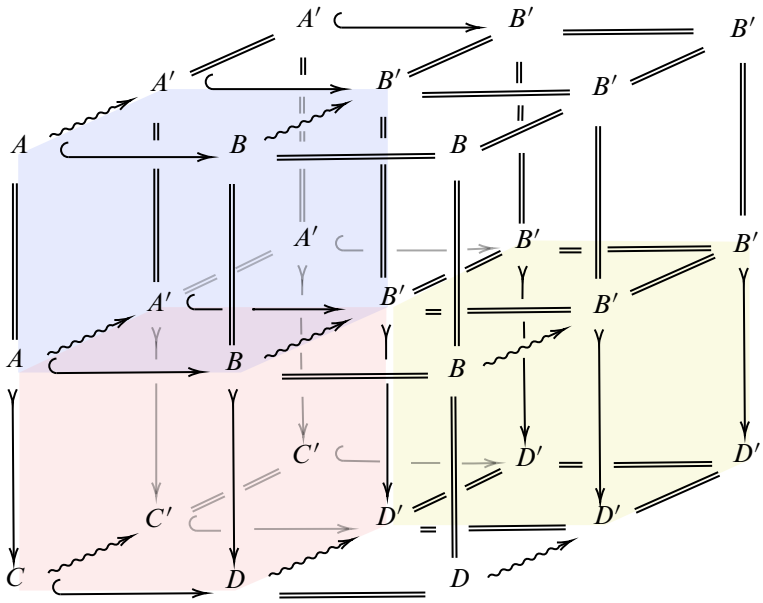


Figure 6

But $[A' \hookrightarrow B']$ was contracted, therefore

(6)
$$\begin{bmatrix} & A' \\ W_0 \nearrow & \\ A & \end{bmatrix} = \begin{bmatrix} A' & \hookrightarrow & B' \\ W_0 \nwarrow & & \nearrow W_1 \\ A & \hookrightarrow & B \end{bmatrix}$$

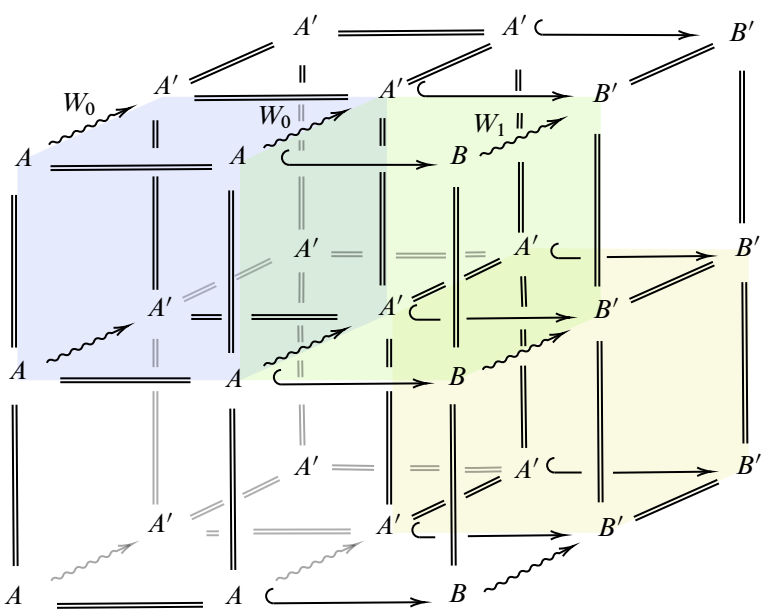


Figure 7

Analogously, one can construct a 2-simplex to obtain the relation

$$[A \hookrightarrow B] + \left[\begin{array}{ccc} & & B' \\ & \overset{W_1}{\nearrow} & \\ B & & \end{array}\right] = \left[\begin{array}{ccccc} & & A' & \hookrightarrow & B' \\ & \overset{W_0}{\nearrow} & & & \overset{W_1}{\nearrow} \\ A & & \hookrightarrow & & B \end{array}\right]$$

so that

$$\left[\begin{array}{ccc} & & B' \\ & \overset{W_1}{\nearrow} & \\ B & & \end{array}\right] = \left[\begin{array}{ccccc} & & A' & \hookrightarrow & B' \\ & \overset{W_0}{\nearrow} & & & \overset{W_1}{\nearrow} \\ A & & \hookrightarrow & & B \end{array}\right] \tag{7}$$

From (6) and (7) we see that cobordisms that fit into a square are in the same class.

We leave out the length of the cobordism in our notation since cobordisms either have constant length 0 over Δ^1, or they have varying nonvanishing length $l\colon \Delta^1 \to (0,\infty)$. The 1-simplices

$$\left[\begin{array}{ccc} & & A \\ & \overset{\mathrm{Id}}{\nearrow} & \\ A & & \end{array}, 0\right]$$

with cobordisms of length 0 are degenerate and these do not contribute any nontrivial loops. Now note that any cobordism W fits into a cube of the shape

$$\left[\begin{array}{ccccc} & & \varnothing & \hookrightarrow & B' \\ & \overset{\varnothing}{\nearrow} & & & \overset{W}{\nearrow} \\ \varnothing & & \hookrightarrow & & B' \end{array}, l\right]$$

This implies that

$$\left[\begin{array}{ccc} & & B' \\ & \overset{W}{\nearrow} & \\ B & & \end{array}, l\right] = \left[\begin{array}{ccc} & & \varnothing \\ & \overset{\varnothing}{\nearrow} & \\ \varnothing & & \end{array}, l\right] \tag{8}$$

Consider any 2-simplex with three sides of the form

$$[l] := \left[\begin{array}{ccc} & & \varnothing \\ & \overset{\varnothing}{\nearrow} & \\ \varnothing & & \end{array}, l\right]$$

where the three functions l are given by

$$\begin{array}{ccccc} & & \varnothing & & \\ & \overset{l_1}{\nearrow} & & \overset{l_3}{\searrow} & \\ \varnothing & & \overset{l_2}{\longrightarrow} & & \varnothing \end{array}$$

Such a 2-simplex exists for nondegenerate sides whenever $l_3(0) > l_1(0)$ and $l_3(1) > l_2(1)$. For instance, we have let l_1 be a generic function $l\colon \Delta^1 \to (0,\infty)$, $l_2 = l(0) + 1 := c$ and $l_3 = 1$, which imposes the relation

$$[l] = [c] + [1].$$

Now consider $l_1 = A$, $l_2 = B$ and $l_3 = C$ with constants $C > A, B$. This imposes the relation

$$[C] = [A] + [B].$$

Taking $A = B = 1$ and $C > 1$ we get $[C] = 2[1]$; thus in particular $[l] = 3[1]$. We also have $[3] = [2] + [1]$ and $[3] = [2] = 2[1]$, so we conclude that $[1] = 0$, and hence $[l] = 0$.

From the above and (8) we deduce that there are no additional generators given by cobordism, since the class of an empty cobordism (of nonzero length) is zero.

Therefore, from (4) and (8), we see that $\pi_1 Y$ is generated by cubes of the form

$$\begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix}$$

Since we established that cubes of the shape

$$\begin{bmatrix} & A' & \hookrightarrow & B' \\ W_0 \nearrow & & & \nearrow W_1 \\ A & & \hookrightarrow & B \end{bmatrix}$$

represent trivial loops, from (4) and (5) we get that

$$[\text{generic 1-simplex}] = \begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix} = \begin{bmatrix} B' \\ \downarrow \\ D' \end{bmatrix} \tag{9}$$

We can deduce from the relations of cubes

$$\partial \begin{bmatrix} A & = & A & \hookrightarrow & B \\ \downarrow & & \downarrow & & \downarrow \\ C & = & C & \hookrightarrow & D \\ \| & & \| & & \| \\ C & = & C & \hookrightarrow & D \end{bmatrix} = \begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix} + [C \hookrightarrow D] - \begin{bmatrix} A & \hookrightarrow & B \\ \downarrow & & \downarrow \\ C & \hookrightarrow & D \end{bmatrix} = 0, \tag{10}$$

$$\partial \begin{bmatrix} A & \hookrightarrow & B & = & B \\ \| & & \| & & \| \\ A & \hookrightarrow & B & = & B \\ \downarrow & & \downarrow & & \downarrow \\ C & \hookrightarrow & D & = & D \end{bmatrix} = [A \hookrightarrow B] + \begin{bmatrix} B \\ \downarrow \\ D \end{bmatrix} - \begin{bmatrix} A & \hookrightarrow & B \\ \downarrow & & \downarrow \\ C & \hookrightarrow & D \end{bmatrix} = 0, \tag{11}$$

that

$$\begin{bmatrix} A & \hookrightarrow & B \\ \downarrow & & \downarrow \\ C & \hookrightarrow & D \end{bmatrix} = \begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix} = \begin{bmatrix} B \\ \downarrow \\ D \end{bmatrix} \tag{12}$$

Note that the composition of embeddings $\varnothing \to A$ and $A \to C$ is an embedding $\varnothing \to C$, and the 2-simplex corresponding to this composition gives the relation

$$\begin{bmatrix} A \\ \downarrow \\ C \end{bmatrix} = \begin{bmatrix} \varnothing \\ \downarrow \\ C \end{bmatrix} - \begin{bmatrix} \varnothing \\ \downarrow \\ A \end{bmatrix} =: [C] - [A]. \tag{13}$$

So we deduce that $\pi_1 Y$ can be generated by 1-simplices of the form

$$\begin{bmatrix} \varnothing \\ \downarrow \\ A \end{bmatrix}$$

that we denote by $[A]$ for A any manifold bundle over Δ^1.

Moreover, whenever A, B, C and D fit into a square

$$\begin{bmatrix} A & \hookrightarrow & B \\ \downarrow & & \downarrow \\ C & \hookrightarrow & D \end{bmatrix}$$

we obtain the square relation

$$[C] - [A] = [D] - [B]. \tag{14}$$

In particular, we have the square

$$\begin{bmatrix} \varnothing & \xhookrightarrow{=} & \varnothing \\ \downarrow & & \downarrow \\ A & \xhookrightarrow{\cong} & A_0 \times [0,1] \end{bmatrix}$$

where $A_0 \times [0,1]$ denotes a bundle over Δ^1 with fiber A_0 that is a trivialization of the bundle A. This shows that any manifold bundle representative can be reduced to a product bundle, which we denote by its fiber, ie $[A_0]$:

$$[A] = [A_0]. \tag{15}$$

We have the pushout square of manifolds

$$\begin{bmatrix} \varnothing & \hookrightarrow & M \\ \downarrow & & \downarrow \\ N & \hookrightarrow & M \sqcup N \end{bmatrix}$$

from which we conclude that

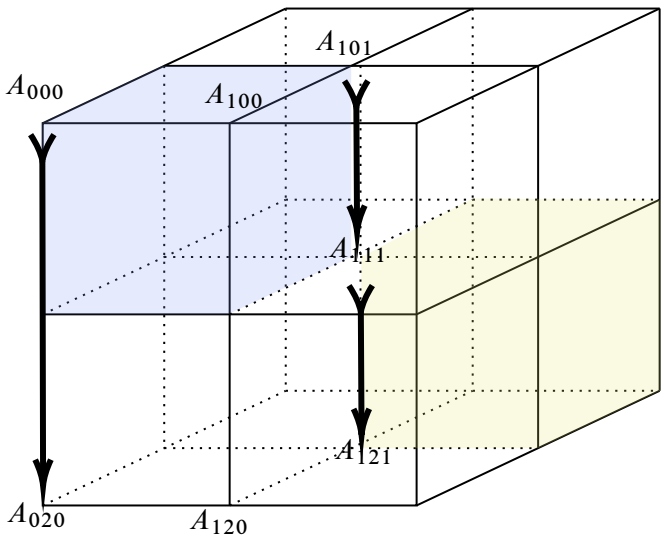
$$[N] = [M \sqcup N] - [M]. \tag{16}$$

Hence the operation in $\pi_1 Y$ corresponds to disjoint union of manifolds.

So far we have established that $\pi_1 Y$ is generated by classes represented by manifolds adhering to the square relation (14). We now establish the full set of relations on the generators.

Consider a generic 2-simplex with vertices labeled as A_{ijk}, where $0 \le i, j, k \le 2$ (see the figure below). From the arguments above, we know that a generic cube can be represented by a cube with one nontrivial

vertical arrow and identities in horizontal directions, where any of the four vertical arrows can be chosen as a representative. This means that the relation arising from a generic 2-simplex can be expressed as (17) as illustrated by



where

$$\begin{bmatrix} A_{000} \\ \downarrow \\ A_{020} \end{bmatrix} = \begin{bmatrix} A_{101} \\ \downarrow \\ A_{111} \end{bmatrix} + \begin{bmatrix} A_{111} \\ \downarrow \\ A_{121} \end{bmatrix}$$

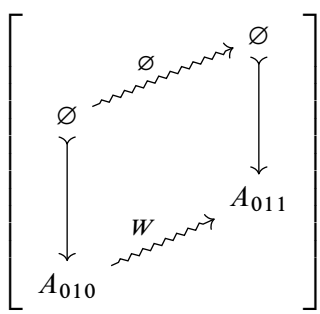
which we can write as

$$[A_{020}] - [A_{000}] = [A_{121}] - [A_{101}]. \tag{17}$$

By (15), we can represent these manifold bundles by a single manifold, for example their fiber over $\{0\}$. Take the lower part of the 2-simplex to be given by the equalities in the vertical direction. Then we also have

$$[A_{010}] - [A_{000}] = [A_{111}] - [A_{101}] = [A_{011}] - [A_{001}], \tag{18}$$

where the last equation follows from (14). Finally consider the cube with $A_{000} = A_{001} = \varnothing$:



We deduce that

$$[A_{010}] - [\varnothing] = [A_{011}] - [\varnothing], \tag{19}$$

and hence we obtain a relation that cobordant manifolds A_{010} and A_{011} represent the same element in $\pi_1 Y$.

The square relation in (14) was shown in [7] to induce the SK^{∂} relation.

At this point we have shown that cobordism and cut and paste relations are relations in $\pi_1 Y$. To show that these are the only relations, we need to check that the generic relation in (17) can be obtained using the cobordism and cut and paste relations.

First note that using the cut and paste relation in a generic cube allows us to write

$$[A_{020}]-[A_{000}]=[A_{120}]-[A_{100}].$$

From the cobordism relation, we know that

$$[A_{120}]-[A_{100}]=[A_{121}]-[A_{101}].$$

Combining these two equations we obtain the generic relation in (17). □

5 $K^{\square}$ is a spectrum

We use Segal's machinery of Γ-spaces to show that the geometric realization of the simplicial space $X_\bullet$ is an infinite loop space, and hence gives an Ω-spectrum. We first briefly recollect the basic definitions of Γ-spaces.

Let Γ^{op} be a skeleton of the category of finite pointed sets. By this we mean that objects in Γ^{op} are given by the pointed sets $n_+ = \{*, 1, \dots, n\}$ for nonnegative integers n and morphisms are pointed maps of sets. The object $0_+ = \{*\}$ is initial and terminal in Γ^{op}. A *Γ-space* is a covariant functor from Γ^{op} to the category of pointed spaces S_*. A functor from Γ^{op} to the category $S^{\Delta^{\mathrm{op}}}$ of simplicial spaces is called a *simplicial Γ-space*.

The i^{th} Segal map is a map $e_i : n_+ \to 1_+$ defined by

$$e_i(j) = \begin{cases} * & \text{if } j \neq i, \\ 1 & \text{if } j = i. \end{cases}$$

A Γ-space $A\colon \Gamma^{\mathrm{op}} \to S_*$ is called *special* if the Segal map

$$\prod_{i=1}^{n} A(e_i)\colon A(n_+) \to \prod_{n} A(1_+)$$

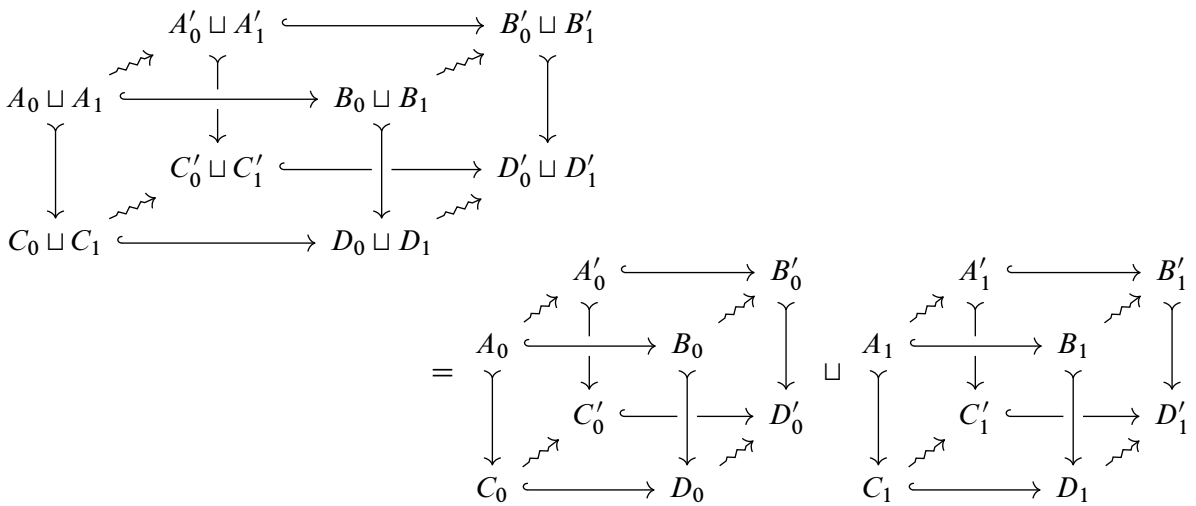
is a weak homotopy equivalence for all positive integers n. In this case the set $\pi_0(A(1_+))$ inherits the structure of an abelian monoid with addition coming from the zigzag

$$A(1_+)\times A(1_+) \xleftarrow[\simeq]{A(e_1)\times A(e_2)} A(2_+) \xrightarrow{A(\nabla)} A(1_+),$$

where the first map is the Segal map and the second map is induced by $\nabla : 2_+ \to 1_+$ with $\nabla(1) = \nabla(2) = 1$. A is called *very special* if it is special and the monoid $\pi_0(A(1_+))$ is a group.

Segal showed how to associate to a very special Γ-space A an Ω-spectrum whose infinite loop space is $A(1_+)$; see [2; 12].

We now construct a Γ-space A such that $A(1_+)$ is the geometric realization of $X_\bullet$. Let us set up some terminology. We say that an element α in X_k is *decomposable* if there exist nonzero elements $\beta, \gamma \in X_k$ such that $\alpha = \beta \sqcup \gamma$, where by a zero element we mean the degenerate k-simplex induced from the zero simplex given by the empty manifold $\varnothing \in X_0$. Recall that any element in X_k is given by a $k \times k \times k$ cube diagram over Δ^k — by the disjoint union of two such cubes we mean the cube given by the vertex-, edge- and facewise disjoint union (if it exists, ie if the corresponding embeddings into $\mathbb{R}^\infty$ are disjoint). An example of a decomposable element in X_1 is depicted below (in this diagram edges and faces of the cube should also be marked by the corresponding disjoint unions, but we have omitted these):



Since we require all manifolds to be compact, every vertex of a cube diagram has only finitely many connected components. Hence every element in X_k can be written as a disjoint union of finitely many indecomposable elements. It is not difficult to see that such a decomposition is unique up to permutation, and we will call the indecomposable summands of an element its *indecomposable components*.

We define $\widetilde{A}(n_+)_\bullet$ to be a simplicial set for which the set of k-simplices $\widetilde{A}(n_+)_k$ is given by pairs (α, L), where $\alpha \neq \varnothing \in X_k$ and L is a labeling of indecomposable components of α by nonbasepoint elements of n_+. For $n = 0$ the simplicial set $\widetilde{A}(0_+)_\bullet$ consists of one point, being the empty manifold $\varnothing$. We treat the $\varnothing \in \widetilde{A}(n_+)_0$ (and the corresponding degenerate simplices) differently than others and don't equip it with any labeling. Note that $\widetilde{A}(1_+)_\bullet$ is just $X_\bullet$.

Lemma 5.1 *The simplicial sets $\{\widetilde{A}(n_+)_\bullet\}_{n\geq 0}$ assemble into a simplicial Γ-set.*

Proof Given a morphism $f : m_+ \to n_+$ of based sets, we define the induced map

$$\widetilde{A}(m_+)_k \to \widetilde{A}(n_+)_k, \quad (\alpha, L) \mapsto (\alpha', L')$$

in the following way. We first relabel the indecomposable components of α, ie the components that had label $i \in \{1, \dots, m\}$ will be labeled by $f(i)$ instead. Next we remove those components that got labeled by the basepoint and call the resulting element (α', L'). □

Lemma 5.2 *The simplicial Γ-set $\widetilde{A}$ is levelwise special in the sense that the Γ-set $n_+ \mapsto \widetilde{A}(n_+)_k$ is special for every k.*

Proof We need to show that for every k, the Segal map is a weak homotopy equivalence

$$\widetilde{A}(n_+)_k \xrightarrow{\prod_{i=1}^n \widetilde{A}(e_i)} \prod_n \widetilde{A}(1_+)_k, \quad (\alpha, L) \mapsto (\alpha_1, \dots, \alpha_n),$$

where $\widetilde{A}(e_i)$ sends (α, L) to α_i, the union of indecomposable components of α that are labeled by i. Note that in $\widetilde{A}(1_+)$ there is no choice for the labeling, all components have to be labeled by the element 1. The image of the Segal map is given by the subset

$$B := \left\{ (\beta_1, \dots, \beta_n) \in \prod_n \widetilde{A}(1_+)_k \;\middle|\; \beta_i \text{ disjoint from } \beta_j \text{ for } i \neq j \right\},$$

where two $k \times k \times k$ diagrams β and γ are called disjoint if the cobordisms (viewed as subsets in $[0,1] \times \mathbb{R}_+ \times \mathbb{R}^\infty$) coming from the corresponding edges of the diagrams β and γ are disjoint. More precisely, if we denote by $\beta_{a,b,[c,c+1]}$ the restriction of the diagram β to the edge that connects the vertices (a,b,c) and $(a,b,c+1)$ and encodes an embedded cobordism, then we require the cobordisms $\beta_{a,b,[c,c+1]}$ and $\gamma_{a,b,[c,c+1]}$ to be disjoint for all $a, b \in \{0, \dots, k\}$ and $c \in \{0, \dots, k-1\}$.

We show that B is a weak deformation retract of the product set $\prod_n \widetilde{A}(1_+)_k$. Analogous to the way it was done in [10] for the usual cobordism category, we use the "shift-map" to make the diagrams disjoint. For the sake of completeness we present the proof here, but we recommend to check [10] for a beautiful visualization of the shift-map.

For $r \in \mathbb{R}$ consider the *shift-map*

$$F_r : \mathbb{R}^\infty \to \mathbb{R}^\infty, \quad (x_1, x_2, \dots) \mapsto (r, x_1, x_2, \dots).$$

It is clear that F_r is homotopic to the identity map. For any nonnegative integer k, the map F_r induces a self-map of $X_k = \widetilde{A}(1_+)_k$ by taking the $k \times k \times k$ diagram over Δ^k $\alpha \in X_k$ and postcomposing all of its vertices, edges and faces (all of these are represented by the embeddings of manifolds/cobordisms) with F_r. Namely, given an embedding of a cobordism W into $[0,1] \times \mathbb{R}_+ \times \mathbb{R}^\infty$, we modify it by postcomposing with

$$\mathrm{id}_{[0,1]} \times \mathrm{id}_{\mathbb{R}_+} \times F_r : [0,1] \times \mathbb{R}_+ \times \mathbb{R}^\infty \to [0,1] \times \mathbb{R}_+ \times \mathbb{R}^\infty.$$

Analogously the postcomposition with F_r is defined for the embedded manifolds, SK-embeddings of manifolds and SK-embeddings of cobordisms. These actions agree and assemble into a new diagram $F_r \circ \alpha \in X_k$. The induced map $X_k \to X_k$ is again homotopic to the identity.

Choose n pairwise different real numbers $r_1, \dots, r_n$ and consider the map

$$\prod_n \widetilde{A}(1_+)_k \to \prod_n \widetilde{A}(1_+)_k$$

given by applying F_{r_i} to the i^{th} factor. This map is the desired deformation retraction. □

Corollary 5.3 *The Γ-space $A\colon \Gamma^{\mathrm{op}} \to S_*$ defined by realizing $\widetilde{A}$ at every n, ie*

$$A(n_+) := \|\widetilde{A}(n_+)_\bullet\|,$$

is special.

Proof A simplicial map that is a degreewise weak homotopy equivalence induces a weak homotopy equivalence on fat realizations, and hence

$$\|\widetilde{A}(n_+)_\bullet\| \xrightarrow[\simeq]{\prod_{i=1}^n \widetilde{A}(e_i)} \prod_n \|\widetilde{A}(1_+)_\bullet\|. \qquad \square$$

Finally, the space $\|\widetilde{A}(1_+)_\bullet\|$ is connected and therefore the Γ-space A is very special. Hence we conclude that $\|X_\bullet\|$ is an infinite loop space as was claimed.

6 Maps from $B\overline{\mathrm{Cob}}^{\partial_{\mathrm{triv}}}_{n+1}$ and $K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$

In this section we consider maps into the constructed spectrum $K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$.

The simplicial set $X^*_{0,0,\bullet}$ is the nerve of a unitalization (in the sense of [4]) of a parametrized version (see [11]) of the cobordism category with trivial boundary $\mathrm{Cob}^{\partial_{\mathrm{triv}}}_{n+1}$ described in Section 2.4. Note that the parametrized cobordism category is defined as a simplicial object in nonunital categories in [11], and we here use a unital version. We write $(N_\bullet \mathrm{Cob}^{\partial_{\mathrm{triv}}}_{n+1})_*$ for the bisimplicial set that is the parametrized nerve of the unital trivial boundary cobordism category.[1]

Let i be the inclusion map

$$i\colon (N_\bullet \mathrm{Cob}^{\partial_{\mathrm{triv}}}_{n+1})_* \to X^*_{\bullet,\bullet,\bullet}.$$

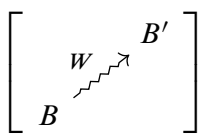
Note that on π_0 of realizations this map is trivial, since $|X^*_{\bullet,\bullet,\bullet}|$ is connected. Denote by $B\mathcal{C}_{n+1}$ the realization $|(N_\bullet \mathrm{Cob}^{\partial_{\mathrm{triv}}}_{n+1})_*|$.

Theorem 6.1 *For $\Omega|i|\colon \Omega B\mathcal{C}_{n+1} \to K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$, the map induced by the inclusion,*

$$\pi_1|i|\colon \pi_1 B\mathcal{C}_{n+1} \to \pi_1|X^*_{\bullet,\bullet,\bullet}| \cong \overline{\mathrm{SK}}^{\partial}_n,$$

is the 0 map.

Proof $B\mathcal{C}_{n+1}$ is the realization of a bisimplicial set, and hence $\pi_1 B\mathcal{C}_{n+1}$ is generated by 1-simplices in the diagonal modulo relations imposed by 2-simplices in the diagonal. Under the map i, any 1-simplex in $(N_\bullet \mathrm{Cob}^{\partial_{\mathrm{triv}}}_{n+1})_*$ is mapped to a cube of the shape



in $X^*_{\bullet,\bullet,\bullet}$, which by the proof of Theorem 4.3 is zero in $|X^*_{\bullet,\bullet,\bullet}|$. Hence $\pi_1|i|$ is the zero map. $\square$

[1] Proposition 3.8 in [4] shows that unitalization induces weak homotopy equivalences on classifying spaces for weakly unital topological categories, although this is a slightly different setting.

Another (tri)simplicial subset of $X^*_{\bullet,\bullet,\bullet}$ that is of interest to us is $X^*_{\bullet,\bullet,0}$, defined in Section 3.4. A spectrum associated to the infinite loop space $\Omega|X^*_{\bullet,\bullet,0}|$ is the topological version of the cut and paste spectrum $K^{\square}(\mathrm{Mfd}^{\partial}_n)$ defined in [7], and will be denoted by $K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$. Here the bar denotes that we are taking topology on manifolds into account by means of the extra simplicial direction.

Recall that $K^{\square}_0(\mathrm{Mfd}^{\partial}_n) \cong \mathrm{SK}^{\partial}_n$. We show that the topological cut and paste spectrum also has SK^{∂}_n as its zeroth homotopy group.

Theorem 6.2 *There is an isomorphism*

$$K^{\square}_0(\overline{\mathrm{Mfd}}^{\partial}_n) \cong \mathrm{SK}^{\partial}_n.$$

Proof We want to compute $\pi_1|X^*_{\bullet,\bullet,0}|$. Write

$$Z = |\mathrm{diag}(X^*_{\bullet,\bullet,0})|.$$

The 2-simplices of the form given by (3) are all in Z; hence we can perform the contraction of vertices as before. Now (10)–(14) in the proof of Theorem 4.3 show that the $\pi_1 Z$ satisfies the SK relations. To conclude that these are all the relations in $\pi_1 Z$, one can show that the relations coming from a generic 2-simplex given by a 2×2 square over Δ^2 can be expressed in terms of the SK relations. □

We denote the map of spectra induced from the inclusion $X^*_{\bullet,\bullet,0} \to X^*_{\bullet,\bullet,\bullet}$ by

$$j : K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n) \to K^{\square\!\!\square}(\overline{\mathrm{Mfd}}^{\partial}_n).$$

Theorem 6.3 *Under the identifications $K^{\square}_0(\overline{\mathrm{Mfd}}^{\partial}_n) \cong \mathrm{SK}^{\partial}_n$ and $K^{\square\!\!\square}_0(\overline{\mathrm{Mfd}}^{\partial}_n) \cong \overline{\mathrm{SK}}^{\partial}_n$, j_0 induces the quotient map*

$$\mathrm{SK}^{\partial}_n \to \overline{\mathrm{SK}}^{\partial}_n.$$

Proof This is evident from comparing the proofs of Theorems 4.3 and 6.2. □

Remark 6.4 Analogously to how it was done in Section 5, one can show that the spaces $K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$ and $\Omega B\mathcal{C}_{n+1}$ are infinite loop spaces. As before, the Γ-space structure comes from the operation of taking the disjoint union of manifolds. One can also check that the obvious map between the corresponding Γ-spaces for $\Omega B\mathcal{C}_{n+1}$, $K^{\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$ and $K^{\square\!\!\square}(\overline{\mathrm{Mfd}}^{\partial}_n)$ restricts to the maps i and j on the first spaces. Therefore $\Omega|i|$ and j are maps of spectra.

References

[1] **M F Atiyah**, **I M Singer**, *The index of elliptic operators, III*, Ann. of Math. 87 (1968) 546–604 MR

[2] **A K Bousfield**, **E M Friedlander**, *Homotopy theory of Γ-spaces, spectra, and bisimplicial sets*, from "Geometric applications of homotopy theory" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 658, Springer (1978) 80–130 MR

[3] **J Campbell**, **J Kuijper**, **M Merling**, **I Zakharevich**, *Algebraic K-theory for squares categories*, preprint (2023) arXiv 2310.02852

[4] **J Ebert**, **O Randal-Williams**, *Semisimplicial spaces*, Algebr. Geom. Topol. 19 (2019) 2099–2150 MR

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

[6] **J Genauer**, *Cobordism categories of manifolds with corners*, Trans. Amer. Math. Soc. 364 (2012) 519–550 MR

[7] **R S Hoekzema**, **M Merling**, **L Murray**, **C Rovi**, **J Semikina**, *Cut and paste invariants of manifolds via algebraic K-theory*, Topology Appl. 316 (2022) art. id. 108105, MR

[8] **U Karras**, **M Kreck**, **W D Neumann**, **E Ossa**, *Cutting and pasting of manifolds*; SK-*groups*, Math. Lecture Series 1, Publish or Perish, Boston (1973) MR

[9] **R C Kirby**, *The topology of 4-manifolds*, Lecture Notes in Math. 1374, Springer (1989) MR

[10] **H K Nguyen**, *On the infinite loop space structure of the cobordism category*, Algebr. Geom. Topol. 17 (2017) 1021–1040 MR

[11] **G Raptis**, **W Steimle**, *Parametrized cobordism categories and the Dwyer–Weiss–Williams index theorem*, J. Topol. 10 (2017) 700–719 MR

[12] **G Segal**, *Categories and cohomology theories*, Topology 13 (1974) 293–312 MR

[13] **W Steimle**, *An additivity theorem for cobordism categories*, Algebr. Geom. Topol. 21 (2021) 601–646 MR

[14] **C T C Wall**, *Non-additivity of the signature*, Invent. Math. 7 (1969) 269–274 MR

[15] **H E Winkelnkemper**, *On equators of manifolds and the action of θ^n*, PhD thesis, Princeton University (1971) Available at `https://www.proquest.com/docview/302535527`

Department of Mathematics, Vrije Universiteit Amsterdam
Amsterdam, Netherlands

Deptartment of Mathematics and Statistics, Loyola University Chicago
Chicago, IL, United States

Laboratoire Paul Painlevé, Université de Lille
Villeneuve d'Ascq, France

`r.s.hoekzema@vu.nl`, `crovi@luc.edu`, `iuliia.semikina@univ-lille.fr`

Received: 3 October 2022 Revised: 12 June 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3315–3340
DOI: 10.2140/agt.2025.25.3315
Published: 1 October 2025

The Curtis–Wellington spectral sequence through cohomology

DANA HUNTER

We study stable homotopy groups through unstable methods applied to their representing infinite loop space Q_0S^0, as pioneered by Curtis and Wellington. Using cohomology instead of homology, we find a width filtration whose subquotients are simple quotients of Dickson algebras, which thus gives a new filtration of stable homotopy groups. We make initial calculations and determine towers in the resulting width spectral sequence. We also make calculations related to the image of J, and prove that the J homomorphism induces a splitting of the indecomposables of the cohomology of Q_0S^0.

55Q40; 55P47, 55Q50

1. Introduction 3315
2. Review 3319
3. Width spectral sequence 3322
4. h_0 towers 3325
5. First analysis of J map 3331
Appendix 3337
References 3339

1 Introduction

Our project, first taken up by Curtis [8] and Wellington [28], is to study stable homotopy groups at the prime two through the unstable Adams spectral sequence for their representing space. We call this the Curtis–Wellington spectral sequence (CWSS), which to our knowledge has not been studied above the zero line for almost forty years.

1.1 Main results

Recall that Q_0S^0 is the zeroth component of the infinite loop space which represents stable homotopy. Barratt and Priddy [3] and Quillen showed that its cohomology is isomorphic to that of the infinite

© 2025 The Author, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

symmetric group. Nakaoka [24] computed the mod-two cohomology of the infinite symmetric group as polynomial. We call its algebra indecomposables, which are a module over the Steenrod algebra, the Nakaoka module $\mathfrak{N}$.

Recall as well that the Dickson algebras are rings of invariants $D_n = \mathbb{F}_2[x_1, \dots, x_n]^{\mathrm{Gl}_n(\mathbb{F}_2)}$, calculated by Dickson as polynomial on generators in degree $2^n - 2^l$. The Steenrod algebra action on the ambient polynomial algebras restricts to the Dickson algebras, which provide a rich and still not fully understood collection of (unstable) modules over the Steenrod algebra. Let D_n^o be the quotient of D_n by all perfect squares.

Theorem 1.1 *There is a width filtration* (see Definition 3.3 below) *of the Nakaoka module $\mathfrak{N}$ whose subquotients are isomorphic to D_n^o as unstable modules over the Steenrod algebra.*

This width filtration is related to composition length in the Dyer–Lashof algebra. Applying a reduction of Bousfield (see Theorem 3.1 below) we have the following.

Corollary 1.2 *There is a width spectral sequence with*

$$E_1^{s,t;n} = \mathrm{Ext}_{\mathcal{U}}^{s,t}(\Sigma^{-1} D_n^o, \mathbb{F}_2)$$

and $d_r : \mathrm{Ext}^{s,t}(\Sigma^{-1} D_n^o) \to \mathrm{Ext}^{s+1,t}(\Sigma^{-1} D_{n+r}^o)$ which converges to the E_2 page of the Curtis–Wellington spectral sequence.

Accordingly, there is a filtration on stable homotopy. In Section 5 we prove that the first submodule in this filtration detects the image of J. Our calculations, made at the prime 2, show that this must be different than the chromatic filtration since the first filtration corresponds to only the image of J and no additional elements.

Our calculations rely on a presentation of the cohomology of $Q_0 S^0$ due to Giusti, Salvatore and Sinha. Through that presentation we can already manage by-hand calculations of the Curtis–Wellington spectral sequence more readily than by previous techniques. Moreover, filtration with Dickson algebra subquotients is particularly amenable to computer calculation, for which we thank Hood Chatham; see Figure 8.

It is elementary to eliminate the possibility of many differentials in the width spectral sequence, so these computer calculations provide a first picture of the E_2 page of the Curtis–Wellington spectral sequence as well. While, disappointingly, the E_2 page of the CWSS is much larger than the classical Adams spectral sequence, we have explored two accessible phenomena.

The first accessible phenomenon is the existence of towers of elements connected by multiplication by h_0. In Section 4 we make calculations to determine the locations of these towers in the resulting width spectral sequence.

Theorem 1.3 *There are infinite towers T_{a_1} in $\mathrm{Ext}^{s,t}(\Sigma^{-1} D_1^o, \mathbb{F}_2)$ only in degrees $t - s = a_1 - 1$ for $a_1 = 4k - 1$, where k is a positive integer.*

Theorem 1.4 *Let $n \geq 2$ be an integer. There are towers $T_{a_1,\dots,a_n}$ in $\mathrm{Ext}^{s,t}(\Sigma^{-1} D_n^o, \mathbb{F}_2)$ in degree $t-s = \left(\sum_{i=1}^n (2^n - 2^{n-i})a_i\right) - 1$ satisfying one of the following two conditions*:

(1) *$a_n = 0$, a_{n-1} is even, at least one other a_i is odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4l$.*

(2) *a_i for $1 \leq i \leq n-2$ are all even, a_{n-1} and a_n are odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4k-1$.*

These results agree with the locations of towers identified by Wellington in the E_2 page of the CWSS, and thus imply that there are no differentials between the towers internal to the width spectral sequence.

The lowest filtration is also accessible, and related to the image of J. In Section 5 we make calculations which show preliminarily that the image of J is compatible with the width filtration. In studying the image of J, we first noticed that the unstable Adams Ext chart for $H^*(BO, \mathbb{F}_2)$ is a shifted version of the unstable Adams Ext chart for the first quotient of the width filtration.

Proposition 1.5 $$\mathrm{Ext}_{\mathcal{U}}^{s,t}(\Sigma^{-1} D_1^o) \cong \mathrm{Ext}_{\mathcal{U}}^{s,t+1}(\Sigma^{-1} \operatorname{Ind} H^*(BO)).$$

While it seems this would be classical, we haven't found any treatment of this in the literature. This motivated us to study the image of the J map on cohomology, and the following.

Theorem 1.6 *The map induced by the image of J on cohomology induces a splitting of the Nakaoka module $\mathfrak{N}$.*

Corollary 1.7 *The algebraic map induced by the image of J on* Ext *induces a splitting of* $\mathrm{Ext}(\mathfrak{N})$.

We also speculate that the map on the E_2 page of the unstable Adams spectral sequence induced by the J map on cohomology agrees with the map on Ext induced by the map on cohomology. If this were the case, then the splitting of Theorem 1.6 would give rise to the standard splitting of homotopy by the image of J.

1.2 Background

Recall that the stable homotopy groups of spheres are isomorphic to the unstable homotopy groups of $Q_0 S^0$, the degree zero component of $\varinjlim \Omega^d S^d$. The unstable Adams spectral sequence for $Q_0 S^0$, which we choose to name the Curtis–Wellington spectral sequence after the only two people to study it globally, was introduced by Curtis [8]. He outlined some first calculations, noticing that Adams filtration lowered and that both the Hopf and Kervaire classes were in filtration zero, leading to the well-known and still open conjecture that these are the only classes to survive on the zero line. But Curtis made some fundamental errors, which Wellington corrected before going on to establish more global properties, including the classification of Bockstein towers (which in particular preclude any upper vanishing lines).

In order to make these calculations, Curtis and Wellington applied a deep connection between stable homotopy and symmetric groups. At the level of homology, this was noticed independently and simultaneously by Barratt and Priddy [3] and Quillen. Briefly, one models the classifying space for the n^{th} symmetric

group as $\varinjlim_d \mathrm{Disks}_n(\mathbb{R}^d)$, a colimit over d of the space of n disks in $\mathbb{R}^d$. Then, given a set of n disks in $\mathbb{R}^d$, associate to it a collapse map from $S^d = \mathbb{R}^d \cup \infty$ to itself which sends the complement of the disks to the basepoint and each interior of a disk homeomorphically onto the $S^d \setminus \infty$. These maps from the space of disks to $\Omega^d S^d$ can be assembled to a map from the colimit. The Barratt–Priddy–Quillen theorem tells us that the resulting map from the classifying space for the infinite symmetric groups to $Q_0 S^0$ is an isomorphism in homology.

Wellington and Curtis then applied the known structure of the homology of $Q_0 S^0$. This homology is the symmetric algebra on the Kudo–Araki–Dyer–Lashof algebra; see Cohen, Lada and May [7]. While algebraic topologists are comfortable with "homological coalgebra", in this case one runs into difficulties because calculations of the homology coproduct as well as the Steenrod coaction (Nishida relations) require regular applications of Adem relations. Wellington had to filter carefully to make things at all tractable. Through nonexplicit methods, Nakaoka [23; 24] had previously shown that the cohomology of the infinite symmetric group and thus $Q_0 S^0$ is polynomial, generated in combinatorially interesting degrees. Finer control of that calculation, and in particular incorporation of the Steenrod algebra action, motivated many authors to study the cohomology of symmetric groups in more detail in the Eighties and Nineties; see for example Adem, Maginnis and Milgram [1], Adem and Milgram [2] and Feshbach [11]. Relatively recently, Giusti, Salvatore and Sinha [12] found a new Hopf ring presentation for the cohomology of symmetric groups, as algebras over the Steenrod algebra, yielding a "skyline diagram" presentation for the cohomology of $\mathcal{S}_\infty$ in the limit. We take their work as a starting point, and ultimately see that it makes the Curtis–Wellington spectral sequence much more accessible than the previous approach through homology.

Plan of the paper

In Section 2 we review work of Giusti, Salvatore and Sinha on the cohomology of symmetric groups. The unstable Adams spectral sequence is typically intractable, with a nonabelian Quillen homology defining its E_2 page. But in Section 3 we apply a standard result of Bousfield [4] in the special case that a cohomology ring is free, as is the case here, equating the E_2 page with Ext in the category of unstable modules over the Steenrod algebra of the desuspended indecomposables. Using the skyline diagram presentation of the cohomology of the infinite symmetric group, these indecomposables are manageable.

Indeed, we show that a filtration by skyline diagram width (which corresponds to composition length in the Kudo–Araki–Dyer–Lashof algebra) yields subquotients which are given by the Dickson algebras, modulo perfect squares. The resulting width spectral sequence is relatively tractable, allowing us, for example, to reveal an error, likely of transcription, in Wellington's Ext charts (at the 11 and 12 stems). We then share computer calculations which imply many more differentials than in the classical Adams spectral sequence, but regular phenomena as well.

In particular, there are Bockstein (h_0) towers, which we classify in the width spectral sequence in Section 4. These occur in the same dimensions as Wellington identified, with considerably more effort, in the E_2

of the CWSS. Thus, there are no differentials in the width spectral sequence with h_0 inverted, and we conjecture no differentials in the width spectral sequence in general, a purely algebraic question. It would be interesting to understand differentials between these h_0 towers and the special roles the resulting classes in homotopy might play.

Based on a remarkable identification of the unstable Adams E_2 for BO and some preliminary calculations, in Section 5 we initiate the study of the J homomorphism. We calculate the induced map on cohomology of the indecomposables and prove that it induces a splitting of the Nakaoka module.

Acknowledgements

I would like to thank my advisor Dev Sinha for all of his help and guidance, Hal Sadofsky for many helpful conversations, Dan Isaksen for confirming interest in the aspects of the Curtis–Wellington spectral sequence which make up the bulk of this work, Haynes Miller for providing references, and Peter May for clarifying some of the history. I would also like to acknowledge the partial support of NSF grant DMS-2039316.

2 Review

Recall that $H^*(Q_0S^0, \mathbb{F}_2) \cong H^*(BS_\infty, \mathbb{F}_2)$. Recent calculations of cohomology of finite symmetric groups by taking all symmetric groups together and considering both cup product and a transfer or induction product gives a concise presentation.

Theorem 2.1 (GSS) *As a Hopf ring, $\bigoplus_n H^*(B\mathcal{S}_n; \mathbb{F}_2)$ is generated by classes $\gamma_{l[n]} \in H^{n(2^l-1)}(B\mathcal{S}_{n2^l})$ along with unit classes on each component. The coproduct of $\gamma_{l[n]}$ is given by*

$$\Delta\gamma_{l[n]} = \sum_{i+j=n} \gamma_{l[i]} \otimes \gamma_{l[j]}.$$

Relations between transfer products of these generators are given by

$$\gamma_{l[n]} \odot \gamma_{l[m]} = \binom{n+m}{n} \gamma_{l[n+m]}.$$

Relations between cup products of generators are that cup products of generators on different components are zero.

These can be presented graphically, as "skyline diagrams". The generator $\gamma_{l[n]}$ is represented by a rectangular block of width $n \cdot 2^{l-1}$ and total area $n \cdot (2^l - 1)$, so that the area of the block corresponds to its degree in cohomology and we get a unique block type for each generator. Cup product is indicated by vertical stacking to make columns, whose placement next to each other denotes transfer product. We also draw in vertical dashed lines separating the block into n equal sections, for purposes of the coproduct. An example of such a diagram can be seen in Figure 1.

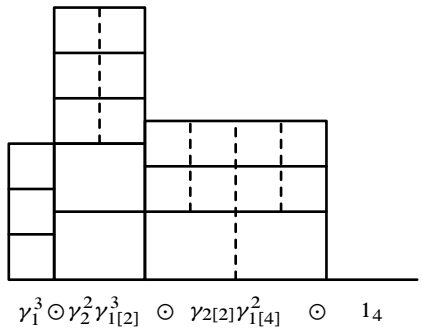


Figure 1: Skyline diagram for $\gamma_1^3 \odot \gamma_2 \gamma_{1[2]}^3 \odot \gamma_{2[2]} \gamma_{1[4]}^2 \odot 1_4$.

The cohomology of the infinite symmetric group is the inverse limit

$$H^*(BS_\infty) = \varprojlim_n H^*(BS_n).$$

The maps $H^*(BS_n) \to H^*(BS_m)$ for $n > m$ take a diagram in $H^*(BS_n)$ which has a "tail" of width greater than or equal to $\frac{1}{2}(n-m)$ — that is, a $\odot$-product factor of 1_k with $k > \frac{1}{2}(n-m)$ — to a diagram in $H^*(BS_m)$ obtained by shortening its tail to make it the appropriate width to be an element of $H^*(BS_m)$. If a class is not such a transfer product with a sufficiently large unit class, it maps to zero. In Figure 1 the diagram has a tail of width 2 and is an element of $H^*(BS_{18})$. Shortening the tail once, it becomes $\gamma_1^3 \odot \gamma_2 \gamma_{1[2]}^3 \odot \gamma_{2[2]} \gamma_{1[4]}^2 \odot 1_2 \in H^*(BS_{16})$. Shortening it again produces $\gamma_1^3 \odot \gamma_2 \gamma_{1[2]}^3 \odot \gamma_{2[2]} \gamma_{1[4]}^2 \in H^*(BS_{14})$ after which point the tail can no longer be shortened and its image in $H^*(BS_{2k})$ is 0 for $k < 7$.

With restriction maps taking this form, the cohomology of $B\mathcal{S}_\infty$ could be viewed through such diagrams with "infinitely long tails", or in monomial form as "$\odot 1_\infty$". As they confer no additional information, we prefer to omit the tails altogether, only using them implicitly when we calculate using the Hopf ring product structures. Using this presentation, we next recall another basic result of Giusti, Salvatore and Sinha, refining a classical result of Nakaoka.

Theorem 2.2 *The cohomology of BS_∞ is a polynomial algebra. Minimal generators of $H^*(BS_\infty)$ as an algebra under cup product are represented graphically by single columns with at least one block type appearing an odd number of times. That is, generators as an algebra are products $\prod_i \gamma_{l[n_i]}^{a_i}$ of Hopf ring generators of the same width, ie $n_i 2^{l_i - 1}$ is equal for all i, with at least one a_i odd.*

These minimal generators form a basis for the indecomposables of the cohomology of BS_∞, which we call the Nakaoka module $\mathfrak{N}$.

The idea of the proof is to consider the product of single-column diagrams which contribute to a skyline diagram. This product results in a sum of diagrams, the widest of which is the original skyline diagram. This shows such products are algebraically independent, and a simple filtration argument shows the

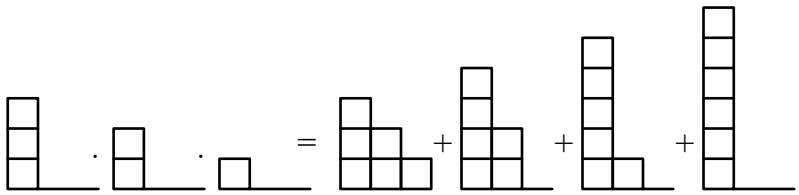


Figure 2: Skyline diagram illustrating $\gamma_1^3 \odot 1_\infty \cdot \gamma_1^2 \odot 1_\infty \cdot \gamma_1 \odot 1_\infty = \gamma_1^3 \odot \gamma_1^2 \odot \gamma_1 \odot 1_\infty +$ lower-width terms.

polynomial ring they generate exhausts the cohomology. For example, consider $\gamma_1^3 \odot \gamma_1^2 \odot \gamma_1 \odot 1_\infty \in H^*(BS_\infty)$, whose skyline diagram has three width-one columns of heights three, two and one. It is the highest-width term in the product of its three columns $\gamma_1^3 \odot 1_\infty$, $\gamma_1^2 \odot 1_\infty$ and $\gamma_1 \odot 1_\infty$, as illustrated in Figure 2.

In the same paper, Giusti, Salvatore and Sinha describe the Steenrod algebra action on the cohomology of symmetric groups in terms of the basis elements $\gamma_{l[2^k]}$. Because there are Cartan formulas for both cup and transfer product, this determines the Steenrod structure on the whole.

Definition 2.3 (i) A gathered block is a monomial of the form $\prod_i \gamma_{l_i[n_i]}^{d_i}$ where the product is the cup product and all the widths $2^{l_i} n_i$ are equal.

(ii) The algebraic degree of a gathered block is the total number of Hopf ring generators $(\gamma_{l_i[n_i]})$ cup-multiplied to make the gathered block.

(iii) The height of one of these skyline diagrams is the largest of the algebraic degrees of its constituent gathered blocks.

(iv) The effective scale of a gathered block, composed of $\gamma_{l[n]}$ cup-multiplied together, is the largest such l that occurs in the block. The effective scale of a gathered monomial is the minimum of the effective scales of its gathered blocks.

(v) A monomial is full-width as long as it is not a nontrivial transfer product of some monomial with a unit class 1_k for some positive integer k.

As an example of the definitions above, consider the monomial $\gamma_1^3 \odot \gamma_2 \gamma_{1[2]}^3 \odot \gamma_{2[2]} \gamma_{1[4]}^2 \odot 1_4 \in H^*(BS_{18})$, whose skyline diagram representation is shown in Figure 1. This monomial is made up of three gathered blocks: γ_1^3 of algebraic degree 3 and effective scale 1, $\gamma_2 \gamma_{1[2]}^3$ of algebraic degree 5 and effective scale 2, and $\gamma_{2[2]} \gamma_{1[4]}^2$ of algebraic degree 3 and effective scale 2. Since 5 is the largest degree and 1 is the smallest effective scale, this diagram has height 5 and effective scale 1. This monomial is not full-width since it is the transfer product of the full-width monomial $\gamma_1^3 \odot \gamma_2 \gamma_{1[2]}^3 \odot \gamma_{2[2]} \gamma_{1[4]}^2$ and the unit class 1_4.

Theorem 2.4 [12, Theorem 8.3] *The Steenrod square* $\mathrm{Sq}^i \gamma_{l[2^k]}$ *is the sum of all full-width monomials of total degree* $2^k(2^l - 1) + i$, *height* 1 *or* 2, *and effective scale at least* l, *with height* 2 *only allowed if the effective scale is* l.

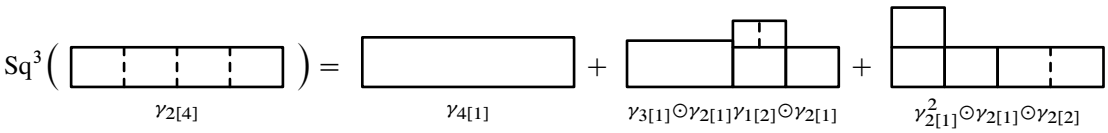


Figure 3: Skyline diagrams for the three summands of $\mathrm{Sq}^3(\gamma_{2[4]})$.

For example, Figure 3 illustrates the three summands of $\mathrm{Sq}^3(\gamma_{2[4]})$:

$$\mathrm{Sq}^3(\gamma_{2[4]}) = \gamma_{4[1]} + \gamma_{3[1]} \odot \gamma_{2[1]}\gamma_{1[2]} \odot \gamma_{2[1]} + \gamma_{2[1]}^2 \odot \gamma_{2[1]} \odot \gamma_{2[2]}.$$

It is straightforward to use Cartan formulas to calculate Steenrod action on $\mathfrak{N}$, the indecomposables. This gives a refinement of Nakaoka's work, which only determined this module additively, and is a much more accessible presentation than of the homology primitives [7; 19]. Indeed, much of Wellington's work on the CWSS is devoted to calculations with these primitives. These calculations are generally simplified or become immaterial through this cohomology approach.

3 Width spectral sequence

In this section, we equate the Curtis–Wellington E_2 with an explicit Ext group in the category of unstable modules over the Steenrod algebra, at which point there is an immediate filtration to develop. Recall that work of [6] and others proves that, for simply connected X with $\pi_*(X)$ of finite type, there is an unstable analog to the Adams spectral sequence with $E_2 \cong \mathrm{Ext}^{s,t}_{\mathcal{U}\mathcal{A}}(H^*(X), \mathbb{F}_2)$ converging to $\pi_*(X)$.

There have been relatively few computations made of the unstable Adams spectral sequence, with some explicit calculations of Curtis and Mahowald [10; 9] and Miller's proof of the Sullivan conjecture [22] being spectacular exceptions. A main roadblock is that the Ext groups which occur, which we call $\mathrm{Ext}_{\mathcal{U}\mathcal{A}}$ for the category of unstable algebras over the Steenrod algebra, are not Ext groups in the usual sense of derived homomorphisms in an abelian category. While Goerss [13] established that they are a "nonabelian" derived Hom, in the sense of Quillen, in the category of simplicial algebras over $\mathcal{A}$, this has not to our knowledge been used in any way for calculations.

To make calculations, one can hope for equivalent Ext calculations in abelian categories. The category of unstable modules over the Steenrod algebra, $\mathcal{U}$, is abelian and there is a free unstable algebra functor $\mathcal{U} \to \mathcal{U}\mathcal{A}$. However, the cohomology of a space is very rarely in the image of this functor, even if it is free. If the cohomology is free only as an algebra, there is still an alternative form of reduction. Let A be an augmented algebra with $\bar{A}$ its augmentation ideal. We let $\mathrm{Ind}\, A$ denote the algebra indecomposables $\bar{A}/(\bar{A}\cdot\bar{A})$. Note that $\Sigma^{-1}\,\mathrm{Ind}\, A$ is naturally in $\mathcal{U}$. The following was originally stated by Bousfield [4] and follows from the composite functor spectral sequence constructed by Miller [22].

Theorem 3.1 *Let P be an unstable algebra over $\mathcal{A}$ that is free as an algebra on* Ind *P. Then*

$$\mathrm{Ext}^{s,t}_{\mathcal{U}\mathcal{A}}(P,k) \cong \mathrm{Ext}^{s,t-1}_{\mathcal{U}}(\Sigma^{-1}\mathrm{Ind}\, P, k).$$

Recall that we refer to $\mathrm{Ind}\, H^*(Q_0 S^0, \mathbb{F}_2)$ as the Nakaoka module, $\mathfrak{N}$. Thus, Theorem 3.1 applies to give the following reduction to a calculation in the abelian category $\mathcal{U}$.

Corollary 3.2 *The E_2 term of the Curtis–Wellington spectral sequence is isomorphic to* $\mathrm{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1}\mathfrak{N})$.

Our graphical skyline diagram presentation, in particular of the indecomposables as stated in Theorem 2.2, points immediately to a filtration of the Nakaoka module.

Definition 3.3 Let F_n be the submodule of $\mathfrak{N}$ of elements of width less than or equal to 2^{n-1}.

This is a submodule as unstable Steenrod modules by the formula from Theorem 2.4. Moreover, F_n/F_{n-1} will be spanned by single columns of width exactly 2^{n-1} with at least one block type appearing an odd number of times.

Proposition 3.4 *The quotient F_n/F_{n-1} is isomorphic to D_n^o, where D_n^o is the quotient of D_n by all perfect squares (just the squares themselves, not the ideal generated by squares).*

Proof Let V_n be the transitive elementary abelian 2-subgroup of S_{2^n}, $(\mathbb{Z}_2)^n$. Corollary 7.6 of [12] follows an argument of Milgram to show the restriction of $\gamma_{l[2^k]}$ with $l+k=n$ to V_n is the Dickson class $d_{k,l}$. Since F_n is the submodule of elements of width less than or equal to 2^{n-1} and F_{n-1} is the submodule of elements of width less than or equal to 2^{n-2}, F_n/F_{n-1} will be spanned by single columns of width exactly 2^{n-1} with at least one block type appearing an odd number of times. The blocks of width 2^{n-1} are $\gamma_{i[2^{n-i}]}$ for i an integer between 1 and n; thus, when restricted to V_n, we get Dickson classes $d_{i,n-i}$, all the elements of D_n. Since at least one block type must appear an odd number of times, the only Dickson algebra elements we cannot get are those where each $d_{i,n-i}$ has an even power, which leaves us with the quotient of D_n by just the perfect squares. □

This filtration allows us to consider only full-width terms in the image of the Steenrod action, substantially simplifying calculations. Assembling the long exact sequences in cohomology associated to the short exact sequences

$$0 \to F_{n-1} \to F_n \to D_n^o \to 0$$

from the filtration described above produces a trigraded spectral sequence converging to the E_2 page of the Curtis–Wellington spectral sequence for stable homotopy.

Theorem 3.5 *The spectral sequence associated to the width filtration has*

$$E_1^{s,t;n} = \mathrm{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1} D_n^o, \mathbb{F}_2)$$

and $d_r : \mathrm{Ext}^{s,t}(\Sigma^{-1} D_n^o) \to \mathrm{Ext}^{s+1,t}(\Sigma^{-1} D_{n+r}^o)$. It converges to $\mathrm{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1}\mathfrak{N})$, and thus the E_2 of the Curtis–Wellington spectral sequence.

Using the well-known Steenrod structure on Dickson algebras, as presented for example in [15], we have been able to make hand calculations out to the 17 stem. Hood Chatham kindly produced an Ext chart

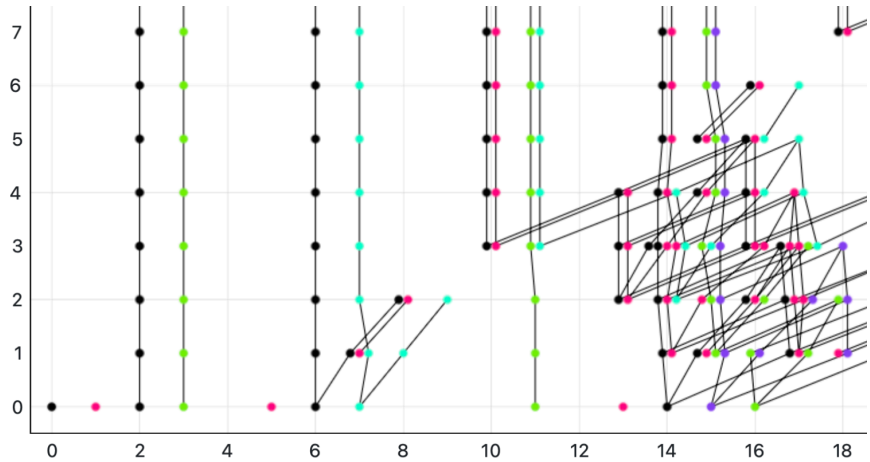


Figure 4: The E_1 page of the width spectral sequence, with width filtration encoded by color: black corresponds to D_1, red to D_2, green to D_3, teal to D_4, and purple to D_5.

illustrating the first page of this spectral sequence, out to the forty-five stem. We share a clip here in Figure 4 and the full chart in the appendix. In this chart, the black connecting lines indicate the partial action of $\mathrm{Ext}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ described in slightly more detail at the start of Section 4. Along the zero line in this width spectral sequence, we see the Steenrod indecomposables of the Dickson algebras, which have been of interest, for example in work of Hung and Peterson [16], and are far from understood. This zero line receives the Hurewicz map for $Q_0 S^0$, as studied by Lannes and Zarati [18].

Wellington made similar computations for $\mathrm{Ext}(H_* Q_0 S^0)$ at the prime 2, including charts out to the 17 stem, which we reproduce with a minor correction in Figures 5 and 6. Comparing Wellington's results to ours, they mainly agree, but our calculations reveal an error in the 11 and 12 stem in the Ext chart. The original chart had a class in bidegree $(12, 4)$ and a d_2 differential to the class in degree $(11, 6)$.

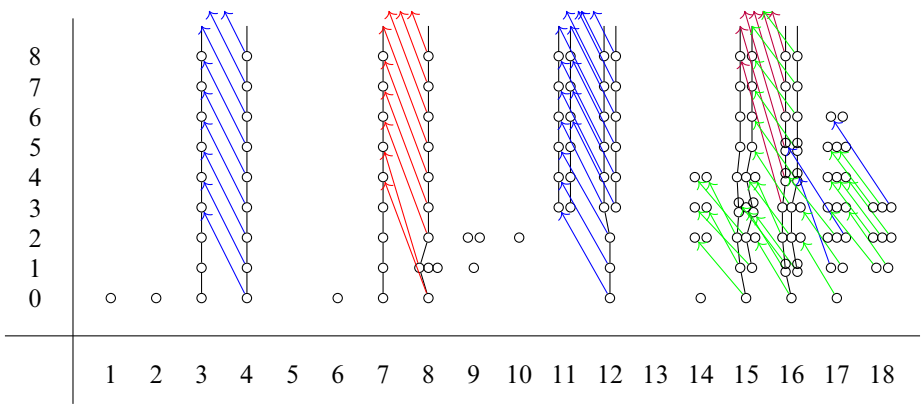


Figure 5: Part of [28, Table 13] (with correction in the 12 stem), depicting the E_2 page and differentials of the CWSS.

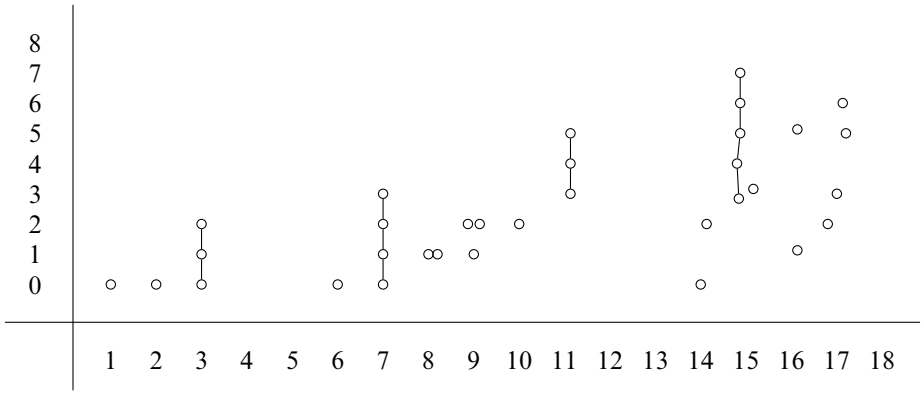


Figure 6: E_∞ page of the Curtis–Wellington spectral sequence up through the 17 stem.

Instead, our hand calculations show that there is a class in degree $(12, 3)$, which appears in the computer calculations in degree $(11, 3)$ after the desuspension. Since we know that the CWSS must converge to stable homotopy, we know that there must be a d_3 differential instead of the d_2 differential. Also note that, with the exception of the class just discussed, which we believe to be just a minor transcription error, every class on the width spectral sequence E_1 is still present on the E_2 page of the CWSS and thus survives the width spectral sequence.

The differentials of the spectral sequence of Theorem 3.5, namely

$$d_r : \mathrm{Ext}^{s,t}(D^o_n) \to \mathrm{Ext}^{s+1,t}(D^o_{n+r}),$$

fix topological degree (t), increase cobar length by 1 (s), and increase filtration by r. In the charts, they will be moving one unit left, one unit up, and in our representation of the third (width) grading by color move between different colors so that, if the source is in D^o_n, the target is in D^o_{n+r}.

4 h_0 towers

One of the immediate differences between the Adams spectral sequence and Curtis–Wellington spectral sequence is the presence of infinite h_0 towers. We classify these in the width spectral sequence.

We begin by defining a partial action of $\mathrm{Ext}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$. In the Yoneda approach to Ext, an element of $\mathrm{Ext}^{s,t}_{\mathscr{A}}(\mathbb{F}_2, \mathbb{F}_2)$ is an equivalence class of extensions of length s from $\mathbb{F}_2$ to $\Sigma^t \mathbb{F}_2$, namely

$$0 \to \Sigma^t \mathbb{F}_2 \to E_1 \to \cdots \to E_s \to \mathbb{F}_2 \to 0,$$

where the E_i are $\mathscr{A}$-modules. An element of $\mathrm{Ext}^{p,q}_{\mathscr{U}}(\Sigma^{-1} D^o_n, \mathbb{F}_2)$ is an extension of length p from $\Sigma^{-1} D^o_n$ to $\Sigma^q \mathbb{F}_2$,

$$0 \to \Sigma^q \mathbb{F}_2 \to F_1 \to \cdots \to F_p \to \Sigma^{-1} D^o_n \to 0,$$

where the F_i are unstable $\mathscr{A}$-modules and the maps $F_i \to F_{i+1}$ are multiplication by elements of $\mathscr{A}$ of excess less than or equal to the degree of the generator of F_{i+1}.

Definition 4.1 Define a partial action of $\mathrm{Ext}_{\mathscr{A}}^{s,t}(\mathbb{F}_2,\mathbb{F}_2)$ on $\mathrm{Ext}_{\mathscr{U}}^{p,q}(\Sigma^{-1}D_n^o,\mathbb{F}_2)$, defined when $t-s+1\le q$, by suspending a representative of the equivalence class of the stable extension q times and concatenating it on the left with a representative of the equivalence class of the unstable extension to give an extension which defines a representative of an element of $\mathrm{Ext}_{\mathscr{U}}^{s+p,t+q}(\Sigma^{-1}D_n^o,\mathbb{F}_2)$:

$$0\to\Sigma^{t+q}\mathbb{F}_2\to\Sigma^q E_1\to\cdots\to\Sigma^q E_s\to\Sigma^q\mathbb{F}_2\to F_1\to\cdots\to F_p\to\Sigma^{-1}D_n^o\to 0.$$

We next recall the Λ-algebra [6; 5], which gives an explicit though computationally involved way to compute some Ext groups over the Steenrod algebra. The Λ-algebra is the graded associative differential algebra with unit over $\mathbb{F}_2$ with

(1) a generator λ_i of degree i for each $i\ge 0$;

(2) for each $i,k\ge 0$, a relation
$$\lambda_i\lambda_{2i+1+k}=\sum_{j\ge 0}\binom{k-1-j}{j}\lambda_{i+k-j}\lambda_{2i+1+j};$$

(3) a differential ∂ given by
$$\partial(\lambda_i)=\sum_{j\ge 1}\binom{i-j}{j}\lambda_{i-j}\lambda_{j-1}.$$

Note that $\Lambda=\bigoplus_{s\ge 0}\Lambda^s$, where Λ^s is generated by monomials λ_I of length s.

To define complexes using the Λ-algebra, we need our action to be on the right. Let $\mathscr{U}_{\mathscr{R}}$ denote the category of unstable right $\mathscr{A}$-modules and continue to denote by $\mathscr{U}$ the category of unstable left $\mathscr{A}$-modules. For M of finite type, $\mathrm{Ext}_{\mathscr{U}_{\mathscr{R}}}^{s,t}(\mathbb{F}_2,M)\cong\mathrm{Ext}_{\mathscr{U}}^{s,t}(M^*,\mathbb{F}_2)$.

The Λ-algebra gives one method to approach calculation of Ext groups; in particular, from [5],

$$\mathrm{Ext}_{\mathscr{U}_{\mathscr{R}}}^{s,t}(\mathbb{F}_2,M)\cong H^s(V(M))_{t-s},$$

where $V(M)$ is the chain complex

$$M\to M\mathbin{\widehat{\otimes}}\Lambda^1\to M\mathbin{\widehat{\otimes}}\Lambda^2\to\cdots,$$

where $M\mathbin{\widehat{\otimes}}\Lambda$ is the subspace of $M\otimes\Lambda$ generated by all $x\otimes\alpha$ with $x\in M_n$ and $\alpha\in\Lambda(n)$, the subspace of Λ generated by allowable λ_I with $i_1<n$ or I empty.

Bousfield [4] defines the following tower complex as a quotient of the chain complex $V(M)$.

Definition 4.2 Let
$$T^s(M)=\begin{cases} f\otimes(\lambda_0)^s & \text{if } s=0,1,\\ M\otimes(\lambda_0)^s\oplus\sum_{k>0}M_{2k}\otimes\lambda_{2k-1}(\lambda_0)^{s-1} & \text{if } s>1,\end{cases}$$
with
$$\delta(x\otimes\lambda_{2k-1}(\lambda_0)^{s-1})=0$$
and
$$\delta(x\otimes(\lambda_0)^s)=\begin{cases} x\cdot\mathrm{Sq}^1\otimes(\lambda_0)^{s+1}+x\cdot\mathrm{Sq}^{2k}\otimes\lambda_{2k-1}(\lambda_0)^s & \text{if } s>0,\ x\in M_{4k},\\ x\cdot\mathrm{Sq}^1\otimes(\lambda_0)^{s+1} & \text{otherwise.}\end{cases}$$

Remark 2.4 of [4] notes that the towers in $H^*(T(M))$ correspond with those in $H^*(V(M))$ and thus also with the towers in $\mathrm{Ext}^{s,t}_{U_R}(\mathbb{F}_2, M)$. Applying these tower detectors to $\mathrm{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1}D^o_n, \mathbb{F}_2)$, we conclude the following two theorems.

Theorem 4.3 *There are infinite towers T_{a_1} in $\mathrm{Ext}^{s,t}(\Sigma^{-1}D^o_1, \mathbb{F}_2)$ only in degrees $t-s = a_1 - 1$ for $a_1 = 4k-1$, where k is a positive integer.*

Theorem 4.4 *Let $n \geq 2$ be an integer. There are towers $T_{a_1,\dots,a_n}$ in $\mathrm{Ext}^{s,t}(\Sigma^{-1}D^o_n, \mathbb{F}_2)$ in degree $t-s = \left(\sum_{i=1}^n (2^n - 2^{n-i})a_i\right) - 1$ satisfying one of the following two conditions:*

(1) *$a_n = 0$, a_{n-1} is even, at least one other a_i is odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4l$.*

(2) *a_i for $1 \leq i \leq n-2$ are all even, a_{n-1} and a_n are odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4k-1$.*

Wellington [28] also used these tower detecting complexes, but of course with his homology approach. Recall that the homology of $Q_0 S^0$ is free under the product induced by loop sum on classes Q^I where I is admissible. Wellington proves there are towers in dimensions $4k-1$ and $4k$ generated by Q^I either in degree $4k$ with excess 0 and some odd index, or Q^I in degree $4k-1$ with final index odd and all others even. Our calculations agree with Wellington's in that the towers in $\mathrm{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1}D^o_n, \mathbb{F}_2)$ for each n correspond with his generated by Q^I with $l(I) = n$. Indeed, in degree $(4k-1)-1$, we have towers $T_{a_1,\dots,a_n}$ where the a_i for $1 \leq i \leq n-2$ are all even, a_{n-1} and a_n are odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4k-1$. In other words, we have a tower corresponding to each integer solution of

$$(2^{n-1} - 2^{n-2})b_1 + \cdots + (2^{n-1} - 1)b_{n-1} + (2^n - 1)c_n + (2^n - 1) = k$$

where $a_i = 2b_i$ for $1 \leq i \leq n-2$, $a_{n-1} = 2b_{n-1} + 1$ and $a_n = 4c_n + 3$. Each of these corresponds to the tower generated by Q^I with $I = (s_1, s_2, \dots, s_n)$ where $s_n = 2k + 1 + 2c_n$ and the other s_i can be computed inductively from right to left with the formula

$$s_i = \frac{1}{2^{(n-1)-i}}\left(k + 1 + a_n + \sum_{j=i}^{n-1} 2^{(n-1)-j} b_j\right) \quad \text{for } 2 \leq i \leq n-1,$$

and finally

$$s_1 = 4k - 1 - \sum_{j=2}^{n} s_j.$$

In degree $4k-1$, we found towers $T_{a_1,\dots,a_n}$ where $a_n = 0$, a_{n-1} is even, at least one other a_i is odd, and $\sum_{i=1}^n (2^n - 2^{n-i})a_i = 4k$. Thus, we have a tower for each integer solution of

$$(2^{n-2} - 2^{n-3})a_1 + \cdots + (2^{n-2} - 1)a_{n-2} + (2^{n-1} - 1)b_{n-1} = k,$$

where $a_{n-1} = 2b_{n-1}$. Each of these corresponds to the tower generated by Q^I with $I = (s_1, s_2, \dots, s_n)$ where we once again find each term in the index working from right to left. First, $s_n = 2k$ and $s_{n-1} = k + b_{n-1}$, then inductively we compute

$$s_i = \frac{1}{2^{(n-1)-i}}\left(k + b_{n-1} + \sum_{j=i}^{n-2} 2^{(n-2)-j} a_j\right) \quad \text{for } 2 \leq i \leq n-1,$$

and finally

$$s_1 = 4k - \sum_{j=2}^{n} s_j.$$

This agreement between our towers in the width spectral sequence and Wellington's imply there are no differentials in the width filtration spectral sequence with h_0 inverted. The only thing that we can definitively say independent of Wellington's results is that there are no possible width spectral sequence differentials until the 60 stem. Indeed, since the differentials in the width spectral sequence increase width filtration, the earliest possible differential would be from some tower from D_3 in the $4k$ stem to some tower from D_4 in the $4k-1$ stem, the first of which is in the 59 stem. Between the fact that these results indicate there are no differentials between towers and that some differentials can be eliminated by hand calculations in low degrees, we wonder whether there are any differentials in the width spectral sequence at all, a purely algebraic question.

Proof of Theorem 4.3 Let d_1 be the generator of D_1. Then elements of D_1^o are d_1^{2i+1} in degree $2i$ after desuspension. Take $\{x_1, x_3, \dots, x_{2i+1}, \dots\}$ as a basis for the linear dual $(\Sigma^{-1} D_1^o)^*$, where x_{2i+1} is dual to d_1^{2i+1}, so each x_{2i+1} is in degree $2i$ in the desuspension. Consider $M = (\Sigma^{-1} D_1^o)^*$ as an unstable right $\mathcal{A}$-module by defining the linear map $x_i \cdot \mathrm{Sq}^k$ as

$$(x_i \cdot \mathrm{Sq}^k)(y) = x_i(\mathrm{Sq}^k\, y) \quad \text{for } y \in D_1^o.$$

We can use Bousfield's tower detector to determine where there are towers in $\mathrm{Ext}^{s,t}_{\mathcal{U}_{\mathcal{R}}}(\mathbb{F}_2, M)$. As defined in Definition 4.2, $T^s(M)$ is constructed so that in degree 2 and above the next degree is constructed from the previous by multiplying by λ_0 on the right and the differential from degree 1 onward is the same as the previous degree but with an extra factor of λ_0 on the right. Thus, it is sufficient to compute $H^2(T(M))$ to determine where the towers are.

We can see in Definition 4.2 that the differential δ only involves Sq^1 and Sq^{2k} and thus in this case is determined by the fact that

$$x_i \cdot \mathrm{Sq}^1 = 0 \quad \text{and} \quad x_{4k+1} \cdot \mathrm{Sq}^{2k} = x_{2k+1}.$$

Elements of $H^2(T(M))$ take three forms. First, all $x_{4k+1} \otimes (\lambda_0)^2$ are not cycles since $\delta(x_{4k+1} \otimes (\lambda_0)^2) = x_{2k+1} \otimes \lambda_{2k-1}(\lambda_0)^2$. Second, all $x_{2k+1} \otimes \lambda_{2k-1}\lambda_0$ are cycles, but also boundaries hit by $x_{4k+1} \otimes \lambda_0$. Finally, all $x_{4k-1} \otimes (\lambda_0)^2$ are cycles and cannot be boundaries since the image of $T^1(M)$ is only elements of the form $x_{2k+1} \otimes \lambda_{2k-1}\lambda_0$. Thus, we get a tower for $x_{4k-1} \otimes (\lambda_0)^s$ in degree $4k-2$ for each positive integer k. □

Proof of Theorem 4.4 We know that D_n is generated by n elements, one in each degree $2^n - 2^i$ for $0 \le i \le n-1$. Let these generators be represented by $d_{2^n-2^i}$. Then an arbitrary basis element of D_n^o is of the form $d^{a_1}_{2^n-2^{n-1}} \cdots d^{a_n}_{2^n-2^0}$ with at least one a_i odd. Let $x_{a_1,\dots,a_n} \in (D_n^o)^*$ denote the linear dual of $d^{a_1}_{2^n-2^{n-1}} \cdots d^{a_n}_{2^n-2^0}$ in degree $\sum_{i=1}^n a_i(2^n - 2^{n-i})$.

Working now in $(\Sigma^{-1}D_n^o)^*$, $x_{a_1,\dots,a_n} \in (D_n^o)^*$ will now be in degree $\sum_{i=1}^n a_i(2^n - 2^{n-i}) - 1$. Let $M = (\Sigma^{-1}D_n^o)^*$ and consider the tower detecting complex $T(M)$. As described in the proof of Theorem 4.3, it is sufficient to calculate $H^2(T(M))$ to determine the location of the towers.

Since the differential δ of Definition 4.2 only uses Sq^1 and Sq^{2k},we need only understand the right action of Sq^1 on an arbitrary $x_{a_1,\dots,a_n}$ and the right action of Sq^{2k} on $x_{a_1,\dots,a_n}$ with $\sum_{i=1}^n a_i(2^n - 2^{n-i}) = 4k+1$. These are given by the formulas

$$x_{a_1,\dots,a_n} \cdot \mathrm{Sq}^1 = \begin{cases} x_{a_1,\dots,a_{n-2},a_{n-1}+1,a_n-1} & \text{if } a_{n-1} \text{ even, } a_n \geq 1, \text{ at least one } a_i \text{ odd,} \\ 0 & \text{otherwise,} \end{cases}$$

and, if $\sum_{i=1}^n a_i(2^n - 2^{n-i}) = 4k+1$,

$$x_{a_1,\dots,a_n} \cdot \mathrm{Sq}^{2k} = \begin{cases} x_{(a_1+2)/2,a_2/2,\dots,a_{n-1}/2,(a_n-1)/2} & \text{if } a_n \text{ odd, } a_i \text{ even,} \\ x_{a_1/2,\dots,a_{j-1}/2,(a_j-1)/2,(a_{j+1}+2)/2,a_{j+2}/2,\dots,a_{n-1}/2,(a_n-1)/2} & \text{if } a_j, a_n \text{ odd, } a_i \text{ even for } i \neq j, n, \\ x_{a_1/2,\dots,a_{n-2}/2,(a_{n-1}-1)/2,(a_n+1)/2} & \text{if } a_{n-1}, a_n \text{ odd, } a_i \text{ even for } i < n-1, \\ 0 & \text{otherwise,} \end{cases}$$

where throughout i and j are between 1 and $n-1$ inclusively.

With these formulas in hand, we analyze $H^2(T(M))$. Our strategy will be to first characterize all cycles in $H^2(T(M))$ and then go through each type of cycle to determine which are boundaries. All of those that are not boundaries will correspond to our tower generators. There are four types of cycles:

(a) $x_{a_1,\dots,a_n} \otimes (\lambda_0)^2$ with $\sum_{i=1}^n a_i(2^n - 2^{n-i}) \neq 4k+1$ and a_{n-1} odd;

(b) $x_{a_1,\dots,a_n} \otimes (\lambda_0)^2$ with a_{n-1} even and $a_n = 0$;

(c) $x_{a_1,\dots,a_n} \otimes (\lambda_0)^2$ with $\sum_{i=1}^n a_i(2^n - 2^{n-i}) = 4k+1$ and a_n, a_{n-1} and some other a_i with $1 \leq i < n-1$ odd;

(d) $x_{a_1,\dots,a_n} \otimes \lambda_{2k-1}\lambda_0$ with $\sum_{i=1}^n a_i(2^n - 2^{n-i}) = 2k+1$.

To see that these are all of the cycles, we look at our formulas for the action of Sq^1 and Sq^{2k}. As long as the degree of $x_{a_i,\dots,a_n}$ is not a multiple of four, the differential on $x_{a_i,\dots,a_n} \otimes (\lambda_0)^2$ only involves Sq^1. So, if $\sum_{i=1}^n a_i(2^n - 2^{n-i}) - 1 \neq 4k$, $x_{a_i,\dots,a_n} \otimes (\lambda_0)^2$ is a cycle when $x_{a_i,\dots,a_n} \cdot \mathrm{Sq}^1 = 0$. This means either a_{n-1} is odd and we get cycles of type (a), $a_n = 0$ and a_{n-1} is even and we get cycles of type (b), or all a_i are even, but then $x_{a_i,\dots,a_n}$ is not an element of $(D_1^o)^*$.

If $x_{a_i,\dots,a_n}$ is in degree $4k$, the differential on $x_{a_i,\dots,a_n} \otimes (\lambda_0)^2$ involves both Sq^1 and Sq^{2k}. Then $\sum_{i=1}^n a_i(2^n - 2^{n-i}) - 1 = 4k$, which means that $a_n > 0$ and a_{n-1} is even. For both Sq^1 and Sq^{2k} to act trivially, a_{n-1} must be odd and some other a_i for $1 \leq i < n-1$ is also odd to give the cycles of type (c).

Finally, we have the elements of the form $x_{a_i,\dots,a_n} \otimes \lambda_{2k-1}(\lambda_0)^{s-1}$, which are all cycles by definition and give us the cycles in class (d).

Beginning with cycles of class (a), we want to determine which are also boundaries. For cycles in (a), we will split into three cases based on the value of $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) \bmod 4$.

If $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) = 4k-1$ and a_{n-1} is odd, then a_n is odd and

$$\delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0) = x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2$$

as long as one of $a_1, \dots, a_{n-2}$ is odd. Thus, we see that the only cycles not hit by boundaries are $x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2$ with a_{n-1}, a_n odd and a_i even for $1 \le i \le n-2$.

If $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) = 4k-2$ and a_{n-1} is odd, then a_n is even and

$$\delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0) = x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2$$

and we see that all these cycles are also boundaries.

If $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) = 4k$ and a_{n-1} is odd with $a_n \ge 2$, then $x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2$ is the boundary of

$$\begin{cases} \delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0 \\ \qquad + x_{a_1+2,a_2,\dots,a_{n-2},a_{n-1},a_n-1} \otimes \lambda_0) & \text{if } a_{n-1} \text{ odd, all other } a_i \text{ even,} \\ \delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0 \\ \qquad + x_{a_1,\dots,a_{j-1},a_j-1,a_{j+1}+2,a_{j+2},\dots,a_{n-2},a_{n-1},a_n+1} \otimes \lambda_0) & \text{if } a_j, a_{n-1} \text{ odd, all other } a_i \text{ even,} \\ \delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0) & \text{if } a_{n-1} \text{ odd, at least two other } a_i \text{ odd,} \end{cases}$$

and we see that all these cycles are also boundaries. This now covers all cases for our class (a) cycles.

Turning to our class (b) cycles, we see that none of these are boundaries. Indeed, they would need to be in the image of some $x_{b_1,\dots,b_n} \otimes (\lambda_0)$ with $x_{b_1,\dots,b_n} \cdot \mathrm{Sq}^1 = x_{a_1,\dots,a_n}$. However, Sq^1 changes the parity of each of the last two indices and is only nonzero if b_{n-1} is even, but then its image a_{n-1} must be odd, a contradiction.

Next, all class (c) cycles are boundaries hit by

$$\delta(x_{a_1,\dots,a_{n-2},a_{n-1}-1,a_n+1} \otimes \lambda_0) = x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2.$$

Thus, all class (c) cycles are boundaries as well.

Finally, for class (d), if $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) = 2k+1$, then

$$\delta(x_{2a_1,\dots,2a_{n-2},2a_{n-1}+1,2a_n-1} \otimes \lambda_0) = x_{a_1,\dots,a_n} \otimes \lambda_{2k-1}\lambda_0.$$

So these too are all boundaries.

This leaves us with only those classes of type (a) corresponding to $x_{a_1,\dots,a_{n-2},a_{n-1},a_n} \otimes (\lambda_0)^2$ with a_{n-1}, a_n odd and a_i even for $1 \le i \le n-2$, and all cycles from class (b) with $x_{a_1,\dots,a_n} \otimes (\lambda_0)^2$ with a_{n-1} even and $a_n = 0$ such that $\sum_{i=1}^{n} a_i(2^n - 2^{n-i}) = 4k$ in $H^2(T(M))$. □

While these calculations immediately give towers in the width filtration spectral sequence, we do not perform the analysis to determine which survive to the E_2 of the CWSS, instead citing agreement with Wellington's towers in the E_2 of the CWSS. His argument is substantially more involved than what was required above, so we would like to find a self-contained argument at some point.

It would be interesting to investigate the elements of homotopy which correspond to these towers, which must of course all support or receive differentials. Would the resulting finite towers at E_∞ be exceptional in any way?

5 First analysis of J map

We start with a fun observation. The Bousfield result [4] equating the E_2 of the unstable Adams spectral sequence with an Ext in unstable modules can be applied to $\operatorname{Ext}^{s,t}_{\mathcal{U}\mathcal{A}}(H^*(BO),\mathbb{F}_2)$, as of course $H^*(BO)$ is polynomial on the Stiefel–Whitney classes. We get

$$\operatorname{Ext}^{s,t}_{\mathcal{U}\mathcal{A}}(H^*(BO),\mathbb{F}_2)\cong\operatorname{Ext}^{s,t-1}_{\mathcal{U}}(\Sigma^{-1}\operatorname{Ind}H^*(BO),\mathbb{F}_2).$$

Up to decomposables, $\operatorname{Sq}^i(w_j)=\binom{j-1}{i}w_{j+i}$, where w_j is the j^{th} Stiefel Whitney class [17; 27; 29]. Preliminary calculations for $\operatorname{Ext}_{\mathcal{U}}$ led to the observation that its Ext chart looked like $\operatorname{Ext}_{\mathcal{U}}(\Sigma^{-1}D_1^o)$, but shifted to the right. Recall that D_1^o is the cohomology of $\mathbb{R}P^\infty$, modulo squares. This then led us to the following isomorphism between familiar modules, which we found surprising.

Proposition 5.1 $$\operatorname{Ext}^{s,t}_{\mathcal{U}}(\Sigma^{-1}D_1^o)\cong\operatorname{Ext}^{s,t+1}_{\mathcal{U}}(\Sigma^{-1}\operatorname{Ind}H^*(BO)).$$

This hints that the lowest degree in the width filtration yields the image of J, an idea that we expand on later in this section. Before proving this proposition, we recall notation from [26]. Define $\Phi\colon\mathcal{U}\to\mathcal{U}$ on $M\in\mathcal{U}$ at the prime 2 by

$$(\Phi M)^n\cong M^{n/2},\quad \operatorname{Sq}^i\Phi x=\Phi\operatorname{Sq}^{i/2}x,$$

where $M^{n/2}$ is trivial if $n/2$ is not an integer. Define $\lambda_M\colon\Phi M\to M$ by $\Phi x\mapsto\operatorname{Sq}^{|x|}x$. Define the functor Ω and its first (and only nontrivial) left derived function Ω_1 by

$$\ker\lambda_M=\Sigma\Omega_1 M\quad\text{and}\quad\operatorname{coker}\lambda_M=\Sigma\Omega M.$$

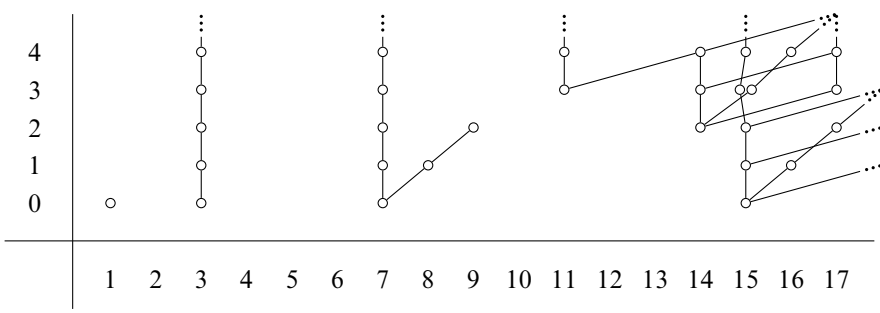


Figure 7: Calculations of $\operatorname{Ext}_{\mathcal{U}}(BO,\mathbb{F}_2)$.

Proposition 5.2 *If $M \in \mathcal{U}$ and λ_M is injective, then* $\mathrm{Ext}^{s,t}_{\mathcal{U}}(\Omega M, \mathbb{F}_2) \cong \mathrm{Ext}^{s,t+1}(M, \mathbb{F}_2)$.

Proof We know that

$$0 \to \ker(\lambda_M) \to \Phi(M) \to M \to \mathrm{coker}(\lambda_M) \to 0$$

is an exact sequence where $\ker(\lambda_M) \cong \Sigma\Omega_1 M$ and $\mathrm{coker}(\lambda_M) \cong \Sigma\Omega M$. Since λ_M is injective, $\ker(\lambda_M) = 0$, so $\Omega_i M = 0$ for all $i > 0$. Let $P_\bullet \to M$ be a free resolution of M. Then $\Omega P_\bullet \to \Omega M$ is a free resolution of ΩM. Thus, $\mathrm{Ext}_{\mathcal{U}}(\Omega M, \mathbb{F}_2) \cong H_s(\mathrm{Hom}^t(\Omega P_\bullet, \mathbb{F}_2))$. Since Ω is left adjoint to Σ, we get

$$H_s(\mathrm{Hom}^t(\Omega P_\bullet, \mathbb{F}_2)) \cong H_s(\mathrm{Hom}^t(P_\bullet, \Sigma\mathbb{F}_2)) \cong H_s(\mathrm{Hom}^{t+1}(P_\bullet, \mathbb{F}_2)) \cong \mathrm{Ext}^{s,t+1}_{\mathcal{U}}(M, \mathbb{F}_2). \qquad \square$$

Proof of Proposition 5.1 The result will follow from Proposition 5.2 if we can show that λ_M is injective for $M = \Sigma^{-1}\,\mathrm{Ind}\, H^*(BO)$ and $\mathrm{coker}\,\lambda_M \cong \Sigma\Sigma^{-1} D_1^o$.

For any $i > 0$, $w_i \in M$, w_i has degree $i-1$ and

$$\lambda_M(w_i) = \mathrm{Sq}_0(w_i) = \mathrm{Sq}^{i-1}(w_i) = \binom{i-1}{i-1} w_{i+i-1} = w_{2i-1},$$

so λ_M is clearly injective. We can also see that the image of λ_M will be the span of w_{2i-1} for $i \geq 1$. Then $\mathrm{coker}\,\lambda_M$ will be $w_{2i} + \mathrm{im}(\lambda_M)$ in degrees $2i-1$. The map $w_{2i} + \mathrm{im}(\lambda_M) \mapsto d_1^{2i-1}$ then defines an isomorphism with $D_1^o \cong \Sigma\Sigma^{-1} D_1^o$, so we conclude that $\Sigma\Sigma^{-1} D_1^o \cong \mathrm{coker}\,\lambda_M$. Thus, applying Proposition 5.2, $\mathrm{Ext}^{s,t}(\Sigma^{-1} D_1^o) \cong \mathrm{Ext}^{s,t+1}(\Sigma^{-1}\,\mathrm{Ind}\, H^*(BO))$. $\square$

The suggestive calculation above for BO could tell us about the delooping of the image of J. But our calculations in this paper are based on the cohomology of QS^0 and not its delooping. For QS^0 itself, the first calculations above are consistent with the following.

Theorem 5.3 *The map induced by the image of J on cohomology induces a splitting of the Nakaoka module $\mathfrak{N}$.*

Corollary 5.4 *The algebraic map induced by the image of J on* Ext *induces a splitting of* $\mathrm{Ext}(\mathfrak{N})$.

To make our calculations of both the map induced by the image of J on cohomology and the induced map on Ext, we begin by studying the map ind on homology. We will use the basis for the homology of SO presented by Hatcher [14],

$$H_*(\mathrm{SO}, \mathbb{F}_2) \cong E[e^1, \ldots, e^n, \ldots],$$

and the basis for $H_*(Q_1 S^0)$ in Theorem 3.44 of Madsen and Milgram [20],

$$H_*(Q_1 S^0, \mathbb{F}_2) \cong E(g_1, \ldots, g_i, \ldots) \otimes P(g_I),$$

where $I = (i_1, \ldots, i_m)$ and runs over all sequences of integers $0 \leq i_1 \leq \cdots \leq i_m$ and $i_1 = 0$ implies $m = 2$ and $i_2 > 0$. We can translate between this basis and the Kudo–Araki–Dyer–Lashof basis of admissible q_I.

Proposition 5.5 [21, Theorem 3.9] *Let ι be the class of the identity map $(S^\infty, *) \to (S^\infty, *)$ in $Q_1 S^0$, let p be any prime, and let I be any admissible sequence of length j. Define*

$$s(Q_I(\iota)) = Q_I(\iota) * (\bar{\iota}) * \cdots * (\bar{\iota}),$$

where there are p^j copies of $\bar{\iota}$, the class of a point in $Q_{-1}S^0$. Then

$$H_*(Q_0 S^0, \mathbb{Z}_p) = \Lambda\{s(Q_I(\iota))\}$$

as an algebra.

In our application, $p = 2$. As discussed in the proof of [21, Theorem C], the product of these generators with ι (shifting to the 1-component) give the homology generators as described by

$$g_i \otimes 1 \mapsto q_i(\iota) * \bar{\iota}, \quad 1 \otimes g_{0,i} \mapsto q_i^2(\iota) * \bar{\iota}^3, \quad 1 \otimes g_I \mapsto q_I(\iota) * \bar{\iota}^{2^m - 1}.$$

The homology map induced by the J homomorphism follows from [21, Corollary 1.5], giving that $J_*\colon H_*(\mathrm{SO}) \to H_*(Q_1 S^0)$ is defined by $e_i \mapsto g_i \otimes 1$, which is $q_i(\iota) * \bar{\iota}$.

We can use this and the pairings between homology and cohomology for both SO and $Q_1 S^0$ to understand the induced map $J^*\colon \Sigma^{-1} \operatorname{Ind} H^*(Q_1 S^0) \to \Sigma^{-1} \operatorname{Ind} H^*(\mathrm{SO})$. For cohomology, use the basis from Hatcher [14],

$$H^*(\mathrm{SO}, \mathbb{F}_2) \cong \bigotimes_{i \text{ odd}} \mathbb{F}_2[\beta_i],$$

where β_i is the linear dual to e^i. To work out the pairing between homology and cohomology for SO, we inductively use the Hopf algebra structure.

In general, the induced map on cohomology is given by the following.

Proposition 5.6 *The map on cohomology indecomposables induced by the J homomorphism is*

$$\gamma_{1[1]}^{2k+1} \mapsto \beta_{2k+1}, \quad \gamma_{1[2]}^{i} \gamma_{2[1]} \mapsto \beta_{2i+3}$$

and all other indecomposables map to 0.

To prove this we make some preliminary calculations.

Lemma 5.7 *Odd-degree elements of $\mathfrak{N}$ pair trivially with any nontrivial product of two or more elements in $H_*(Q_1 S^0)$.*

Proof Let $x \in \mathfrak{N}$ be an indecomposable of odd degree and width 2^k. Any indecomposable of width 2^k in odd degree must contain at least one block of full width; thus, $\psi(x) = x \otimes 1 + 1 \otimes x$. Let q_{I_1} and q_{I_2} be nontrivial elements of $H_*(Q_1 S^0)$. Then

$$\langle x, q_{I_1} * q_{I_2} \rangle = \langle \psi(x), q_{I_1} \otimes q_{I_2} \rangle = 0,$$

using the product and coproduct structure of the homology and cohomology. □

Lemma 5.8
$$\beta^*_{2k+1} = e^{2k+1} + \sum_{a+b=2k+1} e^a \wedge e^b.$$

Proof We begin with the identity $\beta_{2k+1} = (e^{2k+1})^*$, since β_{2k+1} is primitive and pairs nontrivially with only e^{2k+1}. Thus, $\langle \beta_{2k+1}, e^{2k+1} + \sum_{a+b=2k+1} e^a \wedge e^b \rangle = 1$.

We would like to show that $\langle x, e^{2k+1} + \sum_{a+b=2k+1} e^a \wedge e^b \rangle = 0$ for any other $x \in H^*(Q_1 S^0)$. We begin by writing x as the sum of products of primitives

$$x = \sum_j \prod_i \beta^{2^{b_{i,j}}}_{2a_{i,j}+1}.$$

By linearity and degree arguments, it suffices to consider products of the form $x = \prod_{i=1}^n \beta^{2^{b_i}}_{2a_i+1}$ where $\deg x = \sum_i (2a_i+1)\cdot 2^{b_i} = 2k+1$. Then $\langle x, e^a \wedge e^b \rangle = 1$ if there are an odd number of ways to split up the n terms in the product into two products of total degrees a and b; otherwise, $\langle x, e^a \wedge e^b \rangle = 0$.

The number of ways to split the n terms into two groups is

$$\begin{cases} \sum_{i=1}^{\lfloor n/2 \rfloor} \binom{n}{i} & \text{if } n \text{ is odd,} \\ \sum_{i=1}^{n/2-1} \binom{n}{i} + \binom{n}{n/2} & \text{if } n \text{ is even.} \end{cases}$$

In either case, this is equivalent to

$$\frac{1}{2}\sum_{i=0}^{n} \binom{n}{i} - 1 = \frac{2^n}{2} - 1 = 2^{n-1} - 1,$$

which is odd for $n \geq 2$. Certainly not all of these pairs will be distinct, but they will cancel each other out in pairs, leaving an odd number of possibilities. Thus, $\langle x, e^a \wedge e^b \rangle = 1$. We can also calculate that $\langle x, e^{2k+1} \rangle = 1$, which means that

$$\left\langle x, e^{2k+1} + \sum_{a+b=2k+1} e^a \wedge e^b \right\rangle = 1 + 1 = 0$$

and β_{2k+1} is the only term in $H^*(Q_1 S^0)$ that pairs nontrivially with $e^{2k+1} + \sum_{a+b=2k+1} e^a \wedge e^b$. □

In the proofs that follow, we use the notation $[x^n] f(x)$ to denote the coefficient of x^n in $f(x)$.

Lemma 5.9
$$\sum_{i=0}^{m} \binom{n}{i} \equiv_2 \binom{n-1}{m}.$$

Proof We have

$$\sum_{i=0}^{m} \binom{n}{i} = [x^m]\frac{(1+x)^n}{1-x} \equiv_2 [x^m]\frac{(1+x)^n}{1+x} = [x^m](1+x)^{n-1} = \binom{n-1}{m}.$$ □

Lemma 5.10 *We have*

$$\binom{b-1}{k-b} = [x^{k-b}]F_k(x),$$

where

$$F_k(x) = \frac{C(-x)^{-k} - (-xC(-x))^k}{\sqrt{1-4x}} \equiv_2 C(x)^{-k} + (xC(x))^k$$

and $C(x)$ is the generating function for the Catalan numbers

$$C(x) = \frac{1-\sqrt{1-4x}}{2x}.$$

Lemma 5.11 *Fix $k > 0$. For $\lceil \frac{1}{3}(2k+1) \rceil \le j \le k$,*

$$\binom{j-1}{2j-k-1} + \sum_{b=\lceil (k+1)/2 \rceil}^{\lfloor (2k-j)/2 \rfloor} \binom{b-1}{2b-k-1}\binom{j-b-1}{2k-j-2b} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise.} \end{cases}$$

Proof We know that $\binom{b-1}{2b-k-1} = \binom{b-1}{k-b}$. Applying Lemma 5.10, we get

$$\binom{b-1}{k-b} = [x^{k-b}]F_k(x),$$

and

$$\binom{j-b-1}{2k-j-2b} = [x^{2j-(2k+1)+b}]F_{3j-2k-1}(x).$$

Given the range on j, we know that $3j-2k-1 \le k-1$. Thus, $k > 3j-2j-1$, so

$$\begin{aligned}\sum_b \binom{b-1}{2b-k-1}\binom{j-b-1}{2k-j-2b} &= \sum_b ([x^{k-b}]F_k(x))([x^{2j-(2k+1)+b}]F_{3j-2k-1}) \\ &= [x^{2j-k-1}]F_k(x)\cdot F_{3j-2k-1}(x),\end{aligned}$$

which is the coefficient of x^{2j-k-1} in $F_k(x)\cdot F_{3j-2k-1}(x)$. Expanding out this product, we get

$$\begin{aligned}&F_k(x)\cdot F_{3j-2k-1}(x) \\ &\quad= (C(x)^{-k} + x^k C(x)^k)(C(x)^{-3j+2k+1} + x^{3j-2k-1}C(x)^{3j-2k-1}) \\ &\quad= C(x)^{-3j+k+1} + x^k C(x)^{-3j+3k+1} + x^{3j-2k-1}C(x)^{3j-3k-1} + x^{3j-k-1}C(x)^{3j-k-1} \\ &\quad= C(x)^{-(3j-k-1)} + x^{3j-2k-1}(C(x)^{-(3k-3j+1)} + x^{3k-3j+1}C(x)^{3k-3j+1}) + (xC(x))^{3j-k-1} \\ &\quad= F_{3j-k-1} + x^{3j-2k-1}F_{3k-3j+1}.\end{aligned}$$

Then the x^{2j-k-1} coordinate is $[x^{2j-k-1}]F_{3j-k-1} + [x^{k-j}]F_{3k-3j+1} = \binom{j-1}{2j-k-1} + \binom{2k-2j}{k-j}$. Putting everything together, we have

$$\begin{aligned}\binom{j-1}{2j-k-1} + \sum_{b=\lceil (k+1)/2 \rceil}^{\lfloor (2k-j)/2 \rfloor} \binom{b-1}{2b-k-1}\binom{j-b-1}{2k-j-2b} &= \binom{j-1}{2j-k-1} + \binom{j-1}{2j-k-1} + \binom{2k-2j}{k-j} \\ &= \binom{2(k-j)}{k-j},\end{aligned}$$

which is 1 if and only if $j = k$. □

Now we are ready to prove Proposition 5.6.

Proof of Proposition 5.6 Since $H^*(\mathrm{SO})$ has indecomposables in only odd degrees, we only need to understand the map on odd-degree elements in $\mathfrak{N}$ and thus only need to keep track of homology classes that are dual to odd-degree indecomposables. Then, by Lemma 5.7, we can ignore any products of two or more classes in $H_*(Q_1S^0)$ as we work out the maps. Additionally, for each odd degree there is exactly one indecomposable in $H^*(\mathrm{SO})$, namely β_{2k+1}. Then Lemma 5.8 tells us β_{2k+1} is a term in $(e^{2k+1})^*$ and $(e^a \wedge e^b)^*$ but not in the linear dual of any products of three or more terms. So it suffices to look at only products of one and two e^i in $H_*(\mathrm{SO})$ since products of three or more are dual only to decomposables.

The homology map induced by the J homomorphism is $J_*\colon H_*(\mathrm{SO}) \to H_*(Q_1S^0)$, which is defined by $e_i \mapsto g_i \otimes 1 = q_i(\iota) * \bar{\iota}$ in the Kudo–Araki–Dyer–Lashof basis. It follows that

$$e^a \wedge e^b \mapsto (q_a(\iota) * \bar{\iota}) \circ (q_b(\iota) * \bar{\iota}) = \sum_{l=(a-b)/2}^{a/2} \binom{a-l}{a-2l} q_{b+2l-a,a-l} + \text{products}.$$

We consider all $a+b = 2k+1$. Then $(q_{2k+1-2j,j})^* \mapsto \sum (e^a \wedge e^b)^*$, where the sum runs over all $e^a \wedge e^b$ for which $q_{2k+1-2j,j}$ was one of the terms in the image under J_*. Summing over all $a+b=2k+1$, the coefficient for admissible $q_{2k+1-2j,j}$ will be

$$\sum_{i=0}^{2j-k-1} \binom{j}{i} + \sum_{b=\lceil (k+1)/2 \rceil}^{\lfloor (2k-j)/2 \rfloor} \left(\sum_{l=0}^{2b-k-1} \binom{b}{l} \right) \binom{j-b-1}{2j-(2k+1)+b},$$

which simplifies, by Lemma 5.9, mod 2 to

$$\binom{j-1}{2j-k-1} + \sum_{b=\lceil (k+1)/2 \rceil}^{\lfloor (2k-j)/2 \rfloor} \binom{b-1}{2b-k-1} \binom{j-b-1}{2k-j-2b}.$$

By Lemma 5.11, we see that $(q_{1,k})^*$ contains an odd number of $(e^a \wedge e^b)^*$ in its image and thus β_{2k+1} is one of the terms in the image. For all other $q_{2k+1-2j,j}$, there are an even number of $(e^a \wedge e^b)^*$ in the image and thus β_{2k+1} does not appear as one of the terms in the linear dual, which leaves only decomposables. So only $J^*((q_{1,k})^*)$ is nonzero up to decomposables.

Next, for $2a+3b = 2k+1$,

$$\langle \gamma_{1[2]}^a \gamma_{2[1]}^b, q_{1,k} \rangle = \begin{cases} 1 & \text{if } a=1, b=k, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\gamma_{1[2]}^{k-1}\gamma_{2[1]}$ appears as a term in $(q_{1,k})^*$ and is not a term in any other $(q_{i,j})^*$. So

$$J^*(\gamma_{1[2]}^{k-1}\gamma_{2[1]}) = J^*((q_{1,k})^*) + \sum J^*((q_{i,j})^*) = \beta_{2k+1} + 0.$$

Any other $\gamma_{1[2]}^a \gamma_{2[1]}^b$ when expressed as a sum of the $(q_{i,j})^*$ will have only terms with $i > 1$ and thus J^* will contain only decomposables. □

For an example of how this works in degree 15 ($k=7$), see Section A.2.

Theorem 5.3 on the splitting of the Nakaoka module follows from the map on cohomology given in Proposition 5.6.

Proof of Theorem 5.3 At the level of unstable modules, we have the inclusion map $\Sigma^{-1} D_1^o \to \Sigma^{-1}\mathfrak{N}$ and $J^*\colon \Sigma^{-1}\mathfrak{N} \to \Sigma^{-1}\operatorname{Ind} H^*(\mathrm{SO})$, whose composition induces an isomorphism between the image of the first filtration D_1^o in $\mathfrak{N}$ and the indecomposables of $H^*(\mathrm{SO})$. □

5.1 Comparison of width filtration and chromatic filtration

The connection established between the lowest degree in the width filtration and the image of J invites a discussion of how the width filtration compares with the chromatic filtration. The chromatic filtration is a decreasing filtration while the width filtration is an increasing filtration. However, we can still compare the subquotients. The first subquotient in the width spectral sequence is D_1^o, which is exactly the image of J in cohomology, as proved in the previous section. The first subquotient in the chromatic filtration consists of the v_1-periodic elements, which at the prime 2 includes both the image of J and additional classes from the α family in degrees $8k+1$ and $8k+2$ [25]. In the width filtration, the first of these classes shows up in the D_3^o subquotient. So the width filtration differs from the chromatic filtration, and could faithfully rather than "approximately" extend the image of J to a full filtration.

Appendix

A.1 The width spectral sequence

The chart in Figure 8 depicts the E_1 page of the width spectral sequence as an $\mathbb{F}_2$-vector space. There is also a partial action of $\operatorname{Ext}_{\mathscr{A}}(\mathbb{F}_2,\mathbb{F}_2)$, as described in Section 4, with the actions of h_0, h_1 and h_2 indicated with black connecting lines.

A.2 J^* in degree 15

We will illustrate what is happening in the proof of Proposition 5.6 in degree 15 ($k=7$). Using Lemma 5.7 to ignore nontrivial products in $H_*(Q_1 S^0)$ and Lemma 5.8 to ignore products of three or more terms in $H_*(\mathrm{SO})$, the relevant homology maps are

$$J_*(e^{15}) = q_{15}$$

and, for $a+b=15$,

$$J_*(e^a \wedge e^b) = \sum_{l=\lceil (a-b)/2 \rceil}^{a/2} \binom{a-l}{a-2l} q_{b+2l-a,a-l} + \text{products}.$$

We conclude that $J^*(q_{15}^*) = J^*(\gamma_{1[1]}^{15}) = (e^{15})^* = \beta_{15}$ and $J^*((q_{15-2j,j})^*) = \sum (e^a \wedge e^b)^*$, where the sum runs over all $e^a \wedge e^b$ for which $q_{15-2j,j}$ was one of the terms in the image under J_*. Thus, summing over the image of J_* for all $a+b=15$ and looking at the coefficient of $q_{15-2j,j}$ mod 2 will tell us whether $q_{15-2j,j}$ is in the image of an odd or even number of $e^a \wedge e^b$. For $q_{15-2j,j}$ to be admissible,

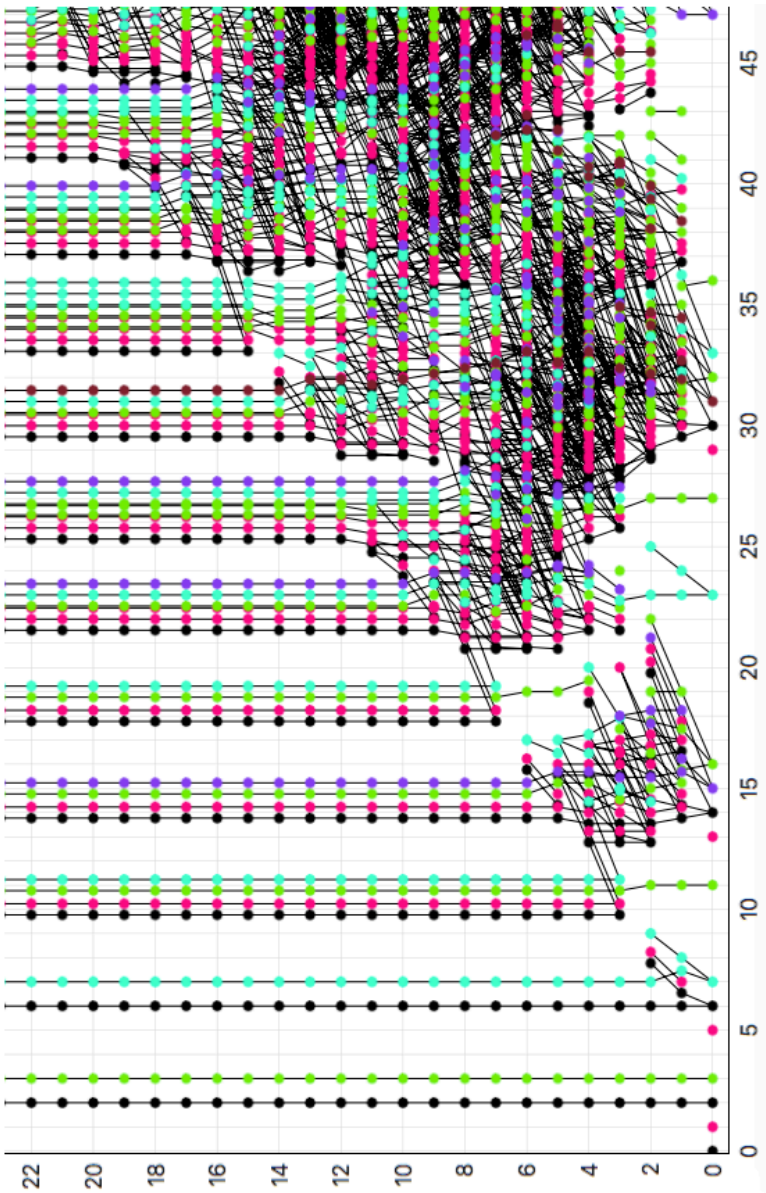


Figure 8: E_1 page of the width spectral sequence. Black corresponds to D_1, red to D_2, green to D_3, teal to D_4, purple to D_5, and brown to D_6.

we need $5 \leq j \leq 7$. We then calculate the coefficient for $q_{5,5}$ is

$$\sum_{i=0}^{2}\binom{5}{i}+\sum_{b=4}^{4}\left(\sum_{l=0}^{2b-8}\binom{b}{l}\right)\binom{5-b-1}{-5+b},$$

which by Lemma 5.9 is

$$\binom{4}{2}+\sum_{b=4}^{4}\binom{b-1}{2b-8}\binom{5-b-1}{-5+b}=0.$$

Repeating the same process for $q_{3,6}$, the coefficient is equivalent mod 2 to

$$\binom{6}{4}+\sum_{b=4}^{4}\binom{b-1}{2b-8}\binom{6-b-1}{-3+b}=1+1=0.$$

And finally, for $q_{1,7}$, the coefficient is equivalent mod 2 to

$$\binom{6}{6}+\sum_{b=4}^{3}\binom{b-1}{2b-8}\binom{7-b-1}{-1+b}=1+0=1.$$

Thus, $q_{1,7}$ is the only element appearing an even number of times as the image of the $e^a\wedge e^b$ and thus only $J^*((q_{1,7})^*)$ is nonzero up to decomposables.

Next, we know that

$$(q_{1,7})^*=\gamma_{1[2]}^6\gamma_{2[1]}+\gamma_{1[2]}^3\gamma_{2[1]}^3+\gamma_{2[1]}^5,\quad (q_{3,6})^*=\gamma_{1[2]}^3\gamma_{2[1]}^3+\gamma_{2[1]}^5\quad\text{and}\quad (q_{5,5})^*=\gamma_{2[1]}^5,$$

so

$$\begin{aligned}J^*(\gamma_{2[1]}^5)&=J^*(q_{5,5}^*)=0,\\ J^*(\gamma_{1[2]}^3\gamma_{2[1]}^3)&=J^*((q_{3,6})^*+(q_{5,5}^*))=0+0,\\ J^*(\gamma_{1[2]}^6\gamma_{2[1]})&=J^*((q_{1,7})^*+(q_{3,6})^*+(q_{5,5}^*))=\beta_{15},\end{aligned}$$

all up to decomposables.

References

[1] **A Adem**, **J Maginnis**, **R J Milgram**, *Symmetric invariants and cohomology of groups*, Math. Ann. 287 (1990) 391–411 MR

[2] **A Adem**, **R J Milgram**, *Cohomology of finite groups*, 2nd edition, Grundl. Math. Wissen. 309, Springer (2004) MR

[3] **M Barratt**, **S Priddy**, *On the homology of non-connected monoids and their associated groups*, Comment. Math. Helv. 47 (1972) 1–14 MR

[4] **A K Bousfield**, *A vanishing theorem for the unstable Adams spectral sequence*, Topology 9 (1970) 337–344 MR

[5] **A K Bousfield**, **E B Curtis**, *A spectral sequence for the homotopy of nice spaces*, Trans. Amer. Math. Soc. 151 (1970) 457–479 MR

[6] **A K Bousfield**, **E B Curtis**, **D M Kan**, **D G Quillen**, **D L Rector**, **J W Schlesinger**, *The* mod $-$ *p lower central series and the Adams spectral sequence*, Topology 5 (1966) 331–342 MR

[7] **F R Cohen**, **T J Lada**, **J P May**, *The homology of iterated loop spaces*, Lecture Notes in Math. 533, Springer (1976) MR

[8] **E B Curtis**, *The Dyer–Lashof algebra and the Λ-algebra*, Illinois J. Math. 19 (1975) 231–246 MR

[9] **E Curtis**, **M Mahowald**, *EHP computations of $E_2(S^n)$*, from "Computers in geometry and topology" (M C Tangora, editor), Lecture Notes in Pure and Appl. Math. 114, Dekker, New York (1989) 105–120 MR

[10] **E B Curtis**, **M Mahowald**, *The unstable Adams spectral sequence for S^3*, from "Algebraic topology" (M Mahowald, S Priddy, editors), Contemp. Math. 96, Amer. Math. Soc., Providence, RI (1989) 125–162 MR

[11] **M Feshbach**, *The mod 2 cohomology rings of the symmetric groups and invariants*, Topology 41 (2002) 57–84 MR

[12] **C Giusti**, **P Salvatore**, **D Sinha**, *The mod-2 cohomology rings of symmetric groups*, J. Topol. 5 (2012) 169–198 MR

[13] **P G Goerss**, *André–Quillen cohomology and the homotopy groups of mapping spaces: understanding the E_2-term of the Bousfield–Kan spectral sequence*, J. Pure Appl. Algebra 63 (1990) 113–153 MR

[14] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR

[15] **N H V Hung**, *The action of the Steenrod squares on the modular invariants of linear groups*, Proc. Amer. Math. Soc. 113 (1991) 1097–1104 MR

[16] **N H V Hu'ng**, **F P Peterson**, *Spherical classes and the Dickson algebra*, Math. Proc. Cambridge Philos. Soc. 124 (1998) 253–264 MR

[17] **S O Kochman**, *An algebraic filtration of $H_*(MO;\mathbb{Z}_2)$*, Proc. Edinburgh Math. Soc. 26 (1983) 313–317 MR

[18] **J Lannes**, **S Zarati**, *Invariants de Hopf d'ordre supérieur et suite spectrale d'Adams*, C. R. Acad. Sci. Paris Sér. I Math. 296 (1983) 695–698 MR

[19] **I Madsen**, *On the action of the Dyer–Lashof algebra in $H_*(G)$*, Pacific J. Math. 60 (1975) 235–275 MR

[20] **I Madsen**, **R J Milgram**, *The classifying spaces for surgery and cobordism of manifolds*, Annals of Mathematics Studies 92, Princeton Univ. Press (1979) MR

[21] **R J Milgram**, *The mod 2 spherical characteristic classes*, Ann. of Math. 92 (1970) 238–261 MR

[22] **H Miller**, *The Sullivan conjecture on maps from classifying spaces*, Ann. of Math. 120 (1984) 39–87 MR Correction in 121 (1984) 605–609

[23] **M Nakaoka**, *Homology of the infinite symmetric group*, Ann. of Math. 73 (1961) 229–257 MR

[24] **M Nakaoka**, *Note on cohomology algebras of symmetric groups*, J. Math. Osaka City Univ. 13 (1962) 45–55 MR

[25] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Math. 121, Academic, Orlando, FL (1986) MR

[26] **L Schwartz**, *Unstable modules over the Steenrod algebra and Sullivan's fixed point set conjecture*, Univ. of Chicago Press (1994) MR

[27] **R E Stong**, *Determination of $H^*(\mathrm{BO}(k,\dots,\infty),Z_2)$ and $H^*(\mathrm{BU}(k,\dots,\infty),Z_2)$*, Trans. Amer. Math. Soc. 107 (1963) 526–544 MR

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

[29] **W-t Wu**, *Les i-carrés dans une variété grassmannienne*, C. R. Acad. Sci. Paris 230 (1950) 918–920 MR

Department of Mathematics, University of Oregon
Eugene, OR, United States
Department of Mathematics, Kalamazoo College
Kalamazoo, MI, United States
Current address: *Mathematics Department, Fort Lewis College*
Durango, CO, United States
dfhunter@fortlewis.edu

Received: 12 October 2022 Revised: 14 June 2023

msp

Algebraic & Geometric Topology **25**:6 (2025) 3341–3384
DOI: 10.2140/agt.2025.25.3341
Published: 1 October 2025

The slices of quaternionic Eilenberg–Mac Lane spectra

BERTRAND J GUILLOU
CARISSA SLONE

We compute the slices and slice spectral sequence of integral suspensions of the equivariant Eilenberg–Mac Lane spectra $H\underline{\mathbb{Z}}$ for the group of equivariance Q_8. Along the way, we compute the Mackey functors $\underline{\pi}_{k\rho}H\underline{\mathbb{Z}}$.

55N91; 55P91, 55Q91

1. Introduction 3341
2. Background 3345
3. Inflation functors 3348
4. Q_8-Mackey functors and Bredon homology 3354
5. Review of the C_4-slices of $\Sigma^n H\mathbb{Z}$ 3361
6. Q_8-slices 3363
7. Homology calculations 3370
8. Slice spectral sequences 3377
References 3383

1 Introduction

Let G be a finite group. The G-equivariant slice filtration was first defined in the context of G-equivariant stable homotopy theory by Dugger in [6]; it came to prominence due to its role in the proof of the Kervaire invariant one problem by Hill, Hopkins and Ravenel [9]. The slice filtration is an analogue in the G-equivariant stable homotopy category of the classical Postnikov filtration of spectra. One can also define a G-equivariant Postnikov filtration; on passage to fixed points with respect to any subgroup $H \leq G$, this recovers the Postnikov filtration of the H-fixed point spectrum. However, there are many equivariant spectra which possess a periodicity with respect to suspension by a G-representation sphere, and this periodicity is not visible in the G-equivariant Postnikov filtration. The slice filtration was devised by Dugger in order to display this periodicity for the case of the C_2-spectrum $K\mathbb{R}$.

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Since the groundbreaking work [9], a number of authors have calculated the slice filtration, as well as the associated slice spectral sequence, for G-spectra of interest. A few cases are understood for an arbitrary finite group G. If $\underline{M}$ is a G-Mackey functor, then the equivariant Eilenberg–Mac Lane spectrum $H_G\underline{M}$ is always a 0-slice [9] (in this article, we use the "regular" slice filtration, as introduced by Ullman in [22]). The slice filtrations of $\Sigma^1 H_G\underline{M}$ and $\Sigma^{-1} H_G\underline{M}$ were described in [22]. The slices of certain suspensions of equivariant Eilenberg–Mac Lane spectra were determined for G an odd cyclic p-group by Angeltveit [1], Hill, Hopkins and Ravenel [10], and Yarnall [25]; for dihedral groups of order $2p$, where p is odd, by Zou [27]; and for the Klein four-group by Guillou and Yarnall [7] and Slone [19]. A majority of the work to date on the slice filtration has focused on cyclic groups of equivariance, with only a few exceptions. We extend the literature by considering in this article the case of the nonabelian group $G = Q_8$.

The choice of the group Q_8 is not arbitrary, as we now explain. Some of the most far-reaching applications of the slice filtration and associated spectral sequence have come in the case of cyclic p-groups of equivariance. In addition to [9], this also includes Hill, Hopkins and Ravenel [11], Hill, Shi, Wang and Xu [12], Meier, Shi and Zeng [18], and Sulyma [20]. In particular, in [12] the authors use slice technology to understand a C_4-equivariant, height 4 Lubin–Tate theory at the prime 2. For each height n, there is a height n Lubin–Tate theory that comes equipped with a continuous action of the height n (profinite) Morava stabilizer group. The homotopy fixed points with respect to this action gives a model for the $K(n)$-local sphere, a central object of study in stable homotopy theory. However, the homotopy fixed points with respect to the entire stabilizer group are quite difficult to calculate. More approachable are the homotopy fixed points with respect to finite subgroups of the stabilizer group. At height 4, the Morava stabilizer group contains a C_4-subgroup (in fact a C_8), which gives the context for [12]. On the other hand, at height $2m$, where m is odd, the Morava stabilizer group contains a Q_8-subgroup. Therefore it is reasonable to expect that Q_8-equivariant slice techniques will eventually shed light on the $K(n)$-local sphere when $n = 2m$ and m is odd.

The focus of our article is the determination of the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$, for $n \geq 0$. We list the slices in Section 6 and describe the associated spectral sequence in Section 8. We rely heavily on the computation of the slices of $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ given by the second author in [19]. The quotient map $Q_8 \to K_4$ allows us to gain insight into the Q_8-equivariant slices from the K_4-case, as we now explain in greater generality.

Given a normal subgroup $N \trianglelefteq G$, there are several constructions that will produce a G-spectrum from a G/N-spectrum. First is the ordinary pullback, or inflation, functor. If $q\colon G \to G/N$ is the quotient, then inflation is denoted $q^*\colon \mathbf{Sp}^{G/N} \to \mathbf{Sp}^G$; it is left adjoint to the N-fixed point functor. This inflation functor plays an important role. For instance $q^*(S^0_{G/N})$ is equivalent to S^0_G. However, from our point of view, this construction has two deficiencies. First, the ordinary inflation does not interact well with the slice filtration. Secondly, the inflation of an $H_{G/N}\underline{\mathbb{Z}}$-module does not have a canonical $H_G\underline{\mathbb{Z}}$-module structure.

On the other hand, the "geometric inflation" functor [8, Definition 4.1; 16, Section II.9]

$$\phi_N^* : \mathbf{Sp}^{G/N} \to \mathbf{Sp}^G,$$

which is right adjoint to the geometric fixed points functor, interacts well with slices. Namely, if N is a normal subgroup of order d and X is a G/N-spectrum, then

$$\phi_N^* P_k^k(X) \simeq P_{dk}^{dk}(\phi_N^* X),$$

by [22, Corollary 4.5] (see also [8, Section 4.2]). However, in general the geometric inflation of an $H_{G/N}\underline{\mathbb{Z}}$-module will not be an $H_G\underline{\mathbb{Z}}$-module.

The third variant is the $\underline{\mathbb{Z}}$-module inflation functor [26, Section 3.2]

$$\Psi_N^* : \mathrm{Mod}_{H_{G/N}\underline{\mathbb{Z}}} \to \mathrm{Mod}_{H_G\underline{\mathbb{Z}}}.$$

By design, the $\underline{\mathbb{Z}}$-module inflation of an $H_{G/N}\underline{\mathbb{Z}}$-module has a canonical $H_G\underline{\mathbb{Z}}$-module structure, though in general this functor does not interact well with the slice filtration.

In some cases, these constructions agree. For instance, if N is a bottleneck subgroup (Definition 3.5) and the underlying spectrum of the G/N-spectrum X is contractible, then $q^*X \simeq \phi_N^* X$ (see the proof of Proposition 3.18). If X is furthermore an $H_{G/N}\underline{\mathbb{Z}}$-module, then the three inflation functors coincide on X (Proposition 3.18).

The above discussion applies to the slices of $\Sigma^n H_{G/N}\underline{\mathbb{Z}}$: all slices, except for the bottom slice, have trivial underlying spectrum. It follows that these inflate to give many of the slices of $\Sigma^n H_G\underline{\mathbb{Z}}$.

Our main result along these lines, Theorem 3.19, describes the higher slices of such an inflated $H_G\underline{\mathbb{Z}}$-module. In the case of $G = Q_8$, $N = Z(Q_8)$, and $G/N = Q_8/Z \cong K_4$, it gives the following:

Theorem 1.1 *Let $n \geq 0$. Then the nontrivial slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$, above level $2n$, are*

$$P_{2k}^{2k}(\Sigma^n H_{Q_8}\underline{\mathbb{Z}}) \simeq \Psi_Z^* P_k^k(\Sigma^n H_{K_4}\underline{\mathbb{Z}}) \simeq \phi_Z^* P_k^k(\Sigma^n H_{K_4}\underline{\mathbb{Z}})$$

for $k > n$. Furthermore,

$$P_n^{2k}(\Sigma^n H_{Q_8}\underline{\mathbb{Z}}) \simeq \Psi_Z^* P_n^k(\Sigma^n H_{K_4}\underline{\mathbb{Z}})$$

for $k \geq n$.

As the slices of $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ were determined by the second author in [19, Section 8], this immediately provides all of the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ above level $2n$ when n is positive. In particular, the slices of $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ occur in slice dimension n to $4(n-4)$, and it follows that the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ occur in slice dimension n to $8(n-4)$.

In more detail, the slices of $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ are fully geometric (ie vanish at proper subgroups) in slice dimension greater than $2(n-3)$ and therefore occur only in slice dimensions a multiple of 4. It follows that the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ are fully geometric in slice dimension greater than $4(n-3)$ and only occur in slice dimensions a multiple of 8. See Proposition 6.10 for a more precise answer.

In slice dimensions from $n+1$ to $2(n-3)$, the second author showed that the K_4-spectrum $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ has some fully geometric slices (in dimensions a multiple of 4) as well as slices detected at the order two subgroups (in dimensions $4k$ if n is odd and $4k+2$ if n is even). It follows that the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$, in slice dimension from $2n+1$ to $4(n-3)$, occur only in slice dimensions a multiple of 4. See Proposition 6.11 for a more precise answer.

The remaining slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$, which occur in slice dimensions from n to $2n$, are then given by analyzing the slice tower of $\Psi_N^*(P_n^n H_K\underline{\mathbb{Z}})$. We perform this analysis in Section 6.1.

Though we have stated Theorem 1.1, as well as Theorem 3.19, for $n \geq 0$, a similar result holds for $n < 0$. The determination of the slices of $\Sigma^n H_{K_4}\underline{\mathbb{Z}}$ in [19] includes the case of $n < 0$, and so the methods in this article would similarly determine the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ for $n < 0$. See also Remark 3.20.

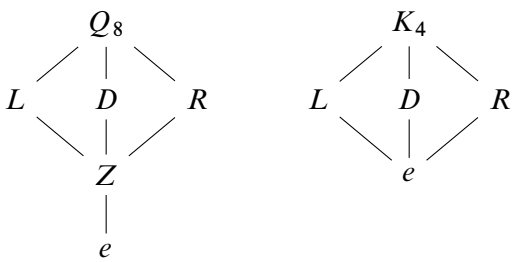
1.1 Notation

Throughout, whenever referencing the slice filtration, we will always mean the "regular" slice filtration of [22].

We will often write simply Q and K to denote the quaternion group Q_8 and Klein four group K_4, respectively. We write Z for the central subgroup of Q of order two generated by $z=-1$. We write

$$L = \langle i \rangle, \quad D = \langle k \rangle, \quad R = \langle j \rangle$$

for the normal, cyclic subgroups of Q of order 4. We also use the same names for the images of these subgroups in $Q/Z \cong K$. In other words, the subgroup lattices of Q_8 and K_4 are



Our nomenclature for the order 4 subgroups of Q_8 amounts to a choice of isomorphism $Q/Z \cong K$.

The sign representation of C_2 will be denoted σ, and we will write $\mathbb{Z}^\sigma$ for the corresponding C_2-module.

1.2 Organization

The paper is organized as follows. In Section 2, we review the representations of C_4, K_4, and Q_8, as well as Mackey functors over C_4 and K_4. Then in Section 3, we introduce three inflation functors from a quotient group G/N of some finite group G as well as several results that will aid in the calculation of the slices of $\Sigma^n H_{Q_8} H\underline{\mathbb{Z}}$. The relevant Q_8-Mackey functors and the homology of $\Sigma^{k\rho_{Q_8}} H_{Q_8}\underline{\mathbb{Z}}$ are found in Section 4. The slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ must restrict to the appropriate slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$; thus, we review this

information in Section 5. We provide some slice towers and describe all slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ in Section 6. We then compute the homotopy Mackey functors of the slices of $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ in Section 7. Finally, we provide some examples of the slice spectral sequence for $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ and $\Sigma^n H_{Q_8}\underline{\mathbb{Z}}$ in Section 8.

1.3 Acknowledgements

The authors are very happy to thank Agnès Beaudry, Michael Geline, Cherry Ng, and Mingcong Zeng for a number of helpful discussions. The spectral sequence charts in Section 8 were created using Hood Chatham's `spectralsequences` package.

The authors were supported by NSF grant DMS-2003204.

2 Background

2.1 Background for C_4

The C_4-sign representation σ_{C_4} is the inflation $p^*\sigma_{C_2}$ of the C_2-sign representation along the surjection $C_4 \to C_2$. We will simply write σ for σ_{C_4}. Then the regular representation for C_4 splits as

$$\rho_{C_4} = 1 \oplus \sigma \oplus \lambda,$$

where λ is the irreducible 2-dimensional rotation representation of C_4. The $RO(C_4)$-graded homotopy Mackey functors of $H_{C_4}\underline{\mathbb{Z}}$ are given in [11]. More specifically, the homotopy Mackey functors of $\Sigma^{k\rho_{C_4}} H_{C_4}\underline{\mathbb{Z}}$, $\Sigma^{k\lambda} H_{C_4}\underline{\mathbb{Z}}$, and $\Sigma^{k\sigma} H_{C_4}\underline{\mathbb{Z}}$ are given in [11, Figures 3 and 6]. Some C_4-Mackey functors that will appear below are displayed in Table 1. All of these Mackey functors have trivial Weyl-group actions.

$\square = \underline{\mathbb{Z}}$	$\circledast = \underline{\mathbb{Z}}^*$	$\underline{\mathbb{Z}}(2,1)$	$\circ = \underline{B}(2,0)$
$\mathbb{Z}$ $1 \circlearrowright 2$ $\mathbb{Z}$ $1 \circlearrowright 2$ $\mathbb{Z}$	$\mathbb{Z}$ $2 \circlearrowright 1$ $\mathbb{Z}$ $2 \circlearrowright 1$ $\mathbb{Z}$	$\mathbb{Z}$ $2 \circlearrowright 1$ $\mathbb{Z}$ $1 \circlearrowright 2$ $\mathbb{Z}$	$\mathbb{Z}/4$ $1 \circlearrowright 2$ $\mathbb{Z}/2$ 0
$\bullet = \underline{g}$	$\bar{\bullet} = \phi^* \underline{f}$	$\blacklozenge = \phi^* \underline{\mathbb{F}_2}$	$\phi^* \underline{\mathbb{F}_2}^*$
$\mathbb{F}_2$ 0 0	0 $\mathbb{F}_2$ 0	$\mathbb{F}_2$ $\downarrow 1$ $\mathbb{F}_2$ 0	$\mathbb{F}_2$ $1 \uparrow$ $\mathbb{F}_2$ 0

Table 1: Some C_4-Mackey functors.

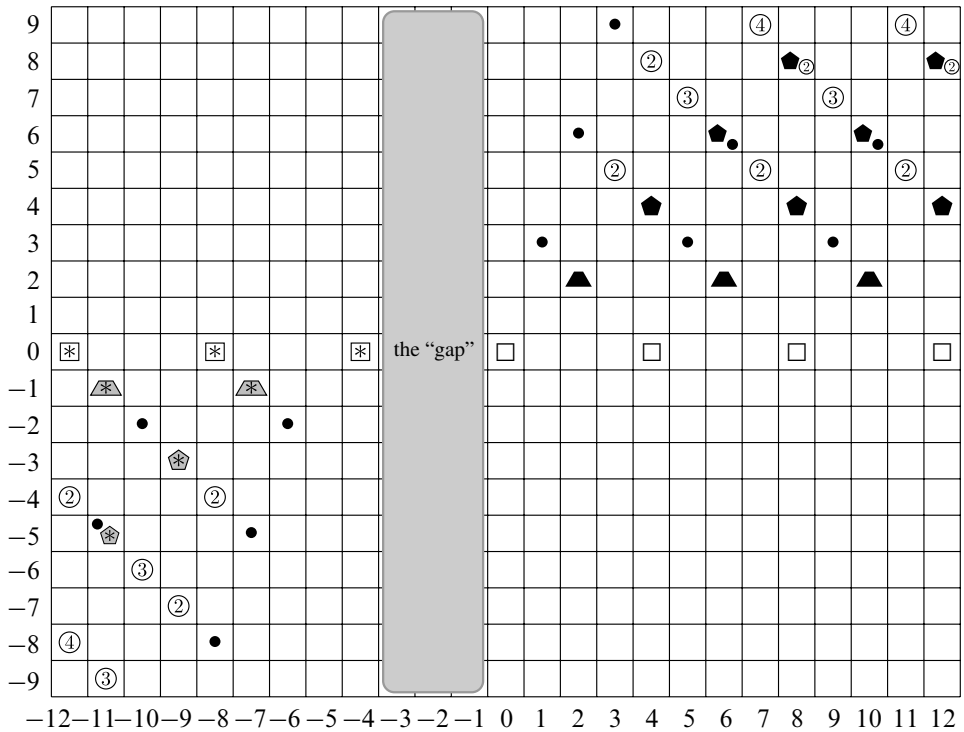


Figure 1: The homotopy Mackey functors of $\bigvee_n \Sigma^{n\rho} H_{K_4}\underline{\mathbb{Z}}$. The Mackey functor $\underline{\pi}_k \Sigma^{n\rho} H_{K_4}\underline{\mathbb{Z}}$ appears in position $(k, 4n-k)$.

2.2 Background for K_4

The Klein 4-group $K_4 = C_2 \times C_2$ has three sign representations, obtained as the inflation along the three surjections $K_4 \to C_2$. We denote these three surjections by p_1, m, and p_2. Then the regular representation of K_4 splits as

$$\rho_{K_4} \cong 1 \oplus p_1^*\sigma \oplus m^*\sigma \oplus p_2^*\sigma.$$

Some K_4-Mackey functors that will appear below are displayed in Table 2.

The homotopy Mackey functors of $\Sigma^{n\rho} H_K\underline{\mathbb{Z}}$ were computed in [19, Section 9]. They are displayed in Figure 1. The homotopy Mackey functors of $\Sigma^{n\rho} H_K\underline{\mathbb{F}_2}$ were computed in [7, Section 7]. They are displayed in Figure 2.

2.3 Background for Q_8

The regular representation of Q splits as

$$\rho_Q \cong \mathbb{H} \oplus \rho_K,$$

where $\mathbb{H}$ is the 4-dimensional irreducible Q_8-representation given by the action of the unit quaternions on the algebra of quaternions and ρ_K is the regular representation of K, inflated to Q along the quotient.

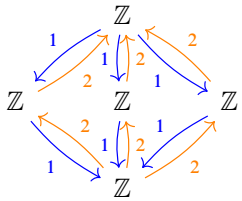
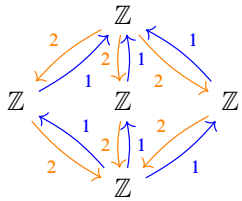
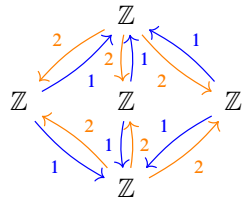
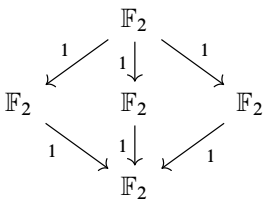
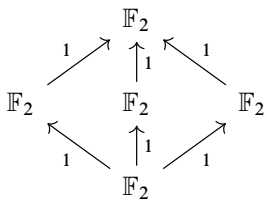
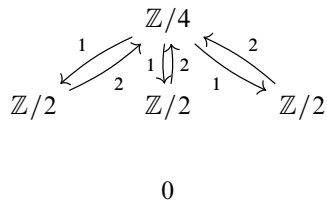
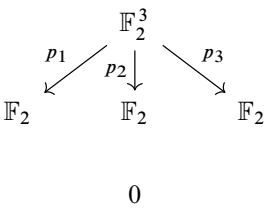
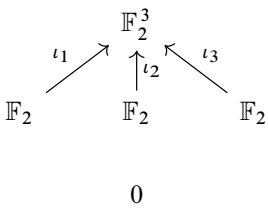
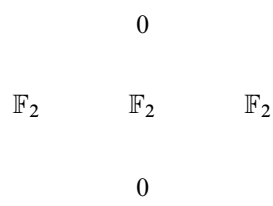
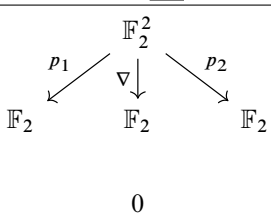
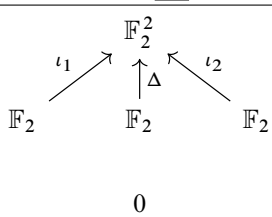
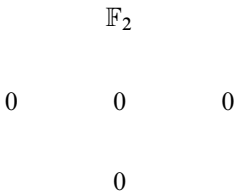
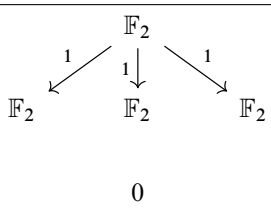
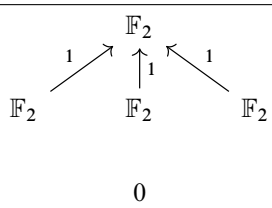
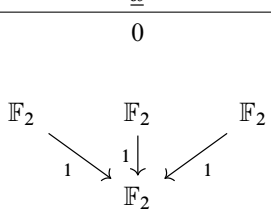
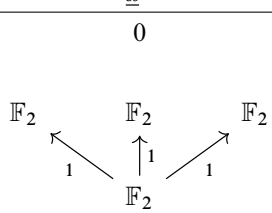
$\square = \underline{\mathbb{Z}}$	$\circledast = \underline{\mathbb{Z}}^*$	$\underline{\mathbb{Z}}(2,1)$
		
$\blacksquare = \underline{\mathbb{F}_2}$	$\circledast = \underline{\mathbb{F}_2}^*$	$\circ = \underline{B}(2,0)$
		
$\pentagon = \phi^*_{LDR}(\underline{\mathbb{F}_2})$	$\circledast = \phi^*_{LDR}(\underline{\mathbb{F}_2})^*$	$\phi^*_{LDR}(\underline{f})$
		
$\blacktriangle = \underline{mg}$	$\circledast = \underline{mg}^*$	$\bullet = \underline{g}$
		
$\underline{m}$	$\underline{m}^*$	
		
$\underline{w}$	$\underline{w}^*$	
		

Table 2: Some K_4-Mackey functors.

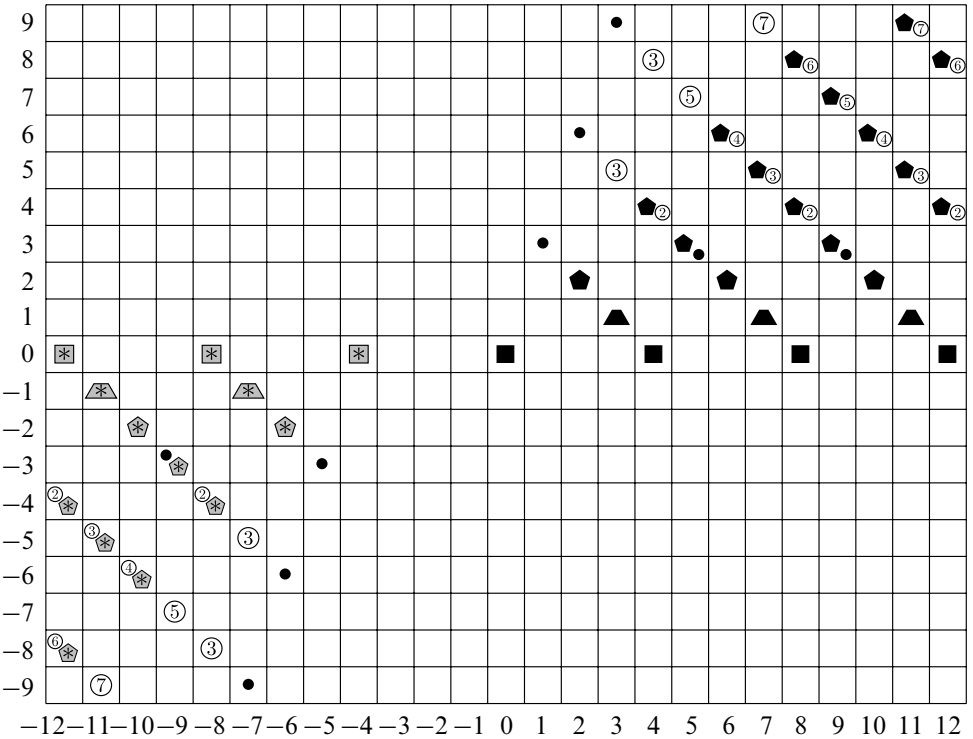


Figure 2: The homotopy Mackey functors of $\bigvee_n \Sigma^{n\rho} H_{K_4}\underline{\mathbb{F}_2}$. The Mackey functor $\underline{\pi}_k \Sigma^{n\rho} H_{K_4}\underline{\mathbb{F}_2}$ appears in position $(k, 4n-k)$.

Denoting by C_4 any of the subgroups L, D, or R of Q_8, we have that

$$\downarrow_{C_4}^{Q_8} \rho_K = 2 + 2\sigma \quad \text{and} \quad \downarrow_{C_4}^{Q_8} \mathbb{H} = 2\lambda.$$

3 Inflation functors

3.1 Inflation and the projection formula

Let $N \trianglelefteq G$ be a normal subgroup and $q\colon G \to G/N$ the quotient map. Recall that there is an induced adjunction

$$\mathbf{Sp}^{G/N} \underset{(-)^N}{\overset{q^*}{\rightleftarrows}} \mathbf{Sp}^G$$

where the pullback functor q^*, called inflation, is strong symmetric monoidal. We will also need a description of the N-fixed points of an Eilenberg–Mac Lane G-spectrum. First note that there is a functor

(3.1) $$\mathbf{Mack}(G) \xrightarrow{q_*} \mathbf{Mack}(G/N)$$

given by

$$q_*(\underline{M})(\overline{H}) = \underline{M}(H),$$

where $\overline{H} = H/N \leq G/N$ whenever $N \leq H$. The functor q_* is denoted by $\beta^!$ in [21, Lemma 5.4]. Then the homotopy Mackey functors of the N-fixed points of a G-spectrum X are given by

$$\underline{\pi}_n(X^N) \cong q_*\underline{\pi}_n(X). \tag{3.2}$$

In the case of an Eilenberg–Mac Lane spectrum this yields an equivalence

$$(H_G\underline{M})^N \simeq H_{G/N}(q_*\underline{M}).$$

The following result will be quite useful.

Proposition 3.3 (projection formula [14, Lemma 2.13; 2, Proposition 2.15]) *Let $N \trianglelefteq G$ be a normal subgroup and $q\colon G \to G/N$ be the quotient map. Then for $X \in \mathbf{Sp}^{G/N}$ and $Y \in \mathbf{Sp}^G$, there is a natural equivalence of G/N-spectra*

$$(q^*X \wedge Y)^N \simeq X \wedge Y^N.$$

We will frequently employ this in the case that $X = S^V$ for some G/N-representation V and $Y = H_G\underline{M}$ for some G-Mackey functor $\underline{M}$. Then the projection formula reads

$$(S^{q^*V} \wedge H_G\underline{M})^N \simeq S^V \wedge H_{G/N}(q_*\underline{M}). \tag{3.4}$$

See also [26, Corollary 5.8].

3.2 Geometric fixed points

For a normal subgroup $N \trianglelefteq G$, we define the family of subgroups $\mathcal{F}[N]$ of G to consist of those subgroups that do not contain N. Recall that the N-geometric fixed points spectrum of a G-spectrum is defined as

$$\Phi^N(X) = (\widetilde{E}\mathcal{F}[N] \wedge X)^N.$$

This notation is simultaneously used to denote the resulting G/N-spectrum as well as the underlying spectrum. The N-geometric fixed points has a right adjoint, given by the geometric inflation functor

$$\phi_N^*(Z) = \widetilde{E}\mathcal{F}[N] \wedge q^*Z.$$

To sum up, we have an adjunction

$$\mathbf{Sp}^G \underset{\phi_N^*}{\overset{\Phi^N}{\rightleftarrows}} \mathbf{Sp}^{G/N}.$$

3.3 Bottleneck subgroups

The subgroup $Z \trianglelefteq Q$ plays an important role in this article. The primary reason is that it satisfies the following property.

Definition 3.5 We say that $N \trianglelefteq G$ is a *bottleneck* subgroup if it is a nontrivial, proper subgroup such that, for any subgroup $H \leq G$, either H contains N or N contains H.

We now demonstrate that bottleneck subgroups only occur in cyclic p-groups or quaternion groups. The following argument was sketched to us by Mike Geline.

Proposition 3.6 *Let $N \trianglelefteq G$ be a bottleneck subgroup of G. Then N is cyclic, and G is either a cyclic p-group or a generalized quaternion group.*

Proof We will refer to a subgroup $H \leq G$ which neither contains N nor is contained in N as "adjacent" to N. The assumption that N is a bottleneck subgroup means precisely that G has no subgroups that are adjacent to N. To see that N must be cyclic, note that if g is not in N, then $N \leq \langle g \rangle$, which implies that N is cyclic.

We next observe that G is necessarily a p-group. This is because if N is contained in some Sylow p-subgroup, then any Sylow q-subgroup, for a different prime q, would be adjacent. It follows that N contains all of the Sylow subgroups and therefore is all of G.

Next, we recall [5, Theorem 4.3] that for a p-group G, the group contains a unique subgroup of order p if and only if G is either cyclic or generalized quaternion. So we will argue that G contains a unique subgroup of order p. The first step is to note that G cannot contain a subgroup isomorphic to $C_p \times C_p$. This is because such a subgroup would necessarily contain N. This would imply that $N \cong C_p$, and then N would have a complement in $C_p \times C_p$, which would be a subgroup adjacent to N in G.

Finally, note that the center $Z(G)$ contains a subgroup of order p. If G has another subgroup of order p, these two would generate a $C_p \times C_p$, contradicting the previous step. □

Remark 3.7 It follows from Proposition 3.6 that if $N \trianglelefteq G$ is a bottleneck subgroup, then G/N is either a cyclic p-group or a dihedral 2-group.

If $N \trianglelefteq G$ is a bottleneck subgroup, then geometric fixed points with respect to G can be computed in terms of geometric fixed points with respect to the quotient group G/N.

Proposition 3.8 *Let $N \trianglelefteq G$ be a bottleneck subgroup. Then $\Phi^G X \simeq \Phi^{G/N} X^N$ for any $X \in \mathbf{Sp}^G$.*

Proof If $N \trianglelefteq G$ is a bottleneck subgroup, then $q^* \widetilde{E\mathcal{P}}_{G/N} \simeq \widetilde{E\mathcal{P}}_G$. Thus

$$\Phi^G X = (\widetilde{E\mathcal{P}}_G \wedge X)^G \simeq ((q^* \widetilde{E\mathcal{P}}_{G/N} \wedge X)^N)^{G/N}.$$

By the projection formula (Proposition 3.3), this is equivalent to

$$(\widetilde{E\mathcal{P}}_{G/N} \wedge X^N)^{G/N} = \Phi^{G/N} X^N.$$ □

Proposition 3.8 also follows from the more general [15, Proposition 9].

3.4 Inflation for $\underline{\mathbb{Z}}$-modules

Given a surjection $q\colon G \to G/N$, the inflation functor

$$\phi_N^*\colon \mathbf{Mack}(G/N) \to \mathbf{Mack}(G)$$

does not send $\underline{\mathbb{Z}}$-modules for G/N to $\underline{\mathbb{Z}}$-modules for G. We now describe a modified inflation functor that exists at the level of $\underline{\mathbb{Z}}$-modules. This functor previously appeared in [26, Section 3.2] and [3, Section 3.10].

Definition 3.9 Let $\mathcal{B}\underline{\mathbb{Z}}_G \subset \mathrm{Mod}_{\mathbb{Z}[G]}$ denote the full subcategory of permutation G-modules. Recall [26, Proposition 2.15] that $\underline{\mathbb{Z}}_G$-modules correspond to additive functors $\mathcal{B}\underline{\mathbb{Z}}_G^{op} \to \mathrm{Ab}$. Then the $\underline{\mathbb{Z}}$-module inflation functor

$$\Psi_N^* \colon \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}} \to \mathrm{Mod}_{\underline{\mathbb{Z}}_G}$$

is defined to be the left Kan extension along the inflation functor $\mathcal{B}\underline{\mathbb{Z}}_{G/N} \to \mathcal{B}\underline{\mathbb{Z}}_G$.

The following is an immediate corollary of the definition as a left Kan extension.

Proposition 3.10 *The functor Ψ_N^* is left adjoint to the functor $q_* \colon \mathrm{Mod}_{\underline{\mathbb{Z}}_G} \to \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$, defined as in (3.1).*

Proposition 3.11 [3, (3.11)] *For $\underline{M} \in \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$, the $\underline{\mathbb{Z}}_G$-module $\Psi_N^*(\underline{M})$ satisfies*

(1) *$q_*(\Psi_N^*(\underline{M}))$ is $\underline{M}$, and*
(2) *$\downarrow_N^G (\Psi_N^*(\underline{M}))$ is the constant Mackey functor at $\underline{M}(e)$.*

Note that Proposition 3.11 completely describes $\Psi_N^*(\underline{M})$ if N is a bottleneck subgroup. The following result states that $\underline{\mathbb{Z}}$-module inflation agrees with ordinary inflation on geometric Mackey functors.

Proposition 3.12 *Let $\underline{M} \in \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$, and let $N \trianglelefteq G$ be a bottleneck subgroup. If $\underline{M}(e) = 0$, then $\Psi_N^* \underline{M} \cong \phi_N^* \underline{M}$.*

Proof This follows immediately from Proposition 3.11. □

Remark 3.13 Note that Proposition 3.12 is not true without the bottleneck hypothesis. For instance, if $N = C_3 \trianglelefteq \Sigma_3$, then $\downarrow_{C_2}^{\Sigma_3} (\Psi_{C_3}^* \underline{M}) \cong \underline{M}$. In particular, it is not true that $\Psi_{C_3}^* \underline{M}$ is concentrated over $N = C_3$ when $\underline{M}(e) = 0$.

We now discuss the extension to equivariant spectra.

Proposition 3.14 *The N-fixed points functor*

$$(-)^N \colon \mathrm{Mod}_{H_G\underline{\mathbb{Z}}} \to \mathrm{Mod}_{H_{G/N}\underline{\mathbb{Z}}}$$

for $H\underline{\mathbb{Z}}$-modules has a left adjoint

$$\Psi_N^* \colon \mathrm{Mod}_{H_{G/N}\underline{\mathbb{Z}}} \to \mathrm{Mod}_{H_G\underline{\mathbb{Z}}}\,.$$

If $N \trianglelefteq G$ is a bottleneck subgroup, then the spectrum-level functor Ψ_N^* extends the functor Ψ_N^* of Definition 3.9, in the sense that

$$\Psi_N^* H_{G/N} \underline{M} \simeq H_G(\Psi_N^* \underline{M}) \tag{3.15}$$

for $\underline{M}$ in $\mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$.

Proof For an $H_{G/N}\underline{\mathbb{Z}}$-module X, the inflation q^*X is canonically a module over $q^*H_{G/N}\underline{\mathbb{Z}}$. We then define the spectrum-level functor Ψ^*_N by the formula

$$\Psi^*_N X = H\underline{\mathbb{Z}} \wedge_{q^*H\underline{\mathbb{Z}}} (q^*X).$$

We leave it to the reader to verify that this is indeed left adjoint to the N-fixed points functor.

To see that (3.15) holds, we show first that this holds on the indecomposable projective $\underline{\mathbb{Z}}_{G/N}$-modules. These are of the form $\uparrow^{G/N}_{K/N} \underline{\mathbb{Z}}$, and the diagram of commuting adjoint functors

$$\begin{array}{ccc} \mathrm{Mod}_{H_{G/N}\underline{\mathbb{Z}}} & \underset{(-)^N}{\overset{\Psi^*_N}{\rightleftarrows}} & \mathrm{Mod}_{H_G\underline{\mathbb{Z}}} \\ \uparrow^{G/N}_{K/N} \big\uparrow \big\downarrow \downarrow^{G/N}_{K/N} & & \uparrow^{G}_{K} \big\uparrow \big\downarrow \downarrow^{G}_{K} \\ \mathrm{Mod}_{H_{K/N}\underline{\mathbb{Z}}} & \underset{(-)^N}{\overset{\Psi^*_N}{\rightleftarrows}} & \mathrm{Mod}_{H_K\underline{\mathbb{Z}}} \end{array}$$

shows that

$$\Psi^*_N\big(H_{G/N} \uparrow^{G/N}_{K/N} \underline{\mathbb{Z}}\big) \simeq \uparrow^G_K \Psi^*_N(H_{K/N}\underline{\mathbb{Z}}) \simeq \uparrow^G_K H_K\underline{\mathbb{Z}} \simeq H_G \uparrow^G_K \underline{\mathbb{Z}} \simeq H_G\Psi^*_N\big(\uparrow^{G/N}_{K/N} \underline{\mathbb{Z}}\big).$$

Since the functor $\Psi^*_N\colon \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}} \to \mathrm{Mod}_{\underline{\mathbb{Z}}_G}$ is exact [26, Lemma 3.14], it follows that if $\mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$ has finite global projective dimension, then (3.15) will hold for any $\underline{\mathbb{Z}}_{G/N}$-module $\underline{M}$. By [4, Theorem 1.7], this is the case precisely when G/N is as described in Remark 3.7. □

Example 3.16 Let $X \in \mathbf{Sp}^{G/N}$ and $\underline{M} \in \mathbf{Mack}(G/N)$, with $\underline{M}(e) = 0$. Again assume that N is a bottleneck subgroup. Then Propositions 3.12 and 3.14 give that

$$\Psi^*_N(X \wedge H_{G/N}\underline{M}) \simeq q^*(X) \wedge \Psi^*_N(H_{G/N}\underline{M}) \simeq q^*(X) \wedge \phi^*_N H_{G/N}\underline{M} \simeq \phi^*_N(X \wedge H_{G/N}\underline{M}).$$

We will employ this equivalence when X is a representation sphere.

Proposition 3.17 *Let $N \trianglelefteq G$ be a bottleneck subgroup. Then for any G/N-representation V and $\underline{\mathbb{Z}}_{G/N}$-module $\underline{L}$, we have*

$$\underline{\pi}_n(\Psi^*_N \Sigma^V H_{G/N}\underline{L}) \cong \Psi^*_N \underline{\pi}_n(\Sigma^V H_{G/N}\underline{L}).$$

Proof Let us write $X = \Psi^*_N \Sigma^V H_{G/N}\underline{L} \simeq \Sigma^{q^*V} H_G \Psi^*_N \underline{L}$. Since N is a bottleneck subgroup, it is enough to describe $\downarrow^G_N \underline{\pi}_n X$ and $q_*\underline{\pi}_n X$. Now

$$\downarrow^G_N \underline{\pi}_n X \cong \underline{\pi}_n \downarrow^G_N X = \underline{\pi}_n \Sigma^{\dim V} H_N \underline{L}(N/N).$$

This is a constant Mackey functor. On the other hand, by (3.2) and (3.4), we have

$$q_*\underline{\pi}_n X \cong \underline{\pi}_n(X^N) \cong \underline{\pi}_n(\Sigma^V H_{G/N}\underline{L}).$$

By Proposition 3.11, this agrees with $\Psi^*_N \underline{\pi}_n(\Sigma^V H_{G/N}\underline{L})$. □

More generally, we have an extension of Proposition 3.12 to $H\underline{\mathbb{Z}}$-modules:

Proposition 3.18 *Let $X \in \mathrm{Mod}_{H\underline{\mathbb{Z}}_{G/N}}$ and let $N \trianglelefteq G$ be a bottleneck subgroup. If the underlying spectrum $\downarrow_e^{G/N} X$ is contractible, then $\Psi_N^*(X) \simeq \phi_N^* X$.*

Proof If the underlying spectrum of X is contractible, then $X \simeq \widetilde{E}(G/N) \wedge X$. The assumption that N is a bottleneck subgroup implies that $E(G/N) = q^*(E(G/N))$ is the universal space for the family of subgroups of N, so $\widetilde{E}(G/N) \wedge \widetilde{E}\mathcal{F}[N] \simeq \widetilde{E}(G/N)$, and it follows that

$$q^* X \simeq \widetilde{E}(G/N) \wedge q^* X \simeq \widetilde{E}(G/N) \wedge \phi_N^*(X) \simeq \phi_N^* X.$$

Now

$$\Psi_N^*(X) = H_G\underline{\mathbb{Z}} \wedge_{q^* H_{G/N}\underline{\mathbb{Z}}} q^*(X) \simeq H_G\underline{\mathbb{Z}} \wedge_{q^* H_{G/N}\underline{\mathbb{Z}}} (\widetilde{E}(G/N) \wedge q^*(X)).$$

Since $\widetilde{E}(G/N)$ is smash idempotent, this can be rewritten as

$$\Psi_N^*(X) \simeq \widetilde{E}(G/N) \wedge H_G\underline{\mathbb{Z}} \wedge_{\widetilde{E}(G/N) \wedge q^* H_{G/N}\underline{\mathbb{Z}}} \widetilde{E}(G/N) \wedge q^*(X).$$

It remains only to show that

$$\widetilde{E}(G/N) \wedge H_G\underline{\mathbb{Z}} \simeq \widetilde{E}(G/N) \wedge q^* H_{G/N}\underline{\mathbb{Z}}.$$

Both sides restrict trivially to an N-equivariant spectrum, so it suffices to show an equivalence on Φ^H, where H properly contains N. Without loss of generality, we may suppose that $H = G$. Since $\Phi^G(\widetilde{E}(G/N)) \simeq S^0$, it suffices to show that

$$\Phi^G H_G\underline{\mathbb{Z}} \simeq \Phi^G q^* H_{G/N}\underline{\mathbb{Z}}.$$

According to Proposition 3.8, the left side is $\Phi^{G/N} H_{G/N}\underline{\mathbb{Z}}$. Similarly, Proposition 3.8 and the projection formula (Proposition 3.3) show that the right side is

$$\Phi^G q^* H_{G/N}\underline{\mathbb{Z}} \simeq \Phi^{G/N}\big(H_{G/N}\underline{\mathbb{Z}} \wedge (S^0_G)^N\big) \simeq \Phi^{G/N} H_{G/N}\underline{\mathbb{Z}} \wedge \Phi^{G/N}(S^0_G)^N \simeq \Phi^{G/N} H_{G/N}\underline{\mathbb{Z}}. \quad \square$$

We now establish our main result, which partially describes the slice filtration of inflated $H\underline{\mathbb{Z}}$-modules. Recall that $P^t_n X = \mathrm{fib}\, P^t X \to P^{n-1} X$ denotes the slice approximation of X that contains the slices of X in dimensions n to t.

Theorem 3.19 *Let $n \geq 0$ and let $N \trianglelefteq G$ be a bottleneck subgroup of order p, a prime. Let $\underline{M} \in \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$ such that $P^n_n \Sigma^n H_{G/N}\underline{M}$ is of the form $\Sigma^V H_{G/N}\underline{L}$, for some G/N-representation V and $\underline{L} \in \mathrm{Mod}_{\underline{\mathbb{Z}}_{G/N}}$. Then the nontrivial slices of the Eilenberg–Mac Lane G-spectrum $\Sigma^n H_G(\Psi_N^*\underline{M})$, above level pn, are*

$$P^{pk}_{pk}\big(\Sigma^n H_G(\Psi_N^*\underline{M})\big) \simeq \Psi_N^* P^k_k(\Sigma^n H_{G/N}\underline{M}) \simeq \phi_N^* P^k_k(\Sigma^n H_{G/N}\underline{M})$$

for $k > n$. Furthermore,

$$P^{pk}_{n}\big(\Sigma^n H_G(\Psi_N^*\underline{M})\big) \simeq \Psi_N^* P^k_n(\Sigma^n H_{G/N}\underline{M})$$

for $k \geq n$.

Proof Applying the functor Ψ_N^* to the slice tower for $\Sigma^n H_{G/N}\underline{M}$ produces a tower of fibrations whose layers are $\Psi_N^* P^k_k(\Sigma^n H_{G/N}\underline{M})$ for $k \geq n$. We wish to show that this is a partial slice tower for

$\Sigma^n H_G(\Psi_N^* \underline{M})$. For $k > n$, the k-slice $P_k^k(\Sigma^n H_{G/N}\underline{M})$ has trivial underlying spectrum. It follows from Proposition 3.18 that

$$\Psi_N^* P_k^k(\Sigma^n H_{G/N}\underline{M}) \simeq \phi_N^* P_k^k(\Sigma^n H_{G/N}\underline{M})$$

for $k > n$. As the geometric inflation of a k-slice, this is a pk-slice.

It remains to show that

$$\Psi_N^* P_n^n(\Sigma^n H_{G/N}\underline{M}) \simeq \Psi_N^* \Sigma^V H_{G/N}\underline{L} \simeq \Sigma^V H_G \Psi_N^* \underline{L}$$

has no slices above level pn. First, note that the restriction of $\Sigma^V H_G \Psi_N^* \underline{L}$ to N is the N-spectrum $\Sigma^n H_N \underline{L(N)}$, where $\underline{L(N)}$ is being considered as a constant N-Mackey functor at the value $\underline{L}(G/N)$. It follows that this N-spectrum has no slices above dimension $|N| \cdot n = pn$. Therefore, to show that $\Sigma^V H_G \Psi_N^* \underline{L}$ is less than or equal to pn, it suffices to show that

$$[G_+ \wedge_H S^{k\rho_H + r}, \Sigma^V H_G \Psi_N^* \underline{L}]^G = 0$$

for any $N < H \leq G$ and integers $r \geq 0$ and k such that $k|H| > pn$. Without loss of generality we consider the case $H = G$.

Denote by U a complement of $\rho_{G/N}$ in ρ_G, so that

$$\rho_G \cong \rho_{G/N} \oplus U.$$

We then have a cofiber sequence

$$S(kU)_+ \wedge S^{k\rho_{G/N}} \to S^{k\rho_{G/N}} \to S^{k\rho_G}$$

and a resulting exact sequence

$$\begin{aligned}[\Sigma^1 S(kU)_+ \wedge S^{k\rho_{G/N}+r}, \Sigma^V H_G \Psi_N^* \underline{L}]^G &\to [S^{k\rho_G + r}, \Sigma^V H_G \Psi_N^* \underline{L}]^G \\ &\to [S^{k\rho_{G/N}+r}, \Sigma^V H_G \Psi_N^* \underline{L}]^G = 0.\end{aligned}$$

We must show that the left term vanishes. Note that the G-action on $S(kU)$ is free, since N is order p. Then the desired vanishing follows from the fact that $\Sigma^1 S(kU)_+ \wedge S^{k\rho_{G/N} - V}$ is G-connected, since $\dim k\rho_{G/N} > \dim V = n$. □

Remark 3.20 While we have chosen to state Theorem 3.19 in the case of $n \geq 0$, an analogous result holds for negative suspensions. There, we have

$$P_{-pk}^{-n}\left(\Sigma^{-n} H_G(\Psi_N^* \underline{M})\right) \simeq \Psi_N^* P_{-k}^{-n}(\Sigma^{-n} H_{G/N}\underline{M})$$

for $k \geq n > 0$. That $\Psi_N^* P_{-n}^{-n}(\Sigma^{-n} H_{G/N}\underline{M}) \simeq \Sigma^{-V} H_G \Psi_N^* \underline{L}$ is at least $-pn$ can be seen from Hill and Yarnall's characterization of slice connectivity in terms of geometric fixed points [13, Theorem 2.5].

4 Q_8-Mackey functors and Bredon homology

We display a number of the Q_8-Mackey functors that will be relevant in Table 3. In these Lewis diagrams, we are using the subgroup lattice of Q_8 as displayed in Section 1.1. We will also often abuse notation and

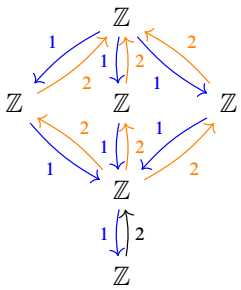
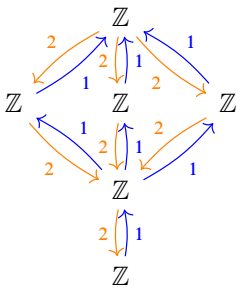
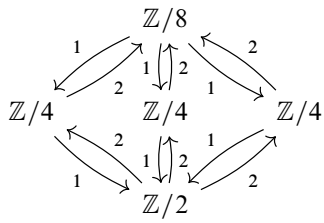
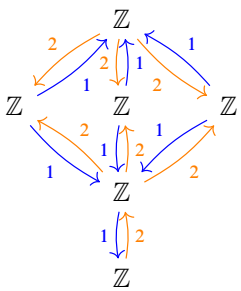
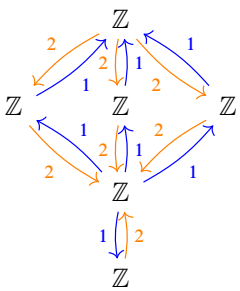
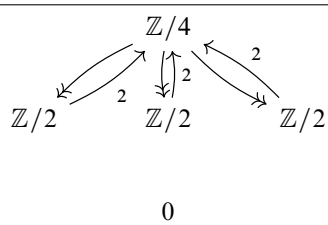
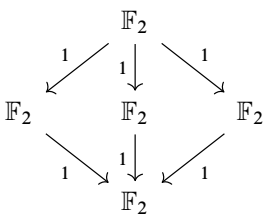
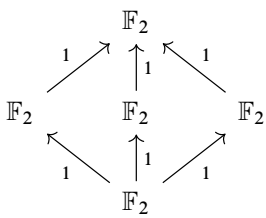
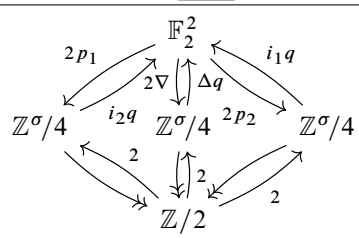
$\square = \underline{\mathbb{Z}}$	$\circledast = \underline{\mathbb{Z}}^*$	$\circ = \underline{B}(3,0)$
		
$\underline{\mathbb{Z}}(3,2) = \Psi_Z^* \underline{\mathbb{Z}}(2,1)$	$\underline{\mathbb{Z}}(3,1) = \Psi_Z^* \underline{\mathbb{Z}}^*$	$\varsigma = \phi_Z^*(\underline{B}(2,0))$
		
$\blacklozenge = \phi_Z^* \underline{\mathbb{F}_2}$	$\circledcirc = \phi_Z^* \underline{\mathbb{F}_2}^*$	$\pentagon = \underline{mgw}$
		

Table 3: Some Q_8-Mackey functors.

write the name for a K_4-Mackey functor, such as $\underline{m}$ or $\underline{mg}$, to denote the resulting inflated Q_8-Mackey functor. We will only write the symbol ϕ_Z^* when it is necessary to resolve an ambiguity, for instance between $\phi_Z^* \underline{\mathbb{F}_2}$ and $\underline{\mathbb{F}_2}$.

In [10, Section 2.1], the authors introduce "forms of $\underline{\mathbb{Z}}$" Mackey functors $\underline{\mathbb{Z}}(i,j)$, where $i \geq j \geq 0$, in the case of $G = C_{p^n}$. From our point of view, Q_8 behaves very similarly to C_8, and we similarly write $\underline{\mathbb{Z}}(i,j)$ for the Mackey functor that looks like $\underline{\mathbb{Z}}^*$ between the subgroups of order 2^i and 2^j and looks like $\underline{\mathbb{Z}}$ outside of this range. We will at times follow [10] in denoting by $\underline{B}(i,j)$ the cokernel of $\underline{\mathbb{Z}}(i,j) \hookrightarrow \underline{\mathbb{Z}}$, although we will often instead use the descriptions given in Proposition 4.1.

These Mackey functors fit together in exact sequences as follows, using notation from Table 3:

Proposition 4.1 *There are exact sequences of Mackey functors*

(1) $\underline{\mathbb{Z}}(3,2) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \underline{g}$,

(2) $\underline{\mathbb{Z}}(3,1) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \phi_Z^* \underline{B}(2,0)$,

(3) $\underline{\mathbb{Z}}(3,1) \hookrightarrow \underline{\mathbb{Z}}(3,2) \twoheadrightarrow \underline{m}^*$,

(4) $\underline{\mathbb{Z}}(2,1) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \underline{m}$,

(5) $\underline{\mathbb{Z}}(1,0) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \phi_Z^* \underline{\mathbb{F}_2}$,

(6) $\underline{\mathbb{Z}}^* \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \underline{B}(3,0)$,

(7) $\underline{mg} \hookrightarrow \underline{mgw} \twoheadrightarrow \underline{w}$.

4.1 $RO(Q_8)$-graded Mackey functor $\underline{\mathbb{Z}}$-homology of a point

We will now compute the homology of $S^{k\rho_Q}$, with coefficients in $\underline{\mathbb{Z}}$, as a Mackey functor. The starting point is that the regular representation of Q splits as

$$\rho_Q \cong \mathbb{H} \oplus \rho_K,$$

where $\mathbb{H}$ is the 4-dimensional irreducible Q-representation given by the action of the unit quaternions on the algebra of quaternions and ρ_K is the regular representation of K, inflated to Q along the quotient. We begin by computing the homology of $S^{k\mathbb{H}}$. See also [17, Section 2] for an alternative viewpoint.

First, Proposition 3.3 and [19, Proposition 9.1] combine to yield the following.

Proposition 4.2 *For $k \geq 0$, the nontrivial homotopy Mackey functors of $\Sigma^{k\rho_K} H_Q\underline{\mathbb{Z}}$ are*

$$\underline{\pi}_n(\Sigma^{k\rho_K} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 4k, \\ \underline{mg} & \text{if } n = 4k-2, \\ \underline{g}^{\frac{1}{2}(4k-n-1)} & \text{if } n \in [2k, 4k-3] \text{ and } n \text{ is odd}, \\ \underline{g}^{\frac{1}{2}(4k-n-4)} \oplus \phi_{LDR}^* \underline{\mathbb{F}_2} & \text{if } n \in [2k, 4k-3] \text{ and } n \text{ is even}, \\ \underline{g}^{n-k+1} & \text{if } n \in [k, 2k-1]. \end{cases}$$

Next, we employ the cofiber sequence

$$S(\mathbb{H})_+ \to S^0 \to S^{\mathbb{H}} \tag{4.3}$$

to obtain the homology of S^{ρ_Q} from that of S^{ρ_K}.

Proposition 4.4 *The nontrivial homotopy Mackey functors of $S(\mathbb{H})_+ \wedge H_Q\underline{\mathbb{Z}}$ are*

$$\underline{\pi}_n(S(\mathbb{H})_+ \wedge H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 3, \\ \underline{mgw} & \text{if } n = 1, \\ \underline{\mathbb{Z}}^* & \text{if } n = 0. \end{cases}$$

Proof Since the action of Q on $S(\mathbb{H})$ is free, we can write down an equivariant cell structure using only free cells. Viewing $S(\mathbb{H})$ as the one-point compactification of $\mathbb{R}^3$, there is a straight-forward cell structure

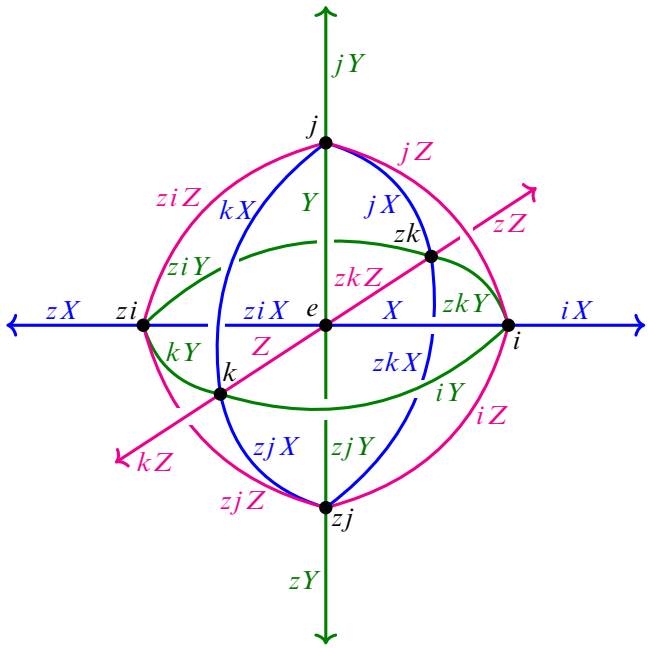


Figure 3: The 1-skeleton of $S(\mathbb{H})$.

in which the subgroups L, D, and R act freely on the x-, y-, and z-axes, respectively. We display the 1-skeleton in Figure 3, and the cell structure is described by the following complex of $\mathbb{Z}[Q]$-modules:

$$\mathbb{Z}[Q]^2 \xrightarrow{\begin{pmatrix} e & j \\ -e & -i \\ e & k \\ -e & -e \end{pmatrix}} \mathbb{Z}[Q]^4 \xrightarrow{\begin{pmatrix} k & e & e & k \\ -e & -e & i & i \\ e & -j & -e & j \end{pmatrix}} \mathbb{Z}[Q]^3 \xrightarrow{(i-e\ \ j-e\ \ k-e)} \mathbb{Z}[Q].$$

This yields an associated complex of induced Mackey functors

$$\underline{\mathbb{Z}[Q]}^2 \to \underline{\mathbb{Z}[Q]}^4 \to \underline{\mathbb{Z}[Q]}^3 \to \underline{\mathbb{Z}[Q]}$$

leading to the claimed homology Mackey functors. □

Remark 4.5 A smaller chain complex for computing the homology of $S(\mathbb{H})$ is given by

$$\mathbb{Z}[Q] \xrightarrow{\binom{i-e}{e-k}} \mathbb{Z}[Q]^2 \xrightarrow{\left(\begin{smallmatrix} e+i & e+k \\ -e-j & -e+i \end{smallmatrix}\right)} \mathbb{Z}[Q]^2 \xrightarrow{(i-e\ \ j-e)} \mathbb{Z}[Q].$$

We gave a less efficient chain complex in the proof of Proposition 4.4 for geometric reasons.

Using (4.3), this immediately yields the following.

Corollary 4.6 *The nontrivial homotopy Mackey functors of* $\Sigma^{\mathbb{H}} H_Q\underline{\mathbb{Z}}$ *are*

$$\underline{\pi}_n(\Sigma^{\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 4, \\ \underline{mgw} & \text{if } n = 2, \\ \underline{B}(3,0) & \text{if } n = 0. \end{cases}$$

We will use this to compute the homology of S^{ρ_Q}, using the following periodicity result.

Proposition 4.7 [23, Proposition 4.1] *For any orientable representation V of dimension d and free Q-space X, the orientation $u_V \in \mathrm{H}_d(S^V;\mathbb{Z})$ induces an equivalence*

$$\Sigma^d X_+ \wedge H_Q\underline{\mathbb{Z}} \simeq \Sigma^V X_+ \wedge H_Q\underline{\mathbb{Z}}.$$

We now compute the homology of S^{ρ_Q}.

Proposition 4.8 *The nontrivial homotopy Mackey functors of $\Sigma^{\rho_Q} H_Q\underline{\mathbb{Z}}$ are*

$$\underline{\pi}_n(\Sigma^{\rho_Q} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 8, \\ \underline{mgw} & \text{if } n = 6, \\ \underline{B}(3,0) & \text{if } n = 4, \\ \underline{mg} & \text{if } n = 2, \\ \underline{g} & \text{if } n = 1. \end{cases}$$

Proof The representation ρ_K is orientable. For example, using the basis $\{1, i, j, k\}$ for $\rho_K = \mathbb{R}[K]$, the matrix $\rho_K(i)$ is given by

$$\rho_K(i) = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

which has determinant equal to 1. By Proposition 4.7, we have

$$\underline{\pi}_n\big(S(\mathbb{H})_+ \wedge \Sigma^{\rho_K} H_Q\underline{\mathbb{Z}}\big) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 7, \\ \underline{mgw} & \text{if } n = 5, \\ \underline{\mathbb{Z}}^* & \text{if } n = 4. \end{cases}$$

The result then follows from the cofiber sequence

$$S(\mathbb{H})_+ \wedge \Sigma^{\rho_K} H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_K} H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_Q} H_Q\underline{\mathbb{Z}}. \qquad \square$$

Corollary 4.6 generalizes as follows.

Proposition 4.9 *The nontrivial homotopy Mackey functors of $\Sigma^{k\mathbb{H}} H_Q\underline{\mathbb{Z}}$, for $k > 0$ are*

$$\underline{\pi}_n(\Sigma^{k\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 4k, \\ \underline{mgw} & \text{if } 0 < n < 4k \text{ and } n \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } 0 \le n < 4k \text{ and } n \equiv 0 \pmod 4. \end{cases}$$

Proof This follows by induction, using the cofiber sequence

$$S(\mathbb{H})_+ \wedge S^{(k-1)\mathbb{H}} \to S^{(k-1)\mathbb{H}} \to S^{k\mathbb{H}}$$

and Proposition 4.7. The latter applies since $\mathbb{H}$, and therefore also $(k-1)\mathbb{H}$, is orientable. $\square$

Combining this with the cofiber sequence

$$S(k\mathbb{H})_+ \wedge \Sigma^{k\rho_K} H_Q\underline{\mathbb{Z}} \to \Sigma^{k\rho_K} H_Q\underline{\mathbb{Z}} \to \Sigma^{k\rho_Q} H_Q\underline{\mathbb{Z}}$$

and Proposition 4.7 gives the following result.

Proposition 4.10 *The nontrivial homotopy Mackey functors of $\Sigma^{k\rho_Q} H_Q\underline{\mathbb{Z}}$, for $k > 0$, are*

$$\underline{\pi}_n(\Sigma^{k\rho_Q} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 8k, \\ \underline{mgw} & \text{if } 4k < n < 8k \text{ and } n \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } 4k \le n < 8k \text{ and } n \equiv 0 \pmod 4, \\ \phi_Z^*\underline{\pi}_n(\Sigma^{k\rho_K} H_K\underline{\mathbb{Z}}) & \text{if } n < 4k, \end{cases}$$

where the latter Mackey functors are listed in Proposition 4.2.

The homotopy Mackey functors of $\Sigma^{k\rho_Q} H_Q\underline{\mathbb{Z}}$ are displayed in Figure 4. When k is negative, the computation follows the same strategy. The initial input, which can again be computed using the chain complex given in Proposition 4.4, is that

$$\underline{\mathrm{H}}^n(S(\mathbb{H});\underline{\mathbb{Z}}) \cong \underline{\pi}_{-n}\big(F(S(\mathbb{H})_+, H_Q\underline{\mathbb{Z}})\big) \cong \begin{cases} \underline{\mathbb{Z}}^* & \text{if } n = 3, \\ \underline{mgw} & \text{if } n = 2, \\ \underline{\mathbb{Z}} & \text{if } n = 0. \end{cases} \tag{4.11}$$

Using this and [19, Proposition 9.2] leads to the following answer.

Proposition 4.12 *The nontrivial homotopy Mackey functors of $\Sigma^{-k\rho_Q} H_Q\underline{\mathbb{Z}}$, for $k > 0$, are*

$$\underline{\pi}_{-n}(\Sigma^{-k\rho_Q} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}}^* & \text{if } n = 8k, \\ \underline{mgw} & \text{if } n \in [4k, 8k] \text{ and } n \equiv 3 \pmod 4, \\ \underline{B}(3,0) & \text{if } n \in [4k+5, 8k] \text{ and } n \equiv 1 \pmod 4, \\ \phi_Z^*\underline{B}(2,0) & \text{if } n = 4k+1, \\ \underline{mg}^* & \text{if } n = 4k-1, \\ \underline{g}^{\frac{4k-n}{2}} & \text{if } n \in [2k+4, 4k-2] \text{ and } n \equiv 0 \pmod 2, \\ \underline{g}^{\frac{4k-n-3}{2}} \oplus \phi_{LDR}^*\underline{\mathbb{F}_2}^* & \text{if } n \in [2k+3, 4k-2] \text{ and } n \equiv 1 \pmod 2, \\ \underline{g}^{n-k-3} & \text{if } n \in [k+4, 2k+2]. \end{cases}$$

Remark 4.13 The "gap theorem" [9, Proposition 3.20] predicts that the groups $\pi_n^Q\Sigma^{-k\rho} H_Q\underline{\mathbb{Z}}$ vanish for $k \ge 0$ and $n \in [-3,-1]$, as indicated in Figure 4. Actually, for $k \ge 2$ the argument there proves more. It tells us that for $k \ge 2$, the cohomology groups $\mathrm{H}^n_Q(S^{k\rho};\underline{M})$ vanish for positive $n \le k+1$. This is equivalent to saying that $\pi_{-n}^Q\Sigma^{-k\rho} H_Q\underline{M}$ vanishes, with the same conditions on k and n.

4.2 Additional homology calculations

We will also need the following auxiliary calculations in Section 6.

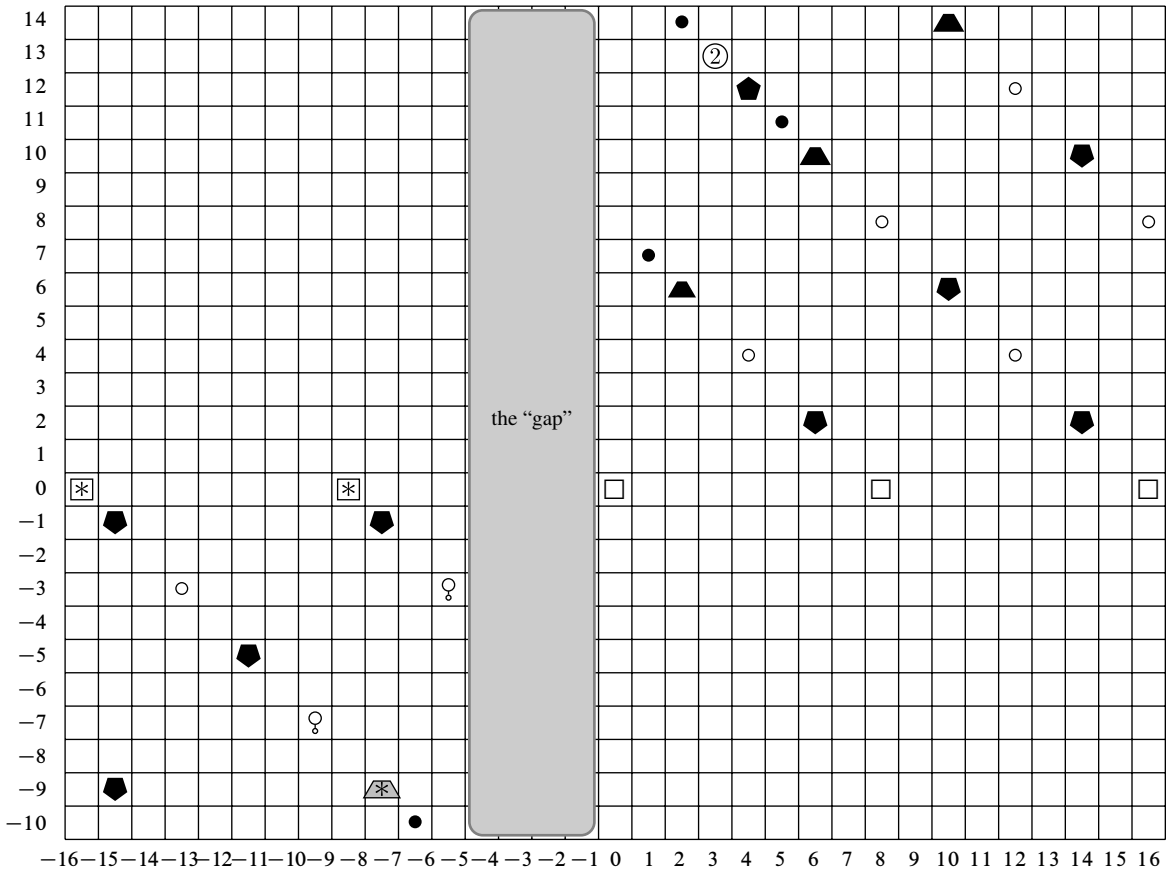


Figure 4: The homotopy Mackey functors of $\bigvee_n \Sigma^{n\rho} H_Q\underline{\mathbb{Z}}$. The Mackey functor $\underline{\pi}_k \Sigma^{n\rho} H_Q\underline{\mathbb{Z}}$ appears in position $(k, 8n-k)$.

Proposition 4.14 *The nontrivial homotopy Mackey functors of* $\Sigma^{\rho_K - \mathbb{H}} H_Q\underline{\mathbb{Z}}$ *are*

$$\underline{\pi}_n(\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \phi_Z^*\underline{\mathbb{F}_2} & \text{if } n = 1, \\ \underline{\mathbb{Z}}^* & \text{if } n = 0. \end{cases}$$

Proof The fiber sequence

$$\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_K} H_Q\underline{\mathbb{Z}} \to F(S(\mathbb{H})_+, \Sigma^{\rho_K} H_Q\underline{\mathbb{Z}}) \simeq \Sigma^4 F(S(\mathbb{H})_+, H_Q\underline{\mathbb{Z}})$$

yields an isomorphism $\underline{\pi}_0(\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \underline{\mathbb{Z}}^*$ and shows that the homotopy vanishes for n outside of $[0, 2]$. Given that the restriction to any C_4, which is the C_4-spectrum $\Sigma^{2+2\sigma-2\lambda} H_{C_4}\underline{\mathbb{Z}}$, has a trivial $\underline{\pi}_2$ [26, Theorem 6.10], the long exact sequence further shows that $\underline{\pi}_2$ vanishes as well, and it implies that we have an extension

$$\underline{w} \hookrightarrow \underline{\pi}_1(\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}) \twoheadrightarrow \underline{g}.$$

It remains to show this is not the split extension. The fiber sequence

$$\uparrow_D^Q \Sigma^{1+2\sigma-2\lambda} H_{C_4}\underline{\mathbb{Z}} \to \Sigma^{1+p_1^*\sigma+p_2^*\sigma-\mathbb{H}} H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}$$

shows that $\underline{\pi}_1(\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}})$ injects into

$$\underline{\pi}_0\big(\uparrow_D^Q \Sigma^{1+2\sigma-2\lambda}H_{C_4}\underline{\mathbb{Z}}\big) \cong \uparrow_D^Q \phi_{C_2}^*\underline{\mathbb{F}_2}.$$

It follows that $\underline{\pi}_1(\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}) \cong \phi_Z^*\underline{\mathbb{F}_2}$. □

Proposition 4.15 *The nontrivial homotopy Mackey functors of* $\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}(3,2)$ *are*

$$\underline{\pi}_n\big(\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}(3,2)\big) \cong \begin{cases} \underline{w} & \text{if } n=1,\\ \underline{\mathbb{Z}}^* & \text{if } n=0.\end{cases}$$

Proof The short exact sequence

$$\underline{\mathbb{Z}}(3,2) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \underline{g}$$

gives rise to a cofiber sequence

$$\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}(3,2) \to \Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_K-\mathbb{H}}H_Q\underline{g} \simeq \Sigma^1 H_Q\underline{g}.$$

Using a naturality square, the second map factors as

$$\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}} \to \Sigma^{\rho_K}H_Q\underline{\mathbb{Z}} \to \Sigma^1 H_Q\underline{g},$$

where the first map is an epimorphism on $\underline{\pi}_1$ by the proof of Proposition 4.14 and the second is an isomorphism on $\underline{\pi}_1$. The conclusion follows. □

Proposition 4.16 *The nontrivial homotopy Mackey functors of* $\Sigma^{\mathbb{H}-\rho_K}H_Q\underline{\mathbb{Z}}(2,0)$ *are*

$$\underline{\pi}_n\big(\Sigma^{\mathbb{H}-\rho_K}H_Q\underline{\mathbb{Z}}(2,0)\big) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n=0,\\ \underline{w}^* & \text{if } n=-2.\end{cases}$$

Proof This follows from Proposition 4.15 by duality. In more detail, Proposition 4.15 gives a fiber sequence

$$\Sigma^1 H_Q\underline{w} \to \Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}(3,2) \to H_Q\underline{\mathbb{Z}}^*.$$

Applying Anderson duality (see [19, Section 2.2]) gives a fiber sequence

$$I(\Sigma^1 H_Q\underline{w}) \leftarrow I\big(\Sigma^{\rho_K-\mathbb{H}}H_Q\underline{\mathbb{Z}}(3,2)\big) \leftarrow I(H_Q\underline{\mathbb{Z}}^*),$$

or in other words

$$\Sigma^{-1}I(H_Q\underline{w}) \leftarrow \Sigma^{\mathbb{H}-\rho_K}H_Q\underline{\mathbb{Z}}(2,0) \leftarrow H_Q\underline{\mathbb{Z}}.$$

But as the Mackey functor $\underline{w}$ is torsion, the Anderson dual is the desuspension of the Brown–Comenetz dual. In other words, $I(H_Q\underline{w}) \simeq \Sigma^{-1}I_{\mathbb{Q}/\mathbb{Z}}H_Q\underline{w} \simeq \Sigma^{-1}H_Q\underline{w}^*$. □

5 Review of the C_4-slices of $\Sigma^n H\mathbb{Z}$

In this section, we review the slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ from [24]. Note that the slices as listed in [24] are written using the classical slice filtration, whereas we use the regular slice filtration. The only difference is a suspension by one. The Mackey functors that appear here were introduced in Table 1.

According to [24, Section 4.2], the C_4-spectrum $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ is an n-slice for $0 \le n \le 4$. For $n \ge 5$, $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ has a nontrivial slice tower. Yarnall's method for determining these slice towers is to splice together suspensions of the cofiber sequences

$$\Sigma^{-1} H_{C_4}\underline{g} \to \Sigma^2 H_{C_4}\underline{\mathbb{Z}} \to \Sigma^{2\sigma} H_{C_4}\underline{\mathbb{Z}},$$
$$\Sigma^{-1} H_{C_4}\phi^*_{C_2}\underline{\mathbb{F}_2}^* \to \Sigma^2 H_{C_4}\underline{\mathbb{Z}} \to \Sigma^{\lambda} H_{C_4}\underline{\mathbb{Z}}(2,1),$$
$$\Sigma^{-1} H_{C_4}\underline{B}(2,0) \to \Sigma^2 H_{C_4}\underline{\mathbb{Z}} \to \Sigma^{\lambda} H_{C_4}\underline{\mathbb{Z}},$$

in combination with the equivalences

$$\Sigma^2 H_{C_4}\underline{\mathbb{Z}} \simeq \Sigma^{2\sigma} H_{C_4}\underline{\mathbb{Z}}(2,1),$$
$$\Sigma^{-1} H_{C_4}\phi^*_{C_2}\underline{\mathbb{F}_2}^* \simeq \Sigma^{-\sigma} H_{C_4}\phi^*_{C_2}\underline{f} \simeq \Sigma^{1-2\sigma} H_{C_4}\phi^*_{C_2}\underline{\mathbb{F}_2}.$$

We first review these slices for odd n.

Proposition 5.1 [24, Theorem 4.2.6] *Let $n \ge 5$ be odd. The bottom slice of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ is*

$$P^n_n(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{\frac{n-5}{4}\rho+4+\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 1 \pmod 8, \\ \Sigma^{\frac{n-3}{4}\rho+3} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 3 \pmod 8, \\ \Sigma^{\frac{n-5}{4}\rho+3+2\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 5 \pmod 8, \\ \Sigma^{\frac{n-3}{4}\rho+2+\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 7 \pmod 8. \end{cases}$$

Proposition 5.2 [24, Lemma 4.2.5] *Let $n \ge 5$ be odd. The nontrivial $4k$-slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ are*

$$P^{4k}_{4k}(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{k\rho} H_{C_4}\underline{B}(2,0) & \text{if } 4k \in [n+1, 2(n-3)] \text{ and } k \text{ is even}, \\ \Sigma^{k\rho} H_{C_4}\phi^*\underline{f} & \text{if } 4k \in [n+1, 2(n-3)] \text{ and } k \text{ is odd}, \\ \Sigma^{k\rho} H_{C_4}\underline{g} & \text{if } 4k \in [2(n-1), 4(n-3)] \text{ and } k \text{ is even}. \end{cases}$$

The $4k$-slices can also be read off of [11, Figure 3]. When n is odd, these are the only nontrivial slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$.

We now recall the slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ for even n.

Proposition 5.3 [24, Theorem 4.2.9] *Let $n \ge 6$ be even. The bottom slice of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ is*

$$P^n_n(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{\frac{n-4}{4}\rho+3+\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 0 \pmod 8, \\ \Sigma^{\frac{n-6}{4}\rho+3+3\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 2 \pmod 8, \\ \Sigma^{\frac{n-4}{4}\rho+4} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 4 \pmod 8, \\ \Sigma^{\frac{n-6}{4}\rho+4+2\sigma} H_{C_4}\underline{\mathbb{Z}} & \text{if } n \equiv 6 \pmod 8. \end{cases}$$

Proposition 5.4 [24, Lemma 4.2.7] *Let $n \ge 6$ be even. The nontrivial $4k$-slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ are*

$$P^{4k}_{4k}(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \Sigma^k H_{C_4}\underline{g}, \quad k \text{ odd}$$

for $4k$ in the range $[n+2, 4n-12]$.

Again, the $4k$-slices can also be read off of [11, Figure 3].

Proposition 5.5 [24, Theorem 4.2.9] *Let $n \geq 6$ be even. The $4k+2$-slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ are*

$$P^{8k+2}_{8k+2}(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \Sigma^{1+2k\rho} H\phi^*\underline{\mathbb{F}_2},$$
$$P^{8k+6}_{8k+6}(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \Sigma^{3+2k\rho} H\phi^*\underline{\mathbb{F}_2},$$

for $8k+2$ or $8k+6$ in the range $[n+2, 2n-6]$.

We may also view these slices through the perspective of the $\underline{\mathbb{Z}}$-module inflation functor. By Theorem 3.19,

$$\Psi^*_{C_2} : \mathrm{Mod}_{H_{C_2}\underline{\mathbb{Z}}} \to \mathrm{Mod}_{H_{C_4}\underline{\mathbb{Z}}}$$

will provide all slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ above level $2n$. Let $r \equiv n \pmod 4$ with $3 \leq r \leq 6$. It follows from [19, Proposition 3.5] that the slices of $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$ in level at least $2n+2r-4$ are

$$P^{4k}_{4k}(\Sigma^n H_{C_4}\underline{\mathbb{Z}}) \simeq \Psi^*_{C_2}\Sigma^k H_{C_2}\underline{g} \simeq \Sigma^k H_{C_4}\underline{g}$$

for $4k \in [2n+2r-4, 4(n-3)]$. The rest of the slices then follow from determining the slices of

$$\Psi^*_{C_2} P^n_n \Sigma^n H_{C_2}\underline{\mathbb{Z}} = \Psi^*_{C_2}\Sigma^{\frac{n-r}{2}\rho_{C_2}+r} H_{C_2}\underline{\mathbb{Z}} \simeq \Sigma^{\frac{n+r}{2}+\frac{n-r}{2}\sigma} H_{C_4}\underline{\mathbb{Z}}.$$

The slice tower for this C_4-spectrum can be found by splicing together the cofiber sequences listed at the start of this section.

6 Q_8-slices

The slices of $\Sigma^n H_K\underline{\mathbb{Z}}$ were determined by the second author in [19, Section 8]. As stated in Theorem 3.19, it follows that the $\underline{\mathbb{Z}}$-module inflation functor

$$\Psi^*_Z : \mathrm{Mod}_{H_K\underline{\mathbb{Z}}} \to \mathrm{Mod}_{H_Q\underline{\mathbb{Z}}}$$

of Proposition 3.14 will produce all slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$ in degree larger than $2n$, as the inflation of the slices of $\Sigma^n H_K\underline{\mathbb{Z}}$ above degree n.

The remaining slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$ will be given as the slices of $\Psi^*_Z\big(P^n_n(\Sigma^n H_K\underline{\mathbb{Z}})\big)$. By [19, Proposition 8.5], these are of the form

$$\Psi^*_Z(\Sigma^{r+j\rho_K} H_K\underline{\mathbb{Z}}) \simeq \Sigma^{r+j\rho_K} H_Q\underline{\mathbb{Z}}, \tag{6.1}$$

where $r \in \{3,4,5\}$, if $n \not\equiv 2 \pmod 4$. In the case $n \equiv 2 \pmod 4$, the same result states that this is

$$\Psi^*_Z\big(\Sigma^{2+j\rho_K} H_K\underline{\mathbb{Z}}(1,0)\big) \simeq \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1).$$

But the cofiber sequence (Proposition 4.1)

$$\Sigma^{1+j\rho_K} H_Q\underline{m} \to \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1) \to \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}} \tag{6.2}$$

reduces the computation of slices of $\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1)$ to the question of the slice tower for $\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}$, given that $\Sigma^{1+j\rho_K} H_Q\underline{m} \simeq \phi^*_Z(\Sigma^{1+j\rho_K} H_K\underline{m})$ is an $8j+4$-slice [19, Proposition 5.7]. We determine the slices of $\Sigma^{r+j\rho_K} H_Q\underline{\mathbb{Z}}$, for $r \in \{2,\dots,5\}$ in Section 6.1.

6.1 Slice towers for $\Sigma^{r+j\rho_K} H_Q\underline{\mathbb{Z}}$

The K_4-spectrum $\Sigma^{r+j\rho_K} H_K\underline{\mathbb{Z}}$ is an n-slice for $r \in \{2,\dots,5\}$ [19, Proposition 7.1]. However, the inflation of this to Q_8 is no longer a slice. We here determine the slice towers of these inflations. Throughout, we will implicitly use Proposition 6.7, which does not rely on the following material.

6.1.1 $r = 2$ First, we observe that $\Sigma^{2+\rho_K} H_Q\underline{\mathbb{Z}}$ is a 6-slice. To see this we first note that it restricts to a 6-slice at every proper subgroup by Proposition 5.3. It therefore remains only to show that it does not have any $8k$-slices for $k \geq 1$. This is equivalent to showing that $\underline{\pi}_{-2}(\Sigma^{\rho_K - k\rho_Q} H_Q\underline{\mathbb{Z}})$ vanishes for $k \geq 1$. In the case $k = 1$, (4.11) shows that $\Sigma^{-\mathbb{H}} H_Q\underline{\mathbb{Z}}$ is (-3)-truncated, in the sense that it has no homotopy Mackey functors above dimension -3. This remains true after further desuspending by copies of ρ_Q.

Next, the tower for $\Sigma^{2+2\rho_K} H_Q\underline{\mathbb{Z}}$ is given by

$$\begin{array}{ccc} P^{14}_{14} = \Sigma^{-1+2\rho_Q} H_Q\underline{w}^* & \longrightarrow & \Sigma^{2+2\rho_K} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ P^{12}_{12} = \Sigma^{1+\rho_Q} H_Q\underline{m} & \longrightarrow & \Sigma^{2+\rho_Q} H_Q\underline{\mathbb{Z}}(2,0) \\ & & \downarrow \\ & & P^{10}_{10} = \Sigma^{2+\rho_Q} H_Q\underline{\mathbb{Z}}(1,0) \end{array}$$

This uses the computation (see Proposition 4.16)

$$\underline{\pi}_n\big(\Sigma^{\mathbb{H}-\rho_K} H_Q\underline{\mathbb{Z}}(2,0)\big) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } n = 0, \\ \underline{w}^* & \text{if } n = -2, \end{cases}$$

to produce the first cofiber sequence.

Finally, for $j \geq 3$, the tower may be obtained by recursively using

$$\begin{array}{ccc} P^{8j-2}_{8j-2} = \Sigma^{-1+j\rho_Q} H_Q\underline{w}^* & \longrightarrow & \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ P^{8j-4}_{8j-4} = \Sigma^{1+(j-1)\rho_Q} H_Q\underline{m} & \longrightarrow & \Sigma^{2+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}(2,0) \\ & & \downarrow \\ P^{8j-6}_{8j-6} = \Sigma^{1+(j-1)\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \longrightarrow & \Sigma^{2+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}(1,0) \\ & & \downarrow \\ & & \Sigma^{2+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}} \end{array}$$

We have proved the following result.

Proposition 6.3 *Let $j \geq 1$. The bottom slice of $\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}$ is*

$$P^{2+4j}_{2+4j}(\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{1+\rho_K+\frac{j-1}{2}\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } j \text{ is odd}, \\ \Sigma^{2+\frac{j}{2}\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } j \text{ is even}. \end{cases}$$

6.1.2 $r = 3$ By (4.11), the cohomology of $S^{\mathbb{H}}$ is given by

$$\underline{\widetilde{H}}^n(S^{\mathbb{H}};\underline{\mathbb{Z}}) \cong \underline{\pi}_{-n}(\Sigma^{-\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}}^* & \text{if } n = 4, \\ \underline{mgw} & \text{if } n = 3. \end{cases}$$

Suspending by $3 + \rho_Q$ leads to the cofiber sequence

$$\begin{array}{ccc} P^8_8 = \Sigma^{\rho_Q} H_Q\underline{mgw} & \longrightarrow & \Sigma^{3+\rho_K} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ & & P^7_7 = \Sigma^{\rho_Q-1} H_Q\underline{\mathbb{Z}}^* \end{array}$$

The tower for $\Sigma^{3+j\rho_K} H_Q\underline{\mathbb{Z}}$, where $j \geq 2$, is then given recursively by

$$\begin{array}{ccc} P^{8j}_{8j} = \Sigma^{j\rho_Q} H_Q\underline{mgw} & \longrightarrow & \Sigma^{3+j\rho_K} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ & & \Sigma^{(j-1)\rho_K+\rho_Q-1} H_Q\underline{\mathbb{Z}}^* \\ & & \| \\ P^{8j-4}_{8j-4} = \Sigma^{2+(j-1)\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \longrightarrow & \Sigma^{3+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}(1,0) \\ & & \downarrow \\ & & \Sigma^{3+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}} \end{array}$$

The last cofiber sequence arises from Proposition 4.1. We have proved the following result.

Proposition 6.4 *Let $j \geq 1$. The bottom slice of $\Sigma^{3+j\rho_K} H_Q\underline{\mathbb{Z}}$ is*

$$P^{3+4j}_{3+4j}(\Sigma^{3+j\rho_K} H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{-1+\frac{j+1}{2}\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } j \text{ is odd}, \\ \Sigma^{3+\frac{j}{2}\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } j \text{ is even}. \end{cases}$$

6.1.3 $r = 4$ The tower for $\Sigma^{4+\rho_K} H_Q\underline{\mathbb{Z}}$ is given by

$$\begin{array}{ccc} P^{12}_{12} = \Sigma^{\rho_Q+1} H_Q\underline{mg} & \longrightarrow & \Sigma^{4+\rho_K} H_Q\underline{\mathbb{Z}} \simeq \Sigma^{2\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\ & & \downarrow \\ P^{10}_{10} = \Sigma^{\rho_Q+1}\underline{w} & \longrightarrow & \Sigma^{2\rho_K} H_Q\underline{\mathbb{Z}}(3,2) \\ & & \downarrow \\ & & P^8_8 = \Sigma^{\rho_Q} H_Q\underline{\mathbb{Z}}^* \end{array}$$

This uses the short exact sequence (Proposition 4.1)

$$\underline{\mathbb{Z}}(3,1) \hookrightarrow \underline{\mathbb{Z}}(3,2) \twoheadrightarrow \underline{m}^*,$$

the equivalence $\Sigma^{\rho_K} H_K\underline{m}^* \simeq \Sigma^2 H_K\underline{mg}$ [7, Proposition 4.8], and the computation (see Proposition 4.15)

$$\underline{\pi}_n\big(\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}(3,2)\big) \cong \begin{cases} \underline{w} & \text{if } n = 1, \\ \underline{\mathbb{Z}}^* & \text{if } n = 0. \end{cases}$$

The tower for $\Sigma^{4+j\rho_K} H_Q\underline{\mathbb{Z}}$, where $j \geq 2$, may then be obtained recursively from

$$\begin{array}{ccc}
P^{8j+4}_{8j+4} = \Sigma^{1+j\rho_Q} H_Q\underline{mg} & \longrightarrow & \Sigma^{4+j\rho_K} H_Q\underline{\mathbb{Z}} \simeq \Sigma^{(j+1)\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\
 & & \downarrow \\
P^{8j+2}_{8j+2} = \Sigma^{1+j\rho_Q}\underline{w} & \longrightarrow & \Sigma^{(j+1)\rho_K} H_Q\underline{\mathbb{Z}}(3,2) \\
 & & \downarrow \\
 & & \Sigma^{(j-1)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}^* \\
 & & \| \\
P^{8j-2}_{8j-2} = \Sigma^{3+(j-1)\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \longrightarrow & \Sigma^{4+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}(1,0) \\
 & & \downarrow \\
 & & \Sigma^{4+(j-2)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}
\end{array}$$

Proposition 6.5 *Let $j \geq 1$. The bottom slice of $\Sigma^{4+j\rho_K} H_Q\underline{\mathbb{Z}}$ is*

$$P^{4+4j}_{4+4j}(\Sigma^{4+j\rho_K} H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{\frac{j+1}{2}\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } j \text{ is odd,} \\ \Sigma^{4+\frac{j}{2}\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } j \text{ is even.} \end{cases}$$

6.1.4 $r = 5$ Here, we start with the slice tower for $\Sigma^5 H_Q\underline{\mathbb{Z}}$, as this is not a slice. The short exact sequence

$$\underline{\mathbb{Z}}(3,1) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \phi_Z^*\underline{B}(2,0)$$

gives rise to a cofiber sequence

$$P^8_8 = \Sigma^{\rho_Q} H_Q\phi_Z^*\underline{B}(2,0) \to \Sigma^5 H_Q\underline{\mathbb{Z}} \simeq \Sigma^{1+\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \to \Sigma^{1+\rho_K} H_Q\underline{\mathbb{Z}}.$$

Now the argument showing that $\Sigma^{2+\rho_K} H_Q\underline{\mathbb{Z}}$ is a 6-slice, given above in Section 6.1.1, also applies to show that $\Sigma^{1+\rho_K} H_Q\underline{\mathbb{Z}}$ is a 5-slice. Thus, this cofiber sequence is the slice tower for $\Sigma^5 H_Q\underline{\mathbb{Z}}$.

Next, the tower for $\Sigma^{5+\rho_K} H_Q\underline{\mathbb{Z}}$ is given by

$$\begin{array}{ccc}
P^{16}_{16} = \Sigma^{2\rho_Q} H_Q\phi_Z^*\underline{B}(2,0) & \longrightarrow & \Sigma^{5+\rho_K} H_Q\underline{\mathbb{Z}} \simeq \Sigma^{1+2\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\
 & & \downarrow \\
P^{12}_{12} = \Sigma^{2+\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \longrightarrow & \Sigma^{1+2\rho_K} H_Q\underline{\mathbb{Z}} \\
 & & \downarrow \\
 & & P^9_9 = \Sigma^{1+\rho_Q} H_Q\underline{\mathbb{Z}}^*
\end{array}$$

where the bottom cofiber sequence arises from the computation (Proposition 4.14)

$$\underline{\pi}_n(\Sigma^{\rho_K-\mathbb{H}} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \phi_Z^*\underline{\mathbb{F}_2} & \text{if } n = 1, \\ \underline{\mathbb{Z}}^* & \text{if } n = 0. \end{cases}$$

The tower for $\Sigma^{5+j\rho_K} H_Q\underline{\mathbb{Z}}$, where $j \geq 2$, may then be obtained recursively from

$$\begin{array}{ccc}
P^{8j+8}_{8j+8} = \Sigma^{(j+1)\rho_Q} H_Q \phi_Z^* \underline{B}(2,0) & \longrightarrow & \Sigma^{5+j\rho_K} H_Q\underline{\mathbb{Z}} \\
 & & \| \\
 & & \Sigma^{1+(j+1)\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\
 & & \downarrow \\
P^{8j+4}_{8j+4} = \Sigma^{2+j\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}}_2 & \longrightarrow & \Sigma^{1+(j+1)\rho_K} H_Q\underline{\mathbb{Z}} \\
 & & \downarrow \\
P^{8j}_{8j} = \Sigma^{j\rho_Q} H_Q \underline{B}(3,0) & \longrightarrow & \Sigma^{1+(j-1)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}^* \\
 & & \downarrow \\
 & & \Sigma^{1+(j-1)\rho_K+\rho_Q} H_Q\underline{\mathbb{Z}}
\end{array}$$

Proposition 6.6 *Let $j \geq 1$. The bottom slice of $\Sigma^{5+j\rho_K} H_Q\underline{\mathbb{Z}}$ is*

$$P^{5+4j}_{5+4j}(\Sigma^{5+j\rho_K} H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{1+\frac{j+1}{2}\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } j \text{ is odd,} \\ \Sigma^{1+\rho_K+\frac{j}{2}\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } j \text{ is even.} \end{cases}$$

6.2 Slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$

In this section, we describe all slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$ for $n \geq 0$.

Proposition 6.7 *The Q_8-spectrum $\Sigma^n H_Q\underline{\mathbb{Z}}$ is an n-slice for $0 \leq n \leq 4$.*

Proof Since this is true after restricting to any C_4 (see Section 5), any higher slices would necessarily be geometric and therefore occur in slice dimension at least 8. But we can show directly that $\Sigma^n H_Q\underline{\mathbb{Z}} < 8$ if $n \in [0,4]$. This follows from the vanishing of $\pi_{\rho_Q}\Sigma^n H_Q\underline{\mathbb{Z}} \cong \pi_{-n}\Sigma^{-\rho_Q} H_Q\underline{\mathbb{Z}}$ as displayed in Figure 4. □

It remains to determine the slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$ when $n \geq 5$. Note that Theorem 3.19 applies by [19, Proposition 8.5]. We first describe the bottom slice.

Proposition 6.8 (the n-slice) *For $n \geq 5$, write $n = 8k + r$, where $r \in [5,12]$. Then the n-slice of $\Sigma^n H_Q\underline{\mathbb{Z}}$ is*

$$P^n_n(\Sigma^n H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{1+\rho_K+k\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } r = 5, \\ \Sigma^{2+\rho_K+k\rho_Q} H_Q\underline{\mathbb{Z}}(3,2) & \text{if } r = 6, \\ \Sigma^{-1+(k+1)\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } r = 7, \\ \Sigma^{(k+1)\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } r = 8, \\ \Sigma^{1+(k+1)\rho_Q} H_Q\underline{\mathbb{Z}}^* & \text{if } r = 9, \\ \Sigma^{2+(k+1)\rho_Q} H_Q\underline{\mathbb{Z}}(1,0) & \text{if } r = 10, \\ \Sigma^{3+(k+1)\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } r = 11, \\ \Sigma^{4+(k+1)\rho_Q} H_Q\underline{\mathbb{Z}} & \text{if } r = 12. \end{cases}$$

Proof By Theorem 3.19, the n-slice of $\Sigma^n H_Q\underline{\mathbb{Z}}$ is the n-slice of the $\underline{\mathbb{Z}}$-module inflation of the n-slice of $\Sigma^n H_K\underline{\mathbb{Z}}$. By [19, Proposition 8.5], writing $n = 4j + r_4$ with $r_4 \in \{2,3,4,5\}$, we have

$$\Psi_Z^* P_n^n(\Sigma^n H_{K_4}\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1) & \text{if } n \equiv 2 \pmod 4, \\ \Sigma^{r_4+j\rho_K} H_Q\underline{\mathbb{Z}} & \text{otherwise.} \end{cases}$$

If $n \not\equiv 2 \pmod 4$, the slice tower was given in Section 6.1. For the case of $n \equiv 2$, since $\Sigma^{1+j\rho_K} H_Q\underline{m} \simeq \phi_Z^*(\Sigma^{1+j\rho_K} H_K\underline{m})$ is an $8j+4$-slice [19, Proposition 5.7], the cofiber sequence (Proposition 4.1)

$$\Sigma^{1+j\rho_K} H_Q\underline{m} \to \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1) \to \Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}, \tag{6.9}$$

combines with the work of Section 6.1.1 to show that

$$P_n^n\big(\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}(2,1)\big) \simeq P_n^n(\Sigma^{2+j\rho_K} H_Q\underline{\mathbb{Z}}).$$

The latter is given in Proposition 6.3. □

Proposition 6.10 (the $8k$-slices) *For $n \geq 5$ and $8k > n$, the $8k$-slice of $\Sigma^n H_Q\underline{\mathbb{Z}}$ is*

$$P_{8k}^{8k}(\Sigma^n H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^k H_Q\underline{g}^{n-k-3} & \text{if } 8k \in [4n-8, 8n-32], \\ \Sigma^{k\rho_Q} H_Q\underline{g}^{\frac{4k-n}{2}} & \text{if } 8k \in [2n+4, 4n-16] \text{ and } n \equiv 0 \pmod 2, \\ \Sigma^{k\rho_Q} H_Q\underline{g}^{\frac{4k-n-3}{2}} \oplus \phi_{LDR}^*\underline{\mathbb{F}_2}^* & \text{if } 8k \in [2n+4, 4n-12] \text{ and } n \equiv 1 \pmod 2, \\ \Sigma^{k\rho_Q} H_Q\underline{mg}^* & \text{if } 8k = 2n+2, \\ \Sigma^{k\rho_Q} H_Q\phi_Z^*\underline{B}(2,0) & \text{if } 8k = 2n-2, \\ \Sigma^{k\rho_Q} H_Q\underline{B}(3,0) & \text{if } 8k \in [n+3, 2n-10] \text{ and } n \equiv 1 \pmod 4, \\ \Sigma^{k\rho_Q} H_Q\underline{mgw} & \text{if } 8k \in [n+1, 2n] \text{ and } n \equiv 3 \pmod 4. \end{cases}$$

Proof This is a translation of Proposition 4.12. Alternatively, the slices above dimension $2n$ follow from Theorem 3.19 and [19, Proposition 8.6]. The slices in dimensions $2n$ and lower follow from the towers computed in Section 6.1. □

Proposition 6.11 (the $8k+4$-slices) *For $n \geq 5$ and $8k+4 > n$, the $8k+4$-slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$ are*

$$P_{8k+4}^{8k+4}(\Sigma^n H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{3+k\rho_Q} H_Q\phi_{LDR}^*\underline{\mathbb{F}_2} & \text{if } 8k+4 \in [2n+4, 4n-12] \text{ and } n \text{ is even}, \\ \Sigma^{2+k\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \text{if } 8k+4 \in [n+1, 2n-4] \text{ and } n \text{ is odd}, \\ \Sigma^{1+k\rho_Q} H_Q\underline{m} & \text{if } 8k+4 \in [n+2, 2n] \text{ and } n \equiv 2 \pmod 4, \\ \Sigma^{1+k\rho_Q} H_Q\underline{mg} & \text{if } 8k+4 \in [n+4, 2n-4] \text{ and } n \equiv 0 \pmod 4. \end{cases}$$

Proof The first case follows from [19, Proposition 8.7]. The remaining cases follow from (6.9) and Section 6.1. □

Proposition 6.12 (the $4k+2$-slices) *Let $n \geq 5$. If n is odd, then $\Sigma^n H_Q\underline{\mathbb{Z}}$ has no nontrivial $4k+2$-slices if $4k+2 > n$. If n is even and $8k+2 > n$, then the $8k+2$-slice of $\Sigma^n H_Q\underline{\mathbb{Z}}$ is nontrivial only if*

$8k+2 \in [n+1, 2n]$, *in which case the slice is*

$$P^{8k+2}_{8k+2}(\Sigma^n H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{1+k\rho_Q} H_Q\underline{w} & \text{if } n \equiv 0 \pmod 4, \\ \Sigma^{1+k\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2} & \text{if } n \equiv 2 \pmod 4. \end{cases}$$

Similarly, if n is even and $8k-2 > n$, the $8k-2$-slice is nontrivial only if $8k-2 \in [n+1, 2n]$, in which case the slice is

$$P^{8k-2}_{8k-2}(\Sigma^n H_Q\underline{\mathbb{Z}}) \simeq \begin{cases} \Sigma^{-1+k\rho_Q} H_Q\phi_Z^*\underline{\mathbb{F}_2}^* & \text{if } n \equiv 0 \pmod 4, \\ \Sigma^{-1+k\rho_Q} H_Q\underline{w}^* & \text{if } n \equiv 2 \pmod 4. \end{cases}$$

Proof According to [19], the K_4-spectrum $\Sigma^n H_K\underline{\mathbb{Z}}$ does not have any nontrivial slices in odd dimensions, except for the n-slice. By Theorem 3.19, this implies that $\Sigma^n H_Q\underline{\mathbb{Z}}$ does not have any $4k+2$-slices above dimension $2n$. The slices in dimensions below $2n$ are given in Section 6.1, using (6.1). □

6.3 Slice towers for $\Sigma^n H_Q\underline{\mathbb{Z}}$

By Proposition 6.7, $\Sigma^n H_Q\underline{\mathbb{Z}}$ is an n-slice for $n \in \{0, \dots, 4\}$. The slice tower for $\Sigma^5 H_Q\underline{\mathbb{Z}}$ was given in Section 6.1.4. We now display a few more examples of slice towers.

Example 6.13 The slice tower for $\Sigma^6 H_Q\underline{\mathbb{Z}}$ is

$$\begin{array}{ccc} P^{16}_{16} = \Sigma^2 H_Q\underline{g} & \longrightarrow & \Sigma^6 H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ P^{12}_{12} = \Sigma^{1+\rho} H_Q\underline{m} & \longrightarrow & \Sigma^{2+\rho_K} H_Q\underline{\mathbb{Z}}(2,1) \\ & & \downarrow \\ & & P^6_6 = \Sigma^{2+\rho_K} H_Q\underline{\mathbb{Z}} \end{array}$$

This follows immediately from combining [19, Example 8.2], (6.9), and Section 6.1.1.

Example 6.14 The slice tower for $\Sigma^7 H_Q\underline{\mathbb{Z}}$ is

$$\begin{array}{ccc} P^{24}_{24} = \Sigma^3 H_Q\underline{g} & \longrightarrow & \Sigma^7 H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ P^{16}_{16} = \Sigma^{2+\rho_Q} H_Q\underline{m} & \longrightarrow & \Sigma^{3+\rho_K} H_Q\underline{\mathbb{Z}}(2,1) \\ & & \downarrow \\ P^8_8 = \Sigma^{\rho_Q} H_Q\underline{mgw} & \longrightarrow & \Sigma^{3+\rho_K} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ & & P^7_7 = \Sigma^{\rho_Q - 1} H_{Q_8}\underline{\mathbb{Z}}^* \end{array}$$

This follows immediately from combining [19, Example 8.3] and Section 6.1.2.

Example 6.15 The slices, but not the slice tower, for $\Sigma^8 H_K\underline{\mathbb{Z}}$ were determined in [19, Section 8]. Let us denote by F the fiber of the map $H_Q\underline{\mathbb{Z}} \to H_Q\phi_{LDR}\underline{\mathbb{F}_2}$ induced by the map of Q_8-Mackey functors $\underline{\mathbb{Z}} \to \phi_{LDR}\underline{\mathbb{F}_2}$ that is surjective at L, D, and R. Then the nontrivial homotopy Mackey functors of F are $\underline{\pi}_0(F) \simeq \underline{\mathbb{Z}}(2,1)$ and $\underline{\pi}_{-1}(F) \cong \underline{g}^2$. The slice tower for $\Sigma^8 H_Q\underline{\mathbb{Z}}$ is

$$\begin{array}{ccc}
P^{32}_{32} = \Sigma^4 H_Q\underline{g} & \longrightarrow & \Sigma^8 H_Q\underline{\mathbb{Z}} \simeq \Sigma^{4+\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\
 & & \downarrow \\
P^{24}_{24} = \Sigma^3 H_Q\underline{g}^2 & \longrightarrow & \Sigma^{4+\rho_K} H_Q\underline{\mathbb{Z}}(2,1) \\
 & & \downarrow \\
P^{20}_{20} = \Sigma^{3+\rho_Q} H_Q\phi^*_{LDR}\underline{\mathbb{F}_2} & \longrightarrow & \Sigma^{4+\rho_K} F \\
 & & \downarrow \\
P^{12}_{12} = \Sigma^{1+\rho_Q} H_Q\underline{mg} & \longrightarrow & \Sigma^{4+\rho_K} H_Q\underline{\mathbb{Z}} \simeq \Sigma^{2\rho_K} H_Q\underline{\mathbb{Z}}(3,1) \\
 & & \downarrow \\
P^{10}_{10} = \Sigma^{\rho_Q+1}\underline{w} & \longrightarrow & \Sigma^{2\rho_K} H_Q\underline{\mathbb{Z}}(3,2) \\
 & & \downarrow \\
 & & P^8_8 = \Sigma^{\rho_Q} H_Q\underline{\mathbb{Z}}^*
\end{array}$$

where the bottom of the tower comes from Section 6.1.3.

7 Homology calculations

In Section 6, we described the slices of $\Sigma^n H_Q\underline{\mathbb{Z}}$. In Section 8 below, we will give the corresponding slice spectral sequences. The E_2-pages of those spectral sequences are given by the homotopy Mackey functors of the slices. We describe those homotopy Mackey functors here.

7.1 The n-slice

We start with the n-slices in the order listed in Proposition 6.8. The homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q\underline{\mathbb{Z}}$ were calculated in Proposition 4.10. We use the same methods to determine the homotopy Mackey functors of $\Sigma^{\rho_K+j\rho_Q} H_Q\underline{\mathbb{Z}}$.

Proposition 7.1 *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{\rho_K+j\rho_Q} H_Q\underline{\mathbb{Z}}$ are*

$$\underline{\pi}_i(\Sigma^{\rho_K+j\rho_Q} H_Q\underline{\mathbb{Z}}) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } i = 8j+4, \\ \underline{mgw} & \text{if } i \in [4j+4, 8j+3] \text{ and } i \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } i \in [4j+4, 8j+3] \text{ and } i \equiv 0 \pmod 4, \\ \phi^*_Z\underline{\pi}_i(\Sigma^{(j+1)\rho_K} H_K\underline{\mathbb{Z}}) & \text{if } i \in [j+1, 4j+3]. \end{cases}$$

See Proposition 4.2 or Figure 1 for the homotopy Mackey functors of $\Sigma^{(j+1)\rho_K} H_K\underline{\mathbb{Z}}$.

We may now use Proposition 7.1 and the exact sequence $\underline{\mathbb{Z}}(3,2) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \underline{g}$ to get the homotopy Mackey functors of $\Sigma^{\rho_K + j\rho_Q} H_Q \underline{\mathbb{Z}}(3,2)$.

Proposition 7.2 *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{\rho_K + j\rho_Q} H_Q \underline{\mathbb{Z}}(3,2)$ are*

$$\underline{\pi}_i(\Sigma^{\rho_K + j\rho_Q} H_Q \underline{\mathbb{Z}}(3,2)) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } i = 8j+4, \\ \underline{mgw} & \text{if } i \in [4j+4, 8j+3] \text{ and } i \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } i \in [4j+4, 8j+3] \text{ and } i \equiv 0 \pmod 4, \\ \phi_Z^* \underline{\pi}_i(\Sigma^{(j+1)\rho_K} H_K \underline{\mathbb{Z}}) & \text{if } i \in [j+2, 4j+3]. \end{cases}$$

The key point here is that the homotopy Mackey functors of $\Sigma^{\rho_K + j\rho_Q} H_Q \underline{\mathbb{Z}}(3,2)$ are the same as that of $\Sigma^{\rho_K + j\rho_Q} H_Q \underline{\mathbb{Z}}$, except that the $\underline{g}$ in degree $j+1$ has been removed.

In Proposition 4.12 we list the homotopy Mackey functors of $\Sigma^{-j\rho_Q} H_Q \underline{\mathbb{Z}}$. Anderson duality then provides us with the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}^*$.

Proposition 7.3 *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}^*$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}^*) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } i = 8j, \\ \underline{mgw} & \text{if } i \in [4j+1, 8j-1] \text{ and } i \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } i \in [4j+1, 8j-1] \text{ and } i \equiv 0 \pmod 4, \\ \phi_Z^* \underline{B}(2,0) & \text{if } i = 4j, \\ \phi_Z^* \underline{\pi}_{i-4}(\Sigma^{(j-1)\rho_K} H_K \underline{\mathbb{Z}}) & \text{if } i \in [j+3, 4j-1]. \end{cases}$$

Finally, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}(1,0)$ follow from the exact sequence

$$\underline{\mathbb{Z}}(1,0) \hookrightarrow \underline{\mathbb{Z}} \twoheadrightarrow \phi_Z^* \underline{\mathbb{F}_2}.$$

Proposition 7.4 *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}(1,0)$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}(1,0)) \cong \begin{cases} \underline{\mathbb{Z}} & \text{if } i = 8j, \\ \underline{mgw} & \text{if } i \in [4j+1, 8j-2] \text{ and } i \equiv 2 \pmod 4, \\ \underline{B}(3,0) & \text{if } i \in [4j+1, 8j-2] \text{ and } i \equiv 0 \pmod 4, \\ \phi_Z^* \underline{B}(2,0) & \text{if } i = 4j, \\ \phi_Z^* \underline{\pi}_i(\Sigma^{j\rho_K} H_K \underline{\mathbb{Z}}) & \text{if } i \in [j, 4j-1]. \end{cases}$$

7.2 The $8k$-slices

We now move on to the $8k$-slices.

Proposition 7.5 *For $j = 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{B}(2,0)$ are*

$$\underline{\pi}_i(\Sigma^{\rho_Q} H_Q \phi_Z^* \underline{B}(2,0)) \cong \begin{cases} \underline{mg} & \text{if } i = 2, \\ \underline{g} & \text{if } i = 1. \end{cases}$$

For $j \geq 2$, they are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{B}(2,0)) \cong \begin{cases} \phi_{LDR}^* \underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j+2, 2j-1], \\ \underline{g}^2 & \text{if } i = j+1, \\ \underline{g} & \text{if } i = j. \end{cases}$$

Proof Because $\phi_Z^* \underline{B}(2,0)$ is a pullback,

$$\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{B}(2,0) \simeq \Sigma^{j\rho_K} H_Q \phi_Z^* \underline{B}(2,0).$$

The exact sequence of K-Mackey functors $\underline{m}^* \to \underline{B}(2,0) \to \underline{g}$ provides us with

$$\Sigma^{j\rho_K} H_K \underline{m}^* \to \Sigma^{j\rho_K} H_K \underline{B}(2,0) \to \Sigma^{j\rho_K} H_K \underline{g}.$$

The conclusion follows from [7, Propositions 4.8 and 7.4] and the resulting long exact sequence in homotopy. □

We may again use this strategy of reducing the calculations from Q to K for determining the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{B}(3,0)$.

Proposition 7.6 *For $j = 1$ the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{B}(3,0)$ are*

$$\underline{\pi}_i(\Sigma^{\rho_K} H_K \underline{B}(3,0)) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4, \\ \underline{mg} & \text{if } i = 2, \\ \underline{g} & \text{if } i = 1. \end{cases}$$

For $j \geq 2$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{B}(3,0)$ are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{B}(3,0)) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus g^{4j-2-i} & \text{if } i \in [2j+2, 4j-2], \\ \underline{g}^{2(k-2)+1} & \text{if } i = 2j+1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus \underline{g}^{2(j-3)+1} & \text{if } i = 2j, \\ \underline{g}^{2(i-j-1)} & \text{if } i \in [j+3, 2j-1], \\ \underline{g}^{i-j+1} & \text{if } i \in [j, j+2]. \end{cases}$$

Proof Because the underlying spectrum of $H_Q \underline{B}(3,0)$ is contractible,

$$\Sigma^{\rho_Q} H_Q \underline{B}(3,0) \simeq \Sigma^{\rho_K} H_Q \underline{B}(3,0).$$

Now, we may consider $\underline{B}(3,0)$ as a pullback $\phi_Z^* B := \underline{B}(3,0)$; thus the calculation is reduced to one of K-Mackey functors. The sequence of K-Mackey functors $\underline{\mathbb{Z}}^* \xrightarrow{2} \underline{\mathbb{Z}} \to \underline{B}$ provides us with

$$\Sigma^{j\rho_K} H_K \underline{\mathbb{Z}}^* \to \Sigma^{j\rho_K} H_K \underline{\mathbb{Z}} \to \Sigma^{j\rho_K} H_K \underline{B}.$$

Except for $i = 4j-2$, the result follows from the associated long exact sequence in homotopy. In degree $4j-2$ we have an extension

$$\underline{mg} \to \underline{\pi}_{4j-2}(\Sigma^{j\rho_K} H\underline{B}) \to \underline{g}.$$

We need to show this is not the split extension. This follows from the exact sequence $\underline{B}(2,0) \to \underline{B} \to \underline{\mathbb{F}_2}$ of K-Mackey functors. □

Proposition 7.7 *For $j = 1$ and $j = 2$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q\underline{mgw}$ are*

$$\underline{\pi}_i(\Sigma^{\rho_Q} H_Q\underline{mgw}) \cong \begin{cases} \phi_Z^*\underline{\mathbb{F}_2} & \text{if } i = 4, \\ \phi_Z^*\underline{B}(2,0) & \text{if } i = 2. \end{cases}$$

and

$$\underline{\pi}_i(\Sigma^{2\rho_Q} H_Q\underline{mgw}) \cong \begin{cases} \phi_Z^*\underline{\mathbb{F}_2} & \text{if } i = 8, \\ \underline{mg} & \text{if } i = 7, \\ \phi_{LDR}\underline{\mathbb{F}_2} & \text{if } i = 6, \\ \underline{g} & \text{if } i = 5, \\ \underline{mg} & \text{if } i = 4, \\ \underline{g} & \text{if } i = 3. \end{cases}$$

For $j \geq 3$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q\underline{mgw}$ are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q\underline{mgw}) \cong \begin{cases} \phi_Z^*\underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}\underline{\mathbb{F}_2} \oplus \underline{g}^{4j-i-2} & \text{if } i \in [2j+2, 4j-2], \\ \underline{g}^{2j-3} & \text{if } i = 2j+1, \\ \underline{g}^{2j-5} \oplus \phi_{LDR}\underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^{2(i-j)-2} & \text{if } i \in [j+2, 2j-1], \\ \underline{g} & \text{if } i = j+1. \end{cases}$$

Proof We first deal with the case $j = 1$. The short exact sequence of Mackey functors

$$\underline{w}^* \hookrightarrow \underline{mgw} \twoheadrightarrow \underline{mg}^*$$

combines with Propositions 7.17 and 7.9 to show that the only nontrivial Mackey functors are $\phi_Z^*\underline{\mathbb{F}_2}$ in degree 4 and an extension of $\underline{m}$ by $\underline{g}$ in degree 2. It remains to see that this extension is $\phi_Z^*\underline{B}(2,0)$. According to Proposition 4.12, the Postnikov tower for $\Sigma^{-\rho_Q} H_Q\underline{\mathbb{Z}}$ is

$$\begin{array}{ccc} \Sigma^{-5} H_Q\phi_Z^*\underline{B}(2,0) & \longrightarrow & \Sigma^{-\rho_Q} H_Q\underline{\mathbb{Z}} \\ & & \downarrow \\ \Sigma^{-7} H_Q\underline{mgw} & \longrightarrow & X \\ & & \downarrow \\ & & \Sigma^{-8} H_Q\underline{\mathbb{Z}}^* \end{array}$$

Desuspending this diagram once by ρ_Q gives a tower for computing the homotopy Mackey functors of $\Sigma^{-2\rho_Q} H_Q\underline{\mathbb{Z}}$. The homotopy Mackey functors for $\Sigma^{-8-\rho_Q} H_Q\underline{\mathbb{Z}}^*$ and $\Sigma^{-5\rho_Q} H_Q\Psi^*\underline{B}(2,0)$ follow, using Anderson duality, from Propositions 4.10 and 7.5. Long exact sequences in homotopy then imply that

$$\underline{\pi}_{-9}(\Sigma^{-7-\rho_Q} H_Q\underline{mgw}) \cong \phi_Z^*\underline{B}(2,0).$$

Dualizing gives that $\underline{\pi}_2(\Sigma^{\rho_Q} H_Q\underline{mgw})$ is $\phi_Z^*\underline{B}(2,0)$.

We now have a fiber sequence

$$\Sigma^4 H_Q\phi_Z^*\underline{\mathbb{F}}_2 \to \Sigma^{\rho_Q} H_Q\underline{mgw} \to \Sigma^2 H_Q\phi_Z^*\underline{B}(2,0). \tag{7.8}$$

Suspending this sequence by ρ_Q immediately gives the homotopy Mackey functors of $\Sigma^{2\rho_Q} H_Q\underline{mgw}$. The same is true in the case $j = 3$, except that we have an extension

$$\underline{g} \hookrightarrow \underline{\pi}_6\Sigma^{3\rho_Q} H_Q\underline{mgw} \twoheadrightarrow \phi_{LDR}\underline{\mathbb{F}}_2.$$

We claim that, more generally, any extension of $\underline{\mathbb{Z}}$-modules

$$\underline{g}^m \hookrightarrow \underline{E} \twoheadrightarrow \phi_{LDR}\underline{\mathbb{F}}_2$$

is necessarily the split extension. To see this, first note that $\phi_{LDR}\underline{\mathbb{F}}_2$ is, by definition, the direct sum $\phi_L^*\underline{\mathbb{F}}_2 \oplus \phi_D^*\underline{\mathbb{F}}_2 \oplus \phi_R^*\underline{\mathbb{F}}_2$. It therefore suffices to show that the only $\underline{\mathbb{Z}}$-module extension of $\phi_L^*\underline{\mathbb{F}}_2$ by $\underline{g}^m$ is the split extension. Since any such extension will vanish at the subgroups D and R, the $\underline{\mathbb{Z}}$-module structure forces the value at Q to be 2-torsion and therefore equal to $\underline{\mathbb{F}}_2{}^{m+1}$. Since there is a nontrivial restriction to the subgroup L, the $\underline{\mathbb{Z}}$-module structure forces the transfer from L to vanish. Thus the extension must be the split extension.

The suspension by $(j-1)\rho_Q$ of (7.8) gives the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q\underline{mgw}$ in degrees $2j+1$ and higher. Now we argue by induction that the Mackey functors for $\Sigma^{j\rho_Q} H_Q\underline{mgw}$ are as claimed, for $j \geq 3$. For instance, since the bottom Mackey functor is

$$\underline{\pi}_j(\Sigma^{(j-1)\rho_Q} H_Q\underline{mgw}) \cong \underline{g},$$

we see by decomposing $\Sigma^{(j-1)\rho_Q} H_Q\underline{mgw}$ using the Postnikov tower that

$$\underline{\pi}_{j+1}(\Sigma^{j\rho_Q} H_Q\underline{mgw}) \cong \underline{g}.$$

The values of the Mackey functors $\underline{\pi}_i$, for $i \leq 2j-2$, follow in a similar way. The values

$$\underline{\pi}_{2j-2}(\Sigma^{(j-1)\rho_Q} H_Q\underline{mgw}) \cong \underline{g}^{2j-7} \oplus \phi_{LDR}\underline{\mathbb{F}}_2,$$

and

$$\underline{\pi}_{2j-1}(\Sigma^{(j-1)\rho_Q} H_Q\underline{mgw}) \cong \underline{g}^{2j-5}$$

give that

$$\underline{\pi}_{2j-1}(\Sigma^{j\rho_Q} H_Q \underline{mgw}) \cong \underline{g}^{2j-4}$$

and that we have an extension of $\underline{\mathbb{Z}}$-modules

$$\underline{g}^{2j-5} \hookrightarrow \underline{\pi}_{2j}(\Sigma^{j\rho_Q} H_Q \underline{\mathbb{Z}}) \twoheadrightarrow \phi_{LDR}\underline{\mathbb{F}_2}.$$

By the argument given above, this must be the split extension. □

The homotopy Mackey functors for the remaining $8k$-slices follow from [19, Propositions 9.5 and 9.8].

Proposition 7.9 [7, Proposition 4.8; 19, Proposition 9.5] *We have the equivalence*

$$\Sigma^{\rho_Q} H_Q \underline{mg}^* \simeq \Sigma^2 H_Q \underline{m}.$$

For $j \geq 2$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{mg}^*$ are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{mg}^*) \cong \begin{cases} \phi^*_{LDR}\underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j+2, 2j-1], \\ \underline{g} & \text{if } i = j+1. \end{cases}$$

Proposition 7.10 [19, Proposition 9.8] *We have equivalences*

$$\Sigma^{j\rho_Q} H\phi^*_{LDR}\underline{\mathbb{F}_2}^* \simeq \begin{cases} \Sigma^2 H\phi^*_{LDR}\underline{f} & \text{if } j = 1, \\ \Sigma^4 H\phi^*_{LDR}\underline{\mathbb{F}_2} & \text{if } j = 2. \end{cases}$$

Then for $j \geq 3$, the nontrivial homotopy Mackey functors of $\Sigma^{j\rho_Q} H\phi^*_{LDR}\underline{\mathbb{F}_2}^*$ are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \phi^*_{LDR}\underline{\mathbb{F}_2}^*) = \begin{cases} \phi^*_{LDR}\underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j+2, 2j-1]. \end{cases}$$

7.3 The $8k+4$-slices

Similarly, the homotopy Mackey functors of the $8k+4$-slices follow from [19, Proposition 9.8] and [7, Corollary 7.2, Propositions 7.3 and 7.4].

Proposition 7.11 [7, Proposition 3.6] *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \phi^*_{LDR}\underline{\mathbb{F}_2}$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \phi^*_{LDR}\underline{\mathbb{F}_2}) \cong \begin{cases} \phi^*_{LDR}\underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j, 2j-1]. \end{cases}$$

Proposition 7.12 [7, Corollary 7.2] *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus g^{4j-2-i} & \text{if } i \in [2j, 4j-2], \\ \underline{g}^{2(i-j)+1} & \text{if } i \in [j, 2j-1]. \end{cases}$$

Proposition 7.13 [7, Proposition 7.3] *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{m}$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{m}) \cong \begin{cases} \phi_{LDR}^* \underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j+1, 2j-1], \\ \underline{g} & \text{if } i = j. \end{cases}$$

Proposition 7.14 [7, Proposition 7.4] *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{mg}$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{mg}) \cong \begin{cases} \phi_{LDR}^* \underline{\mathbb{F}_2} & \text{if } i = 2j, \\ \underline{g}^3 & \text{if } i \in [j+1, 2j-1], \\ \underline{g}^2 & \text{if } i = j. \end{cases}$$

7.4 The $4k+2$-slices

The homotopy Mackey functors of the $4k+2$-slice $\Sigma^{1+k\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}$ are given in Proposition 7.12. The homotopy Mackey functors of the remaining $4k+2$-slices are as follows.

Proposition 7.15 [7, Proposition 4.8, Corollary 7.2] *We have the equivalence $\Sigma^{\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}^* \simeq \Sigma^4 H_Q \phi_Z^* \underline{\mathbb{F}_2}$. For $j \geq 2$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}^*$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \phi_Z^* \underline{\mathbb{F}_2}^*) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus g^{4j-2-i} & \text{if } i \in [2j+2, 4j-2], \\ \underline{g}^{2(i-j)-5} & \text{if } i \in [j+3, 2j+1]. \end{cases}$$

Finally, we have the homotopy of $\Sigma^{j\rho_Q} H_Q \underline{w}$ and $\Sigma^{j\rho_Q} H_Q \underline{w}^*$.

Proposition 7.16 *For $j \geq 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{w}$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{w}) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus g^{4j-2-i} & \text{if } i \in [2j, 4j-2], \\ \underline{g}^{2(i-j)+1} & \text{if } i \in [j+1, 2j-1]. \end{cases}$$

Proof The underlying spectrum of $\Sigma^{j\rho_Q} H_Q \underline{w}$ is contractible; thus,

$$\Sigma^{j\rho_Q} H_Q \underline{w} \simeq \Sigma^{j\rho_K} H_Q \underline{w}.$$

Then, because $\underline{w}$ is a pullback over Z, the calculation is essentially K-equivariant. Consider the short exact sequence of K-Mackey functors $\underline{w} \to \underline{\mathbb{F}_2} \to \underline{g}$ and the corresponding cofiber sequence $\Sigma^{j\rho_K} H_K \underline{w} \to \Sigma^{j\rho_K} H_K \underline{\mathbb{F}_2} \to \Sigma^{j\rho_K} H_K \underline{g}$. The statement follows immediately from the resulting long exact sequence in homotopy. □

Proposition 7.17 *For $j = 1$, the homotopy Mackey functors of $\Sigma^{j\rho_Q} H_Q \underline{w}^*$ are*

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{w}^*) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4, \\ \underline{g} & \text{if } i = 2. \end{cases}$$

For $j \geq 2$, they are

$$\underline{\pi}_i(\Sigma^{j\rho_Q} H_Q \underline{w}^*) \cong \begin{cases} \phi_Z^* \underline{\mathbb{F}_2} & \text{if } i = 4j, \\ \underline{mg} & \text{if } i = 4j-1, \\ \phi_{LDR}^* \underline{\mathbb{F}_2} \oplus g^{4j-2-i} & \text{if } i \in [2j+2, 4j-2], \\ \underline{g}^{2(i-j)-5} & \text{if } i \in [j+3, 2j+1], \\ \underline{g} & \text{if } i = j+1. \end{cases}$$

Proof The proof is the same as that in Proposition 7.16, except that we start with the exact sequence of K-Mackey functors $\underline{g} \to \underline{\mathbb{F}_2}^* \to \underline{w}^*$. □

8 Slice spectral sequences

Here we include the slice spectral sequences for $\Sigma^n H_Q \underline{\mathbb{Z}}$ for several values of n between 5 and 15. See Figures 5–9. In some cases, we use the restriction to the C_4-subgroups to determine some of the slice differentials.

The grading is the same as that in [9, Section 4.4.2]. The Mackey functor $\underline{E}_2^{t-n,t}$ is $\underline{\pi}_n P_t^t(X)$. We also follow the Adams convention, where $\underline{\pi}_n P_t^t(X)$ has coordinates $(n, t-n)$ and the differential

$$d_r : \underline{E}_r^{s,t} \to \underline{E}_r^{s+r,t+r-1}$$

points left one and up r.

$\square = \underline{\mathbb{Z}}$	$= \phi_Z^* \underline{\mathbb{F}_2}$	$= \phi_{LDR}^* \underline{\mathbb{F}_2}$	$= \underline{mgw}$
$\circ = \underline{B}(3,0)$	$= \phi_Z^* \underline{B}(2,0)$	$= \underline{mg}$	$= \underline{g}^n$

Table 4: Symbols for Q-Mackey functors.

$\square = \underline{\mathbb{Z}}$	$\blacklozenge = \phi^*_{C_2}\underline{\mathbb{F}}_2$	$\circ = \underline{B}(2,0)$	$\bullet = \underline{g}$

Table 5: Symbols for C_4-Mackey functors.

The Q-Mackey functors that appear in these spectral sequences are listed in Table 4. We also display some companion C_4-slice spectral sequences, and the C_4-Mackey functors that appear are listed in Table 5.

Example 8.1 In the spectral sequences for $\Sigma^5 H_Q\underline{\mathbb{Z}}$, $\Sigma^6 H_Q\underline{\mathbb{Z}}$, and $\Sigma^7 H_Q\underline{\mathbb{Z}}$, because we must be left with

$$\underline{\pi}_n(P^n_n \Sigma^n H_Q\underline{\mathbb{Z}}) \cong \underline{\mathbb{Z}},$$

all differentials are forced.

Example 8.2 For $\Sigma^8 H_Q\underline{\mathbb{Z}}$, the pattern of differentials emanating from the Mackey functor

$$\underline{\pi}_6(P^8_8 \Sigma^8 H_Q\underline{\mathbb{Z}})$$

is forced; no other pattern of differentials wipes out all classes in this region. The shorter differentials clearing out the smaller region are then similarly forced.

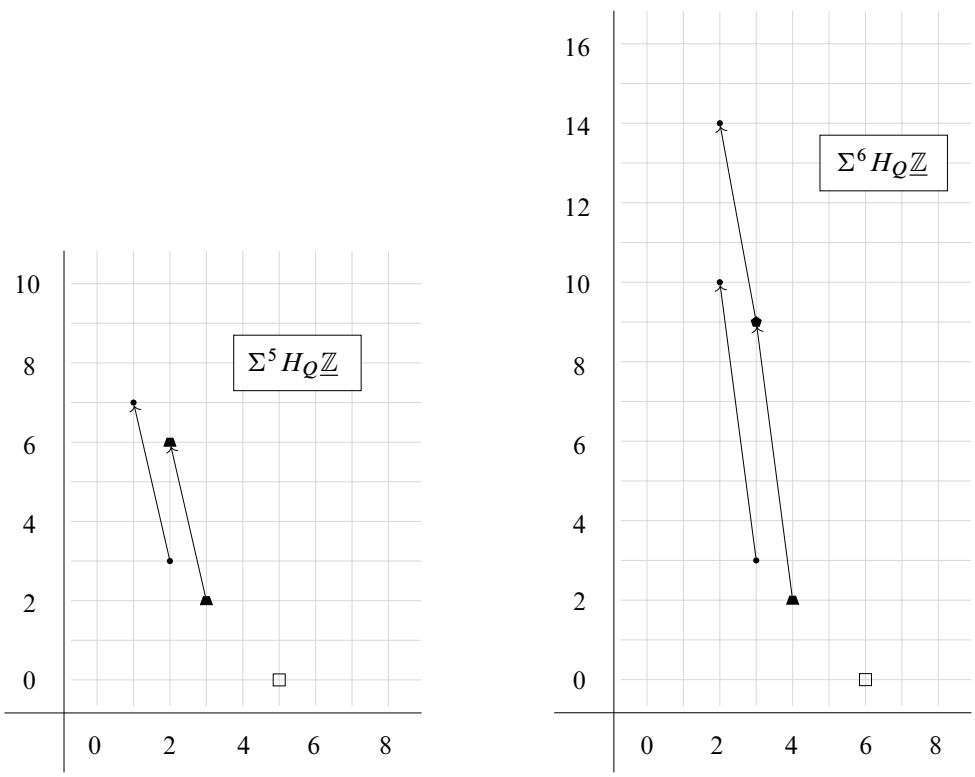


Figure 5

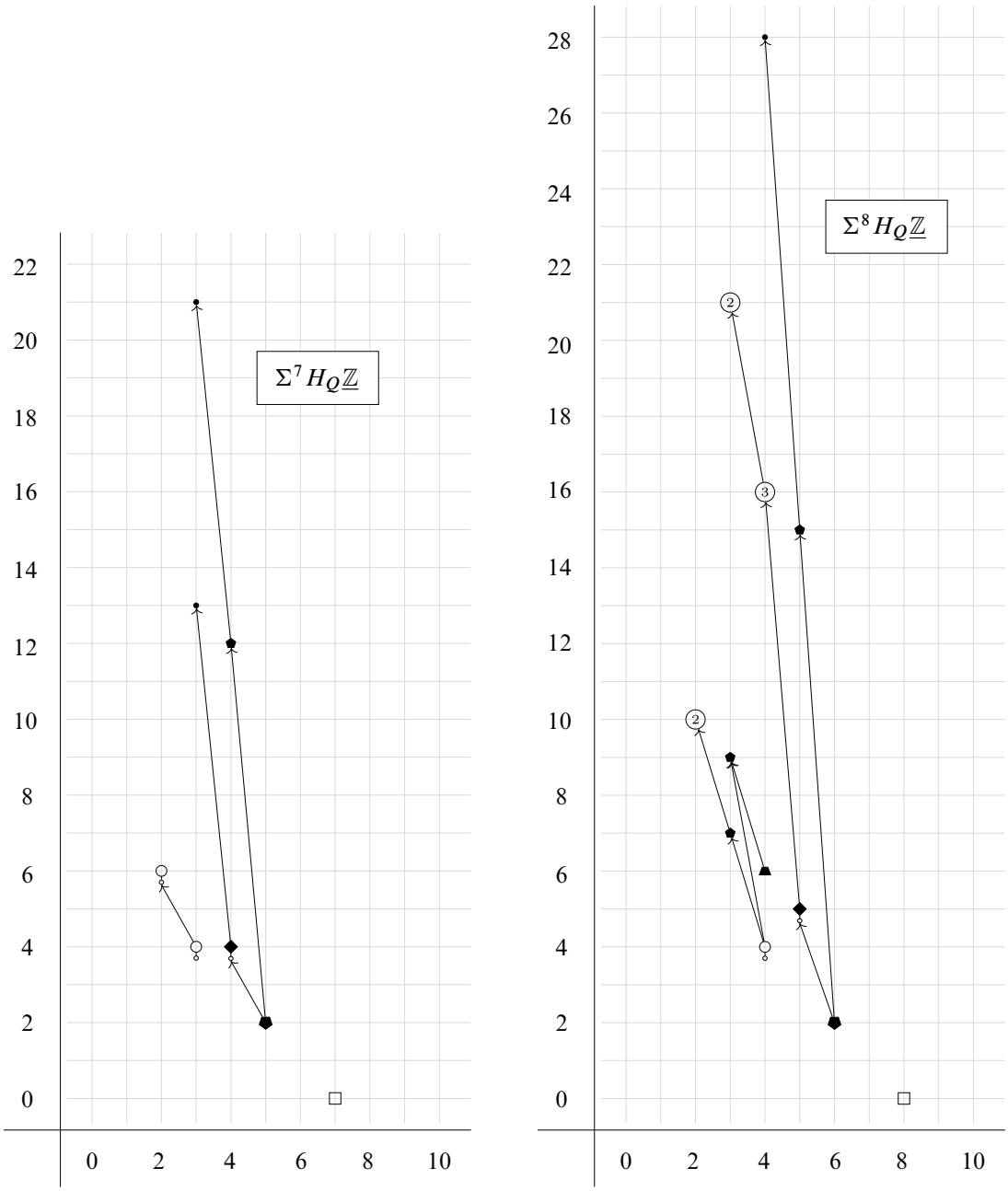


Figure 6

Example 8.3 In the cases of $\Sigma^n H_Q\underline{\mathbb{Z}}$ for $n = 10$, 12, and 15, we also display the corresponding slice spectral sequence for $\Sigma^n H_{C_4}\underline{\mathbb{Z}}$, where we use C_4 to indiscriminately refer to any of the subgroups $L, D, R \leq Q$. The slice differentials in the C_4-case force many of the slice differentials for the Q-equivariant spectra.

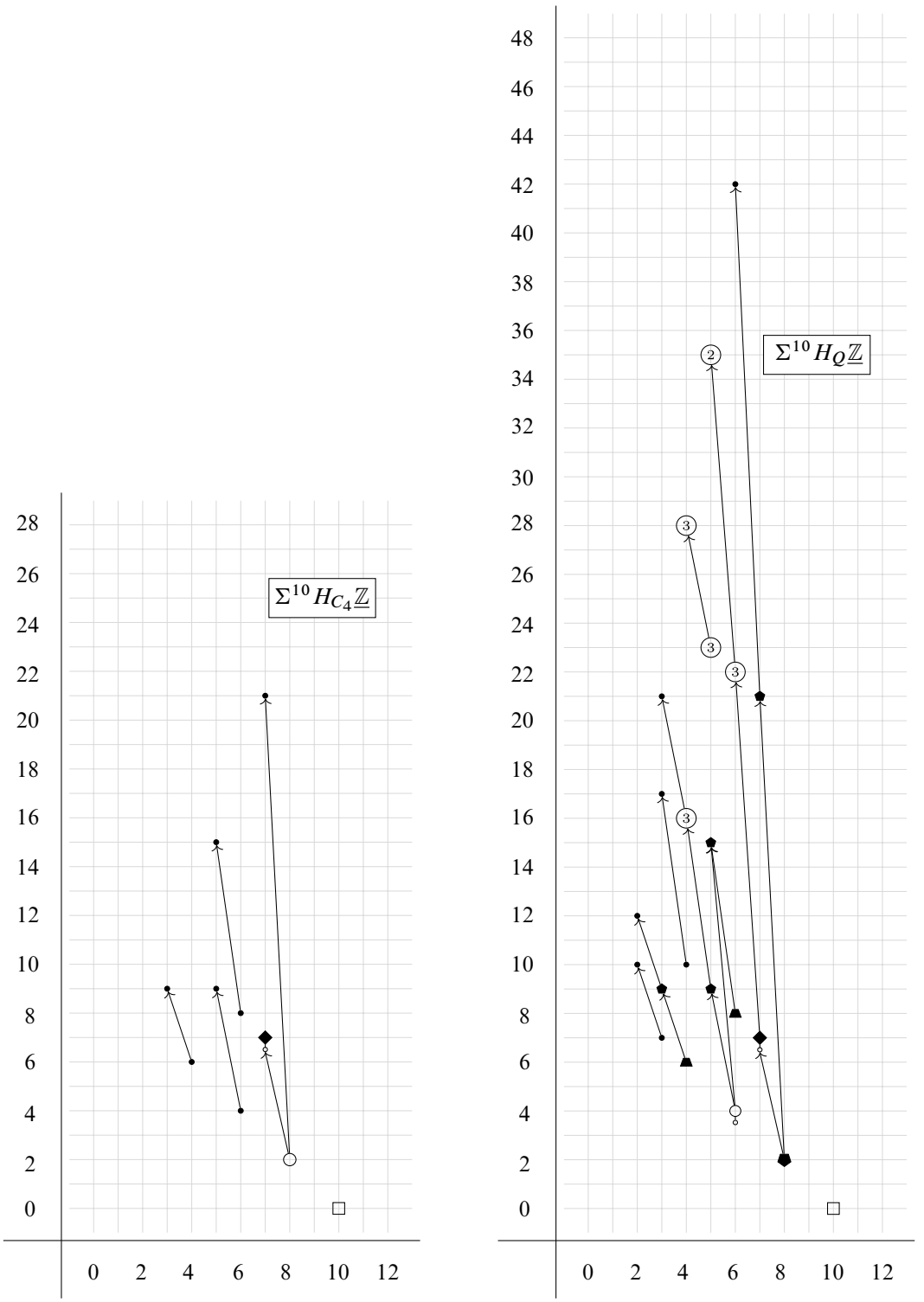

$\Sigma^{10} H_{C_4}\underline{\mathbb{Z}}$
$\Sigma^{10} H_Q\underline{\mathbb{Z}}$
0 2 4 6 8 10 12
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48
2
3

Figure 7

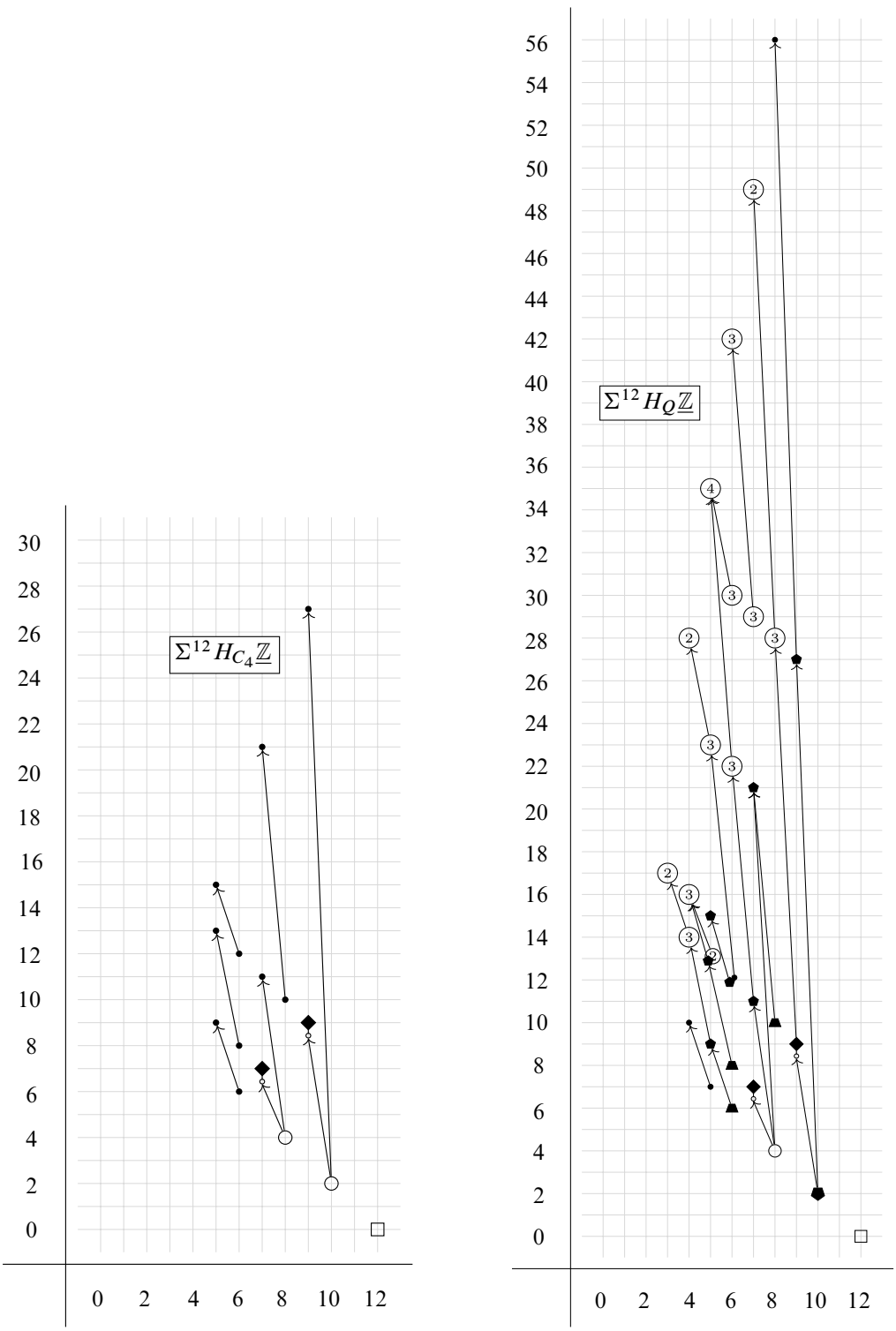

$\Sigma^{12} H_{C_4}\underline{\mathbb{Z}}$
$\Sigma^{12} H_Q\underline{\mathbb{Z}}$

Figure 8

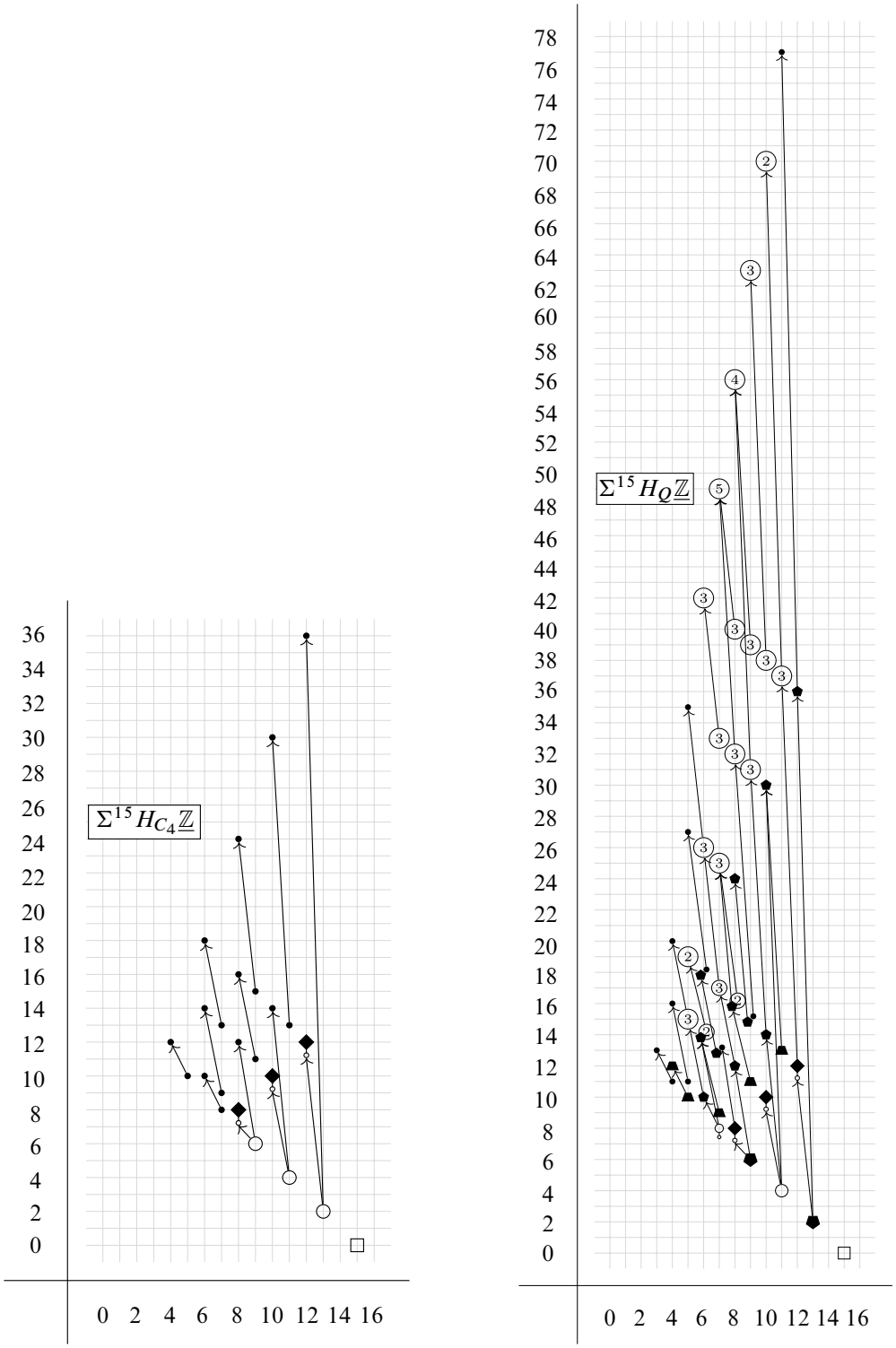

$\Sigma^{15} H_{C_4}\underline{\mathbb{Z}}$
$\Sigma^{15} H_{Q}\underline{\mathbb{Z}}$

Figure 9

References

[1] **V Angeltveit**, *The slice spectral sequence for the cyclic group of order p*, preprint (2021) arXiv 2111.06559

[2] **P Balmer**, **I Dell'Ambrogio**, **B Sanders**, *Grothendieck–Neeman duality and the Wirthmüller isomorphism*, Compos. Math. 152 (2016) 1740–1776 MR Zbl

[3] **S Basu**, **S Ghosh**, *Non-trivial extensions in equivariant cohomology with constant coefficients*, preprint (2021) arXiv 2108.12763

[4] **S Bouc**, **R Stancu**, **P Webb**, *On the projective dimensions of Mackey functors*, Algebr. Represent. Theory 20 (2017) 1467–1481 MR Zbl

[5] **K S Brown**, *Cohomology of groups*, Grad. Texts in Math. 87, Springer (1982) MR Zbl

[6] **D Dugger**, *An Atiyah–Hirzebruch spectral sequence for KR-theory*, K-Theory 35 (2005) 213–256 MR Zbl

[7] **B Guillou**, **C Yarnall**, *The Klein four slices of $\Sigma^n H\underline{\mathbb{F}}_2$*, Math. Z. 295 (2020) 1405–1441 MR Zbl

[8] **M A Hill**, *The equivariant slice filtration: a primer*, Homology Homotopy Appl. 14 (2012) 143–166 MR Zbl

[9] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *On the nonexistence of elements of Kervaire invariant one*, Ann. of Math. 184 (2016) 1–262 MR Zbl

[10] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *The slice spectral sequence for certain $RO(C_{p^n})$-graded suspensions of $H\underline{\mathbb{Z}}$*, Bol. Soc. Mat. Mex. 23 (2017) 289–317 MR Zbl

[11] **M A Hill**, **M J Hopkins**, **D C Ravenel**, *The slice spectral sequence for the C_4 analog of real K-theory*, Forum Math. 29 (2017) 383–447 MR Zbl

[12] **M A Hill**, **X D Shi**, **G Wang**, **Z Xu**, *The slice spectral sequence of a C_4-equivariant height-4 Lubin–Tate theory*, Mem. Amer. Math. Soc. 1429, Amer. Math. Soc., Providence, RI (2023) MR Zbl

[13] **M A Hill**, **C Yarnall**, *A new formulation of the equivariant slice filtration with applications to C_p-slices*, Proc. Amer. Math. Soc. 146 (2018) 3605–3614 MR Zbl

[14] **P Hu**, **I Kriz**, *The homology of BPO*, from "Recent progress in homotopy theory", Contemp. Math. 293, Amer. Math. Soc., Providence, RI (2002) 111–123 MR

[15] **S Kriz**, *Notes on equivariant homology with constant coefficients*, Pacific J. Math. 309 (2020) 381–399 MR Zbl

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

[17] **Y Lu**, *On the $RO(G)$-graded coefficients of Q_8 equivariant cohomology*, Topology Appl. 307 (2022) art. id. 107921 MR Zbl

[18] **L Meier**, **X D Shi**, **M Zeng**, *The localized slice spectral sequence, norms of real bordism, and the Segal conjecture*, Adv. Math. 412 (2023) art. id. 108804 MR Zbl

[19] **C Slone**, *Klein four 2-slices and the slices of $\Sigma^{\pm n} H\underline{\mathbb{Z}}$*, Math. Z. 301 (2022) 3895–3938 MR Zbl

[20] **Y J F Sulyma**, *Equivariant aspects of topological Hochschild homology*, PhD thesis, University of Texas at Austin (2019) Available at `https://www.proquest.com/docview/2473436390`

[21] **J Thévenaz**, **P Webb**, *The structure of Mackey functors*, Trans. Amer. Math. Soc. 347 (1995) 1865–1961 MR Zbl

[22] **J Ullman**, *On the slice spectral sequence*, Algebr. Geom. Topol. 13 (2013) 1743–1755 MR Zbl

[23] **S Waner**, *Periodicity in the cohomology of universal G-spaces*, Illinois J. Math. 30 (1986) 468–478 MR Zbl

[24] **C M Yarnall**, *The slices of $S^n \wedge H\underline{\mathbb{Z}}$ for cyclic p-groups*, PhD thesis, University of Virginia (2013) Available at https://www.proquest.com/docview/1433854798

[25] **C Yarnall**, *The slices of $S^n \wedge H\underline{\mathbb{Z}}$ for cyclic p-groups*, Homology Homotopy Appl. 19 (2017) 1–22 MR Zbl

[26] **M Zeng**, *Equivariant Eilenberg–Mac Lane spectra in cyclic p-groups*, preprint (2017) arXiv 1710.01769

[27] **Y Zou**, *$RO(D_{2p})$-graded slice spectral sequence for $H\underline{\mathbb{Z}}$*, PhD thesis, University of Rochester (2018) Available at http://hdl.handle.net/1802/34283

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

Rochester, NY, United States

bertguillou@uky.edu, carissa.slone4@gmail.com

https://www.ms.uky.edu/~guillou/, https://sites.google.com/view/carissaslone/home

Received: 19 January 2023 Revised: 25 April 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3385–3427
DOI: 10.2140/agt.2025.25.3385
Published: 1 October 2025

Cocycles of the space of long embeddings and BCR graphs with more than one loop

LEO YOSHIOKA

The purpose of this paper is to construct nontrivial cocycles of the space $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ of long embeddings. We construct the cocycles through configuration space integrals, associated with Bott–Cattaneo–Rossi graphs with more than one loop. As an application, we give a nontrivial cycle of long embeddings in the path component of the trivial long embedding, for odd n, j with $n - j \geq 2$ and $j \geq 3$. This cycle is constructed from a chord diagram on directed lines. The nontriviality is shown by cocycle-cycle pairing, described by pairing between graphs and chord diagrams.

57K16, 57K45, 57M15, 57R40

Introduction 3385
1. Background 3387
2. Definitions 3388
3. Main result 3397
4. Construction of the cocycles 3399
5. Review of Sakai and Watanabe's construction of the wheel-like cycles 3404
6. Construction of the cycles 3407
7. Some lemmas for computing cocycle-cycle pairing 3417
8. Proof of the main result 3423
References 3425

Introduction

A long embedding is an embedding of $\mathbb{R}^j$ into $\mathbb{R}^n$, which is standard outside a disk in $\mathbb{R}^j$. A long n-knot is a long embedding $\mathbb{R}^n \to \mathbb{R}^{n+2}$. In this paper, we produce some geometric nontrivial cocycles of the space $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ of long embeddings through integrals associated with graphs. This kind of approach, which we call configuration space integrals, has its quantum field theoretical origin in E Witten's work [39] and its mathematical formulation has been developed by M Kontsevich [18], D Bar-Natan [5], R Bott [6], C Taubes [7], G Kuperberg, D Thurston [19], A S Cattaneo and C A Rossi [11] and some others from the

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

1990s to 2000s. Recently, this approach attracted much attention thanks to T Watanabe [36], who in 2018 disproved the 4-dimensional Smale conjecture based on a series of his works [33; 34; 35].

Bott [6] and Cattaneo and Rossi [11] defined 1-loop graphs (graphs with $b_1 = 1$) with two types of edges and vertices. We call these graphs BCR graphs. Using specific linear combinations of BCR graphs called graph cocycles, they defined configuration space integral invariants for long n-knots. K Sakai and Watanabe [27; 29] generalized their construction and systematically gave nontrivial higher cocycles of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ for $n - j \geq 3$ from BCR graph cocycles.

The graphs used in this paper are similar to theirs, but our graphs have more than one loop ($b_1 \geq 2$). Using such higher-loop graphs plays an essential role in getting cocycles of higher degree when the codimension $n - j$ is exactly two. The framework of applying higher-loop BCR graphs was already considered by Sakai [27]. However, no nontrivial cocycle corresponding to higher-loop graphs has been given so far.

In this paper, we show the following. (See Section 3.)

Theorem 0.1 *Let n, j be odd and satisfy $n - j \geq 2$ and $j \geq 3$. Then the linear combination H of 2-loop BCR graphs gives an explicit nontrivial $3n - 2j - 7$ (co)cycle of $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$, the space of long embeddings modulo immersions.*

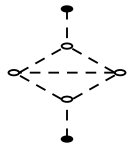
This theorem is shown by cocycle-cycle pairing. We systematically construct cycles of the space of long embeddings, which we call generalized ribbon cycles, from chord diagrams on directed lines. This diagrammatic construction makes it possible to compute cocycle-cycle pairing through pairing between graphs and diagrams. This diagrammatic computation is analogous to the diagrammatic description of cohomology-homology pairing of configuration spaces that D P Sinha introduced in [31].

As a corollary, we give a nontrivial S^2-family of trivial long 3-knots. Budney and Gabai [9] and Watanabe [37] already showed that the $(n-1)^{\mathrm{st}}$ homotopy group of the path component of the trivial long n-knots, $\pi_{n-1}(\mathrm{Emb}(\mathbb{R}^n, \mathbb{R}^{n+2})_\iota)$ ($n \geq 2$), has an infinite-rank subgroup, by different approaches from ours. However, their families of long n-knots are constructed through $\mathrm{Diff}_\partial(D^{n+1} \times S^1)$ or $\mathrm{Emb}_\partial(D^1, D^{n+1} \times S^1)$. It would be interesting to compare our construction of families with theirs. Note that our approach has the possibility to produce an infinite-rank subgroup of $\pi_{n-1}(\mathrm{Emb}(\mathbb{R}^n, \mathbb{R}^{n+2})_\iota)$ using more general 2-loop graphs.

Moreover, our graph cocycles with more than one loop further support the interesting similarity between the BCR graph complex and the graph complex introduced by G Arone and V Turchin [2; 3]. Their complex arises from a homotopy theoretical approach to embedding spaces: embedding calculus developed by Goodwillie, Klein and Weiss [14; 15; 38]. The significant point of their complex is that its graph homology has the full information about the dimension of the rational homology of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ for $n \geq 2j + 2$.[1] They mentioned in [3] that some of their graphs are very similar to BCR graphs with no

[1] In [13], Fresse, Turchin and Willwacher generalized the results of [3] to $n - j \geq 3$. Note that their approach consists of the Goodwillie–Weiss embedding calculus tower and its mapping space model. The latter is applicable to any codimension, though the former works for $n - j \geq 3$.

more than one loop. The point there was that Arone and Turchin's graph complex allows graphs with any number of loops. The 2-loop graph cocycle we give in this paper has the 2-loop graph



similar to a graph which appears as a nontrivial element in [3]. Since BCR graph cohomology can give information for any codimension more than one, we would be able to say that the stability of high codimensions that is described by this graph also survives in the range $n - j = 2$.

Although we can apply our framework to the case when n and j are even, we have got no nontrivial cocycles of the space of long embeddings. The main cause is that we have not got a nontrivial graph cocycle for n, j even. But one will be able to guess nontrivial cocycles from the computation of the 2-loop part of Arone and Turchin's graph homology [12]. Note that the graph homology is infinite dimensional.

The paper is organized as follows. First, in Section 1, we review the framework of configuration space integrals associated with BCR graphs. After that, some definitions and notation are given in Section 2. The main result is stated in Section 3. Then in Section 4, we construct the well-defined cocycles of $\mathrm{Emb}(\mathbb{R}^n, \mathbb{R}^{n+2})$. After reviewing Sakai and Watanabe's construction of cycles in Section 5, we construct cycles from chord diagrams on directed lines in Section 6. In Section 7, we show preliminary lemmas used in the proof of the main result. Finally, in Section 8, we give the proof of the main result by pairing the cocycles and the cycles of Sections 4 and 6, described by graph-diagram pairing.

Acknowledgements

The author is deeply grateful to the author's supervisor, Mikio Furuta, for his continuous support and encouragement for years. The author would like to thank Tadayuki Watanabe for explaining much of his work; discussion with Watanabe was highly motivating during the preparation of this paper. The author thanks Keiichi Sakai for the great opportunity to talk and discuss at the Shinshu Topology Seminar. Sakai also taught the author a lot about embedding spaces. The author also thanks Masahito Yamazaki and Tatsuro Shimizu for helpful suggestions and stimulating conversations. The author thanks Victor Turchin for sharing the papers [12; 13] and for advice for better exposition.

This research was supported by the Forefront Physics and Mathematics Program to Drive Transformation (FoPM), a World-leading Innovative Graduate Study (WINGS) program, at the University of Tokyo.

1 Background

1.1 The BCR graph complex and the de Rham complex of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$

Sakai [27] defined the BCR graph complex $(\mathcal{D}_g^{k,l}, \delta)$ and constructed a linear map from this space to the de Rham complex of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ through configuration space integrals. This correspondence

gives a higher dimensional analog (in the sense $j \geq 2$) of the correspondence developed by Cattaneo, Cotta-Ramusino and Longoni [10]. Refer to Sections 2.2 and 2.3 for the definition of the BCR graph complex (and the definitions of k, g and l).

Theorem 1.1 [27, Theorem 1.2] *Let $k \geq 1$ and $g \geq 0$. Configuration space integrals give a linear map*

$$I : \mathcal{D}_g^{k,l} \to \Omega_{dR}^{k(n-j-2)+(g-1)(j-1)+l}(\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)).$$

Moreover, if either

- *$n-j$ is even, $n-j \geq 2$, $j \geq 2$ and $g = 0$, or*
- *n and j are odd, $n > j \geq 3$ and $g = 1$,*

then the map I gives a cochain map.

Based on the above theorem, it is expected that we can obtain nontrivial cohomology classes of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ if the graph cohomology is nontrivial and the map on the cohomology level induced from I satisfies injectivity.

From this perspective, Sakai and Watanabe conducted further research on $g = 1$ (and $l = 0$) in [29], where configuration space integral invariants are put together into the universal one. Their work is based on the theory of finite type invariants for ribbon n-knots developed by Habiro, Kanenobu and Shima [17; 16].

Theorem 1.2 [29, Theorem 1.1] *Let (n, j, k) satisfy some conditions. Let $\mathcal{A}_1^k$ be the quotient of the space of 1-loop graphs by some relations (such as the STU relations of Jacobi diagrams). Then there exists a linear and injective map*

$$\alpha : \mathcal{A}_1^k \to H_{(n-j-2)k}(\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n), \mathbb{R}).$$

In fact, the universal configuration space integral invariant z_k gives a closed $(n-j-2)k$ form, and detects the nontriviality of the above cycles. We review the construction of the cycles in Section 5.

Remark 1.3 In Theorem 1.1, we can observe that when $n - j = 2$ we have to use graphs with $g \geq 2$ or $l \geq 1$ to obtain cocycles of higher degree.

Remark 1.4 The case $l \geq 1$ is out of the scope of this paper. For the case $j = 1$, some progress has been made by Longoni [23], Sakai [26; 28], and Pelatt and Sinha [25]. In these works, new types of cycles are constructed generalizing the case $l = 0$.

2 Definitions

2.1 The space of long embeddings

First, We define the space of long embeddings and some related spaces.

Definition 2.1 A long embedding is an embedding $\mathbb{R}^j \to \mathbb{R}^n$ which coincides with the standard linear embedding $\iota\colon \mathbb{R}^j \subset \mathbb{R}^n$ outside a disk in $\mathbb{R}^j$. A long embedding $\mathbb{R}^n \to \mathbb{R}^{n+2}$ is called a long n-knot. We equip the space $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ of long embeddings with the induced topology from the weak C^∞ topology. We define the space of long immersions $\mathrm{Imm}(\mathbb{R}^j, \mathbb{R}^n)$ similarly.

Remark 2.2 Budney showed in [8] that $\mathrm{Emb}(S^j, S^n) \simeq \mathrm{SO}_{n+1} \times_{\mathrm{SO}_{n-j}} \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$. That is, the difference between the two spaces $\mathrm{Emb}(S^j, S^n)$ and $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ is given by the Stiefel manifold $\mathrm{SO}_{n+1}/\mathrm{SO}_{n-j}$, whose homotopy groups are well-studied.

Remark 2.3 We often consider $\mathrm{Emb}_\partial(D^j, D^n)$, the space of embeddings $D^j \to D^n$ which are standard near the boundary, instead of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$. The homotopy types of these two spaces are equivalent.

For technical reasons, sometimes we think of the space of embeddings modulo immersions.

Definition 2.4 We write $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ for the homotopy fiber of the inclusion

$$\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n) \hookrightarrow \mathrm{Imm}(\mathbb{R}^j, \mathbb{R}^n)$$

at the standard linear embedding $\iota\colon \mathbb{R}^j \subset \mathbb{R}^n$. That is, an element of $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ is a one-parameter family $\{\overline{K}_t\}_{t\in[0,1]}$ of long immersions which satisfies $\overline{K}_0 = \iota$, $\overline{K}_1 \in \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$. Let

$$r\colon \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$$

be the natural projection.

Notation 2.5 Write $\iota\colon \mathbb{R}^j \to \mathbb{R}^n$ for the standard linear embedding. The path component of $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ which ι belongs to is denoted by $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)_\iota$. We often call this component the unknot component. Write $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)_\iota$ for the component of the trivial family of the standard linear immersion.

2.2 BCR graphs

We move on to the definition of BCR graphs and their complex. See Section 3 for examples of BCR graphs. Although we basically follow the definitions of [6; 11; 27; 29; 32], we restate them since there are some differences in conventions among these papers.

Definition 2.6 A BCR graph is a graph which satisfies the following conditions.

(1) There are two types of vertices, white and black.

(2) There are two types of edges, dashed and solid.

(3) Each white vertex has only dashed edges, and its valency is at least three. Each black vertex has at least one dashed edge and has an arbitrary number of solid edges.

(4) White vertices have no small loop:

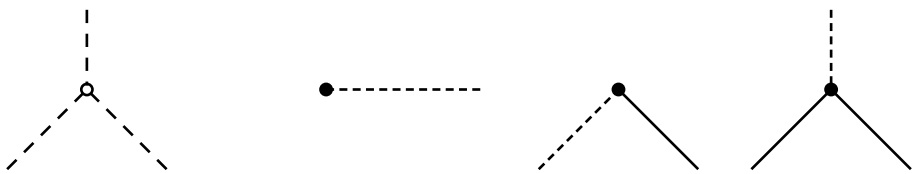


Black vertices have no small loop consisting of a solid edge only:



(See Remark 2.8 below.)

Example 2.7 Here are typical examples of vertices of BCR graphs:



Remark 2.8 Black vertices are allowed to have small loops:



and double loops:



We write the set of small loops of Γ as $L_s(\Gamma)$ and the set of double loops as $L_d(\Gamma)$. Set $L(\Gamma) = L_s(\Gamma) \sqcup L_d(\Gamma)$. In our case ($n$ and j are odd), double loops vanish by the third relation in Definition 2.19.

Notation 2.9 Let Γ be a BCR graph. We denote the set of solid edges of Γ by $E_\eta(\Gamma)$, and the set of dashed edges by $E_\theta(\Gamma)$. Edges of small loops and double loops are included in these sets. The set of edges $E(\Gamma)$ is decomposed as

$$E(\Gamma) = E_\eta(\Gamma) \sqcup E_\theta(\Gamma).$$

We denote the set of black vertices of Γ by $B(\Gamma)$, and the set of white vertices by $W(\Gamma)$. The set of vertices $V(\Gamma)$ is decomposed as

$$V(\Gamma) = B(\Gamma) \sqcup W(\Gamma).$$

Definition 2.10 (defect of a graph) Let v be a vertex of a BCR graph Γ. We define the defect $l(v)$ of v as

$$l(v) = \begin{cases} \#E_\theta(v) - 3 & \text{if } v \in W(\Gamma), \\ \#E_\theta(v) - 1 & \text{if } v \in B(\Gamma). \end{cases}$$

Here $E_\theta(v)$ is the number of dashed edges v has. The defect $l(\Gamma)$ of Γ is defined as

$$l(\Gamma) = \sum_{v \in V(\Gamma)} l(v),$$

and is equal to

$$2\#E_\theta(\Gamma) - 3\#W(\Gamma) - \#B(\Gamma).$$

Remark 2.11 The defect of a graph measures how the graph is different, in terms of edge contractions, from graphs in which each black vertex has exactly one dashed edge and each white vertex has exactly three dashed edges.

Definition 2.12 (order of a graph) The order $k(\Gamma)$ of a BCR graph Γ is given by

$$k(\Gamma) = \#E_\theta(\Gamma) - \#W(\Gamma).$$

Remark 2.13 When the defect is 0, we have $k = \frac{1}{2}\#V(\Gamma)$.

Definition 2.14 (g-loop graphs) A BCR graph is a g-loop graph if the first Betti number of any component is g. Here, we consider small and double loops as points, and we do not count them as loops.

Definition 2.15 (labels of a graph) We give a label of Γ by an ordering of $V(\Gamma)$, an ordering of $E(\Gamma)$, an orientation of each edge $e \in E(\Gamma)$ and an ordering of $L_s(\Gamma)$. Here, an ordering of a set X with m elements is a bijective map from $\{1, 2, \dots, m\}$ to X. We choose an ordering so that black vertices and solid edges come first.

Definition 2.16 (orientations of a graph) An orientation s of a graph Γ is a choice

$$s \in \det\left(\mathbb{R}V(\Gamma) \oplus \bigoplus_{e \in E(\Gamma)} \mathbb{R}H(e)\right).$$

Here $H(e)$ is the set of two half-edges of e. A labeled graph determines an orientation of the underlying graph.

Remark 2.17 In this paper, we only handle the case n and j are odd. The above definitions of labels and orientations are only for this case. We can define orientation of graphs so that the definition depends only on the parities of n and j.

Remark 2.18 An ordering of $L_s(\Gamma)$ is used to define the configuration space integral associated with Γ. However, since we define configuration space integrals for graphs of defect 0, we do not use this data in the rest of this paper.

2.3 The BCR graph complex

Definition 2.19 We define relations of the vector space spanned by labeled BCR graphs of order k and defect l. We denote the quotient space by $\mathcal{D}^{k,l}$.

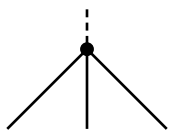
(1) Two labels with the same orientation are identified, and ones with opposite orientations are given opposite signs.

(2) $\Gamma = 0$ for any graph Γ with a double edge:




(3) $\Gamma = 0$ for any graph Γ with a double loop. (On the other hand, a graph with small loops may survive.)

(4) When $l = 0$, we set $\Gamma = 0$ for any "nonadmissible" graph Γ. A graph Γ is nonadmissible
   (i) if it has a black vertex whose eta-valency is at least three, such as



   (ii) if it has a multiple edge



Notation 2.20 We write the subspace of $\mathcal{D}^{k,l}$ generated by graphs with g loops as $\mathcal{D}_g^{k,l}$. (Here, we count neither small nor double loops as loops.)

Remark 2.21 The fourth condition is just to make the computation of the top cohomology easier, and it follows the conventions of [29]. So far, the author has yet to find precise reasons to exclude the possibility that we can construct a graph cocycle (of defect 0) essentially using these nonadmissible graphs. But one possible reason is that we can change nonadmissible graphs to admissible graphs by the Arnold relation (see Section 2.4) with respect to solid edges.

Remark 2.22 One possible variant of $\mathcal{D}$ is the graph complex $\overline{\mathcal{D}}$ that is the same as $\mathcal{D}$ for $l \geq 0$ but has defect (-1) part $\overline{\mathcal{D}}^{-1}$. Here a graph with defect (-1) is a graph that has vertices of defect 0 and exactly one black vertex with three solid edges and no dashed edge. Contractions of the three solid edges adjacent to the black vertex yield the Arnold relation.

Next, we define the coboundary map $\delta\colon \mathcal{D}_g^{k,l} \to \mathcal{D}_g^{k,l+1}$ of the graph complex.

Definition 2.23 (contraction of a BCR graph [27]) We define the contraction Γ/e of a BCR graph Γ at an edge e as follows.

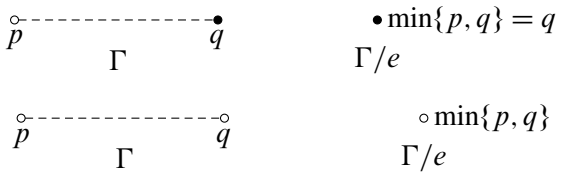
(a) When $e = (p, q) \in E_\eta(\Gamma)$ (except multiple edges).

Collapse e to the black vertex at the ends of e with a smaller label, and then reassign labels to vertices of the collapsed graph in a natural way (that is, assign $\min\{p, q\}$ to this collapsed vertex, and subtract 1 from all vertex-labels greater than $\max\{p, q\}$):



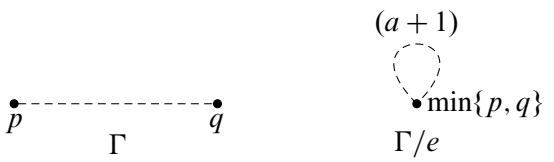
(b) When $e=(p,q)\in E_\theta(\Gamma)$ and at least one of p and q is white.

If exactly one is white, collapse e to the black vertex. If both are white, collapse e to the white vertex with a smaller label. The way to reassign labels is similar to (a):



(c) When $e=(p,q)\in E_\theta(\Gamma)$ and both p and q are black vertices.

We construct a small loop.[2] If the number of small loops of Γ is a, the label of this loop is $(a+1)$:



(d) When $e=(p,q)$ is the solid edge of a multiple edge.

We construct a double loop, which vanishes by the third relation in Definition 2.19. We do not perform contractions of dashed edges of multiple edges.

Definition 2.24 (coboundary map [27]) We define the coboundary map $\delta\colon \mathcal{D}_g^{k,l}\to\mathcal{D}_g^{k,l+1}$ of $\mathcal{D}_g^{k,l}$ as

$$\delta(\Gamma)=\sum_{e\in E(\Gamma)\setminus\{\text{loops}\}}\sigma(e)\Gamma/e.$$

Here, the sign $\sigma(e)$ of an edge e is defined as

$$\sigma(e=(p,q))=\begin{cases}(-1)^q & \text{if } p<q,\\ (-1)^{p+1} & \text{if } p>q.\end{cases}$$

Proposition 2.25 *$(\mathcal{D}_g^{k,*},\delta)$ is a cochain complex.*

Proof This follows from the fact that reversing the order of two contractions changes the sign. □

2.4 Comparing with the Arone–Turchin graph complex

Arone and Turchin [2; 3] defined two graph complexes $\mathcal{C}^{j,n}$ and $HH^{j,n}$ (j and n come from $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$). These complexes both arise from a homotopy theoretical approach: the description of the space of long embeddings as a space of derived maps between modules over an operad.

[2] Though [27] introduces sign of small loops, it seems unnecessary.

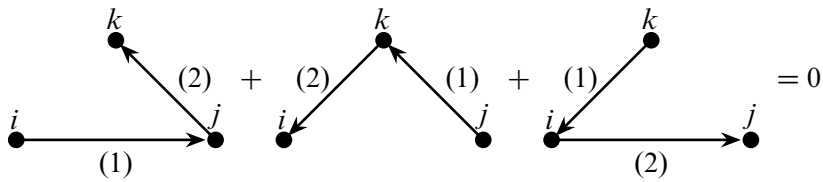
The complex $\mathscr{E}^{j,n}$ consists of so-called hairy graphs with two types of vertices and with one type of edges. Graphs of $HH^{j,n}$ have two types of edges as graphs of $\mathscr{D}$. The differences between $HH^{j,n}$ and our complex $\mathscr{D} = \mathscr{D}^{j,n}$ are:

(1) Their graphs have only black vertices.

(2) Their graphs do not have cycles that consist of one type of edges. In particular, there are no small loops, double loops or double edges.

(3) Their coboundary map is the sum of contractions of only solid edges.

(4) Their complex has another relation called the Arnold relation [1] as follows (the edges in the figure must be the same type, either dashed or solid):



Note that our convention for solid and dashed edges are opposite from [2; 3]: we follow the convention that solid edges correspond to $\mathbb{R}^j$ of the domain.

2.5 Configuration space integrals

From now on, for simplicity, we assume that graphs used to define configuration space integrals are connected and satisfy $l = 0$.

Notation 2.26 We write $C_s(\mathbb{R}^n)$ for the configuration space of s points in $\mathbb{R}^n$:

$$C_s(\mathbb{R}^n) = (\mathbb{R}^n)^s \setminus \text{all diagonals}.$$

Definition 2.27 (configuration space) Let $K\colon \mathbb{R}^j \to \mathbb{R}^n$ be a long embedding. Then we define the configuration space $C_{s,t}(K, \mathbb{R}^n)$ as

$$C_{s,t}(K, \mathbb{R}^n) = \{(x_1, x_2, \dots, x_s, x_{s+1}, \dots, x_{s+t}) \in (\mathbb{R}^n)^{s+t} \setminus \text{all diagonals} \mid \forall i \in \{1, \dots, s\}, x_i \in K\}.$$

Hence $C_{s,t}(K, \mathbb{R}^n)$ is the pullback of the following diagram:

$$\begin{array}{ccc} C_{s,t}(K, \mathbb{R}^n) & \longrightarrow & C_{s+t}(\mathbb{R}^n) \\ \downarrow & & \downarrow \text{restriction} \\ C_s(\mathbb{R}^j) & \xrightarrow{K^{\times s}} & C_s(\mathbb{R}^n) \end{array}$$

Definition 2.28 (direction map) Let Γ be a labeled BCR graph, $s = \#B(\Gamma)$ and $t = \#W(\Gamma)$. Let $e = (i, j)$ be an edge of Γ. Depending on if $e \in E_\theta(\Gamma)$ or $e \in E_\eta(\Gamma)$, the direction map $\phi_e = \phi_{ij}$ from $C_{s,t}(K, \mathbb{R}^n)$ to a sphere is defined as follows:

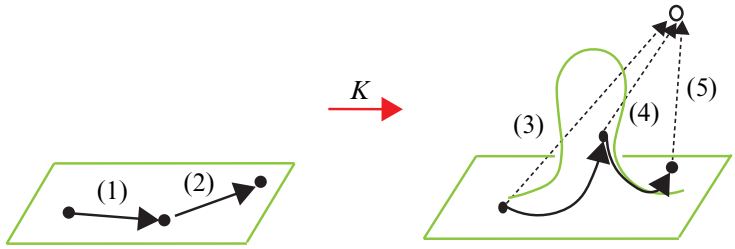


Figure 1: Graph and its direction maps.

- If $e=(i,j)\in E_\theta(\Gamma)$,
$$\phi_{ij}: C_{s,t}(K,\mathbb{R}^n)\to C_2(\mathbb{R}^n)=C_{0,2}(\varnothing,\mathbb{R}^n)\to S^{n-1}.$$
- If $e=(i,j)\in E_\eta(\Gamma)$,
$$\phi_{ij}: C_{s,t}(K,\mathbb{R}^n)\to C_2(\mathbb{R}^j)=C_{2,0}(K,\mathbb{R}^n)\to S^{j-1}.$$

In both cases, the first map is the restriction

$$(x_1,x_2,\dots,x_s,x_{s+1},\dots,x_{s+t})\mapsto(x_i,x_j),$$

and the second is defined as

$$(x,y)\mapsto\frac{y-x}{\|y-x\|}.$$

Intuitively, dashed edges measure the direction between two points in $\mathbb{R}^n$, while solid edges measure the direction in $\mathbb{R}^j$.

Definition 2.29 (propagator) We define the differential form $\omega(K,\Gamma)$ on $C_{s,t}(K,\mathbb{R}^n)$ of degree $(n-1)\#E_\theta(\Gamma)+(j-1)\#E_\eta(\Gamma)$ by

$$\omega(K,\Gamma)=\bigwedge_{e\in E_\eta(\Gamma)}\phi_e^*\omega_{S^{j-1}}\bigwedge_{e\in E_\theta(\Gamma)}\phi_e^*\omega_{S^{n-1}}.$$

Here, the order of the wedge products is arbitrary, since both $\omega_{S^{j-1}}$ and $\omega_{S^{n-1}}$ are even forms when n and j are odd. We can use the order of edges, for example.

Remark 2.30 The forms $\phi_e^*\omega_{S^{j-1}}$ and $\phi_e^*\omega_{S^{n-1}}$ are often called propagators. Leturcq [22] has given more flexible propagators to define invariants of embeddings into rational homology n-spheres.

The next proposition ensures the convergence of integrals we later define.

Proposition 2.31 [4; 7, Proposition 1.2; 30] *There exists a canonical compactification $\overline{C}_{s,t}(K,\mathbb{R}^n)$ of $C_{s,t}(K,\mathbb{R}^n)$ such that ϕ_{ij} can be extended to the whole $\overline{C}_{s,t}(K,\mathbb{R}^n)$. The codimension 1 part of $\partial\overline{C}_{s,t}(K,\mathbb{R}^n)$ is*

$$\bigsqcup_{A\subset\{1,\dots,s,s+1,\dots,s+t,\infty\}}\widetilde{C}_A(K,\mathbb{R}^n),$$

where $\widetilde{C}_A(K,\mathbb{R}^n)$ is "the configuration space with A being infinitely close". (See Section 4.3 for details.)

Definition 2.32 We classify $\widetilde{C}_A(K,\mathbb{R}^n)$ ($A\subset\{1,\dots,s,s+1,\dots,s+t,\infty\}$ and $\#A\geq 2$) into four types as follows (Γ_A is the subgraph of Γ generated by vertices of A).

- **Infinite faces** $\infty\in A$.
- **Principal faces** $\#A=2$.
- **Hidden faces** $\#A\geq 3$ and Γ_A is not the whole Γ; $\Gamma_A\subsetneq\Gamma$.
- **Anomalous faces** $\#A\geq 3$ and Γ_A is the whole Γ; $\Gamma_A=\Gamma$.

Notation 2.33 We write $E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)$ for the bundle over $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ whose fiber at a long embedding K is the configuration space $C_{s,t}(K,\mathbb{R}^n)$. More precisely, $E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)$ is the pullback of the following diagram:

$$\begin{array}{ccc} E_{s,t}(\mathbb{R}^j,\mathbb{R}^n) & \longrightarrow & C_{s+t}(\mathbb{R}^n) \\ \downarrow & & \downarrow{\scriptstyle \text{restriction}} \\ C_s(\mathbb{R}^j)\times\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n) & \xrightarrow{\text{evaluation}} & C_s(\mathbb{R}^n) \end{array}$$

By abusing notation, we also write $E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)$ for the bundle whose fiber is $\overline{C}_{s,t}(K,\mathbb{R}^n)$. We write $\widetilde{E}_A(\mathbb{R}^j,\mathbb{R}^n)$ for the bundle over $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ whose fiber at a long embedding K is the configuration space $\widetilde{C}_A(K,\mathbb{R}^n)$.

As in Definition 2.28, we can construct the direction map

$$\phi_e\colon E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)\to S^{n-1}\quad\text{or}\quad \phi_e\colon E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)\to S^{j-1}$$

depending on the type (solid or dashed) of e.

Notation 2.34 We write $\omega(\Gamma)$ for the form on $E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)$ defined by pulling back the standard volume forms $\omega_{S^{n-1}}$ and $\omega_{S^{j-1}}$ (in the same way as Definition 2.29).

Definition 2.35 (configuration space integrals) Let $C_{s,t}$ (and $\overline{C}_{s,t}$) be a typical fiber of the bundle $\pi\colon E_{s,t}(\mathbb{R}^j,\mathbb{R}^n)\to\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$. Let Γ be a labeled BCR graph. We define the configuration space integral associated with Γ by the fiber integral

$$I(\Gamma)=\int_{\overline{C}_{s,t}}\omega(\Gamma)=\int_{C_{s,t}}\omega(\Gamma).$$

The form $I(\Gamma)$ on $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ is sometimes written as $\pi_*\omega(\Gamma)$.

By Stokes' theorem (see [7, Appendix] and [32, Appendix A.2]), we have

$$\begin{aligned} dI(\Gamma)&=\pi_*d\omega(\Gamma)+(-1)^r\pi_*^{\partial}\omega(\Gamma)\\ &=\int_{\overline{C}_{s,t}}d\omega(\Gamma)+(-1)^r\int_{\partial\overline{C}_{s,t}}\omega(\Gamma)=(-1)^r\int_{\partial\overline{C}_{s,t}}\omega(\Gamma)=(-1)^r\sum_A\int_{\widetilde{C}_A}\omega(\Gamma)\end{aligned}$$

where r is the degree of $\int_{\partial \overline{C}_{s,t}} \omega(\Gamma)$. Hence for the closedness of $I(\Gamma)$, we should check that the integral over each $\widetilde{C}_A$ vanishes or is canceled.

Contributions of infinite faces vanish for dimensional reasons. (Recall that all our embeddings are standard outside a disk.)

Theorem 2.36 [27, Theorem 5.10, Lemma 5.18] *If $\infty \in A$, the integral $\int_{\widetilde{C}_A} \omega(\Gamma)$ vanishes.*

On the other hand, contributions of principal faces are canceled when we sum up the configuration space integrals for graphs in a graph cocycle (with their coefficients).

Theorem 2.37 [27, Theorem 5.11] *Let $\#A = 2$. The integral $\int_{\widetilde{C}_A} \omega(\Gamma)$ survives only when $A = e$ for some edge e of Γ. Moreover,*

$$\int_{\widetilde{C}_e} \omega(\Gamma) = -\int_{C_{V(\Gamma/e)}} \omega(\delta_e \Gamma).^{3}$$

Here $C_{V(\Gamma/e)}$ is the configuration space of vertices of Γ/e.

3 Main result

Recall $r : \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ is the natural projection defined in Definition 2.4.

Theorem 3.1 *Let $(n, j) = (\text{odd}, \text{odd})$ with $j \geq 3$.*

(1) *The linear combination of 2-loop BCR graphs of order 3*

$$H = \sum_{i=1}^{21} w(\Gamma_i)\Gamma_i$$

is a graph cocycle. For the definition of Γ_i and its coefficient $w(\Gamma_i)$, see Figure 2.

(2) *Let $\bar{c}(H)$ be the correction term defined in Definition 4.6. Then*

$$\bar{z}_2^3 = r^* I(H) + \bar{c}(H) \in \Omega_{dR}^{3n-2j-7}\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$$

is closed and $[\bar{z}_2^3] \in H_{dR}^{3n-2j-7}\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ is nontrivial. In fact we can explicitly construct $\psi \in H_{3n-2j-7}\,\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ and its lift $\bar{\psi}$, so that the pairing $\bar{z}_2^3(\bar{\psi})$ is nontrivial. See Figure 10 for ψ, which is constructed from the presentation in Figure 13.

(3) *Moreover, if $H_{dR}^{3n-j-6}(V_{n,j}(\mathbb{R})) = 0$, there is a nontrivial element $z_2^3 \in H_{dR}^{3n-2j-7}(\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n))$. Here $V_{n,j} = V_{n,j}(\mathbb{R})$ is the real Stiefel manifold $\mathrm{SO}(n)/\mathrm{SO}(n-j)$.*

Remark 3.2 [24, Theorem 3.14] When $(n, j) = (\text{odd}, \text{odd})$,

$$H_{dR}^*(V_{n,j}) \cong \bigwedge(e_{n-j}, e_{2(n-j)+3}, e_{2(n-j)+7}, \ldots, e_{2(n-j)+2j-3}).$$

In particular, if $j = 3$ then $H_{dR}^{3n-j-6}(V_{n,j}(\mathbb{R})) = 0$.

[3] We follow the inward normal convention for orientations of boundaries, which causes the sign of the right-hand side.

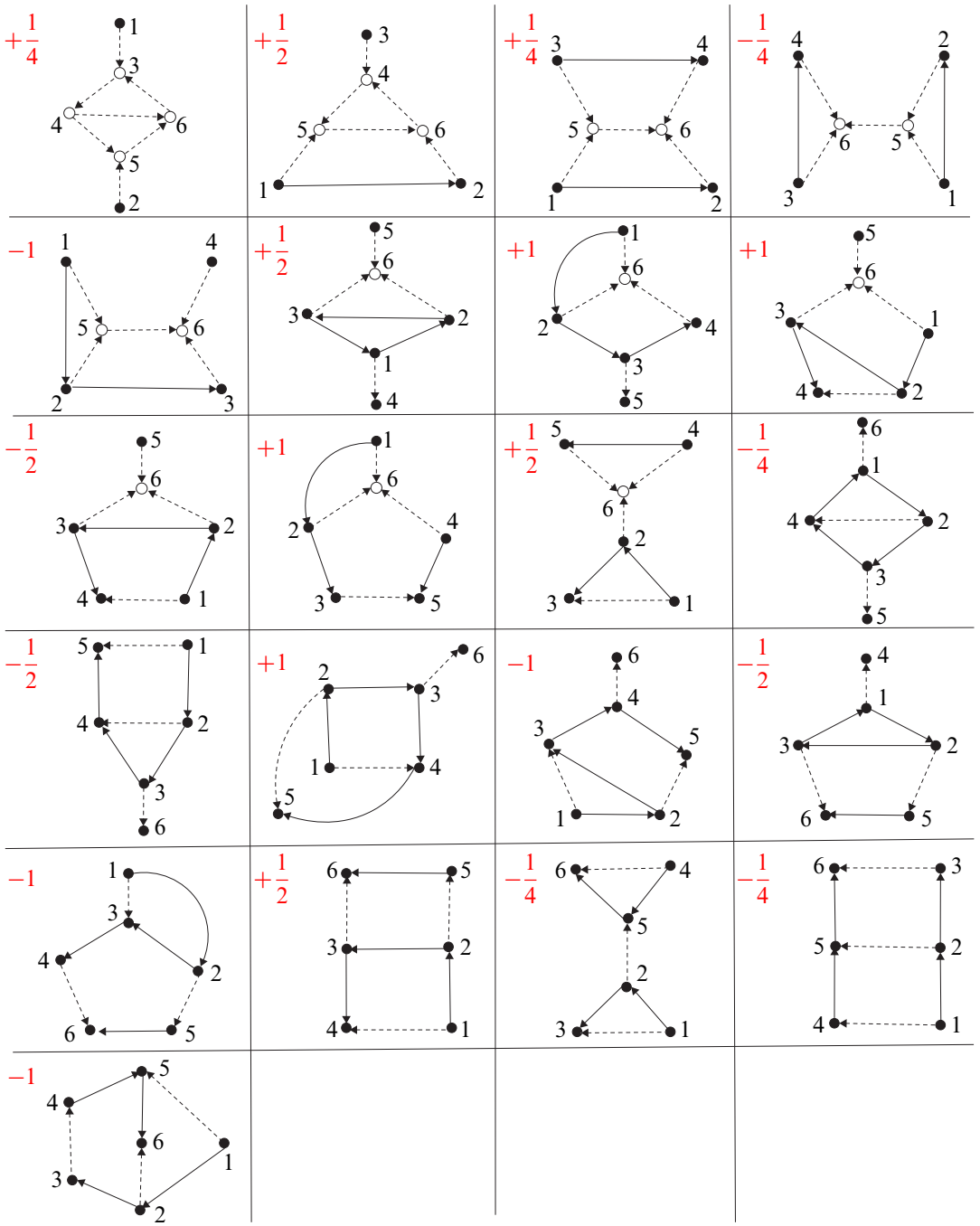


Figure 2: Nontrivial graph cocycle $H(\text{odd}, \text{odd})$.

Remark 3.3 A similar graph to the first graph Γ_1 of Figure 2 appears in [3, Table C] as one of the rational homology generators of $\mathscr{E}^{m,n}$, the graph complex of hairy graphs. One can also check that $HH^{m,n}$ has the last graph as one of the rational generators.

We can take the above nontrivial cycle ψ from the unknot component, and also over the 2-sphere when the codimension is 2.

Corollary 3.4 $\pi_2(\mathrm{Emb}(\mathbb{R}^3, \mathbb{R}^5)_\iota) \otimes \mathbb{Q} \neq 0.$

4 Construction of the cocycles

In this section, we show $r^*I(H) + \bar{c}(H)$ is closed. According to the paragraphs after Definition 2.35, what to show is the vanishing of contributions of principal, hidden and anomalous faces.

4.1 H is a graph cocycle

We can show that H is a graph cocycle, by direct computation of the coboundary map δ (defined in Definition 2.24). In this computation, we need to put together isomorphic graphs with different labels, taking their orientations into account. This will be the most difficult part when computing the graph cohomology through computer calculus.

Note that for the top degree, δ is described by a matrix (see Figure 3). We can easily see that the matrix is decomposed into blocks depending on the type of edges contracted:

(A) ○- - -○

(B) ●- - -●

(C) ○- - -●

(D) ●——●

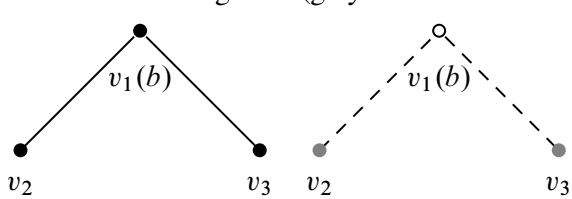
4.2 Vanishing of contributions of hidden faces

Here, we show that contributions of hidden faces vanish. The fundamental approach is the following proposition. The cancellation using the symmetry as below was first introduced by Kontsevich [18]. Lescop gave a detailed explanation of it in [20; 21].

Proposition 4.1 *Let $s = \#B(\Gamma)$, $t = \#W(\Gamma)$ and $\widetilde{C}_A \subset \partial C_{s,t}(K, \mathbb{R}^n)$ be a hidden face. If the subgraph Γ_A of Γ satisfies one of the following conditions, the contribution of the integral over $\widetilde{C}_A$ vanishes. (Below, u and b stand for univalent and bivalent, respectively.)*

(1) *Γ_A is not connected.*

(2) *Γ_A has a bivalent vertex of the following form (gray vertices can be both black and white):*



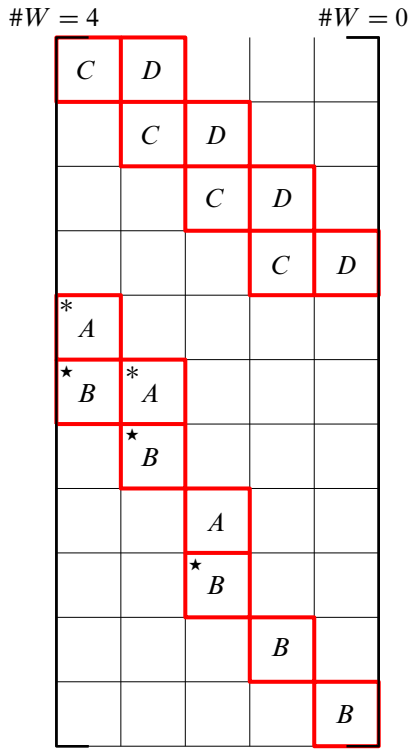
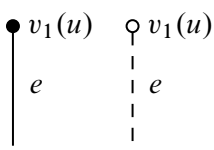


Figure 3: An example of the matrix of δ. For our cocycle H, the blocks with $*$ are 0. The blocks with $\star$ do not appear.

(3) Γ_A *has a univalent vertex of the following form*:



Proof For (2), we use the involution of v_1 with respect to the middle point of v_2 and v_3:

$$v_1 \mapsto v_2 + v_3 - v_1.$$

When $(n, j) = (\text{odd}, \text{odd})$, this involution reverses the orientation, while the integrand form does not change.

For (3), the group $\mathbb{R}_+$ acts on $\widetilde{C}_A$ by rescaling the edge e. Since the form $\omega(\Gamma)$ can be constructed through the projection

$$\widetilde{C}_A \to \widetilde{C}_A/\mathbb{R}_+,$$

where the quotient space on the right-hand side has less dimension, we can see the integral over $\widetilde{C}_A$ is zero.

For (1), we can use the action of $\mathbb{R}_+$ that increases or decreases the distance between two components. □

Unfortunately, our graph cocycle H has hidden faces that satisfy none of the above conditions. However, since $(n, j) = (\text{odd}, \text{odd})$ we can apply the following proposition that was shown in [27].

Proposition 4.2
- *If $n - j =$ even and Γ_A is a 0-loop (namely, $b_1(\Gamma_A) = 0$), the contribution of $\widetilde{C}_A$ vanishes.*
- *If $(n, j) = (\text{odd}, \text{odd})$ and Γ_A is a 1-loop (namely, $b_1(\Gamma_A) = 1$), the contribution of $\widetilde{C}_A$ vanishes.*

Remark 4.3 The second vanishing argument is also useful to cancel the contribution of the anomalous face, when $(n, j) = (\text{odd}, \text{odd})$ and the first Betti number of the graph is odd.

Proof The first statement is a consequence of similar involutions and rescaling to those of Proposition 4.1. We review the involution used to prove the second statement. This is an involution of all the vertices of A with respect to one black vertex of A:

$$(x_1, x_2, \dots, x_s, x_{s+1}, x_{s+2}, \dots, x_{s+t}) \mapsto (x_1, 2x_1 - x_2, \dots, 2x_1 - x_s, 2\iota(x_1) - x_{s+1}, 2\iota(x_1) - x_{s+2}, \dots, 2\iota(x_1) - x_{s+t}).$$

Here, $\iota\colon \mathbb{R}^j \to \mathbb{R}^n$ is the linear injective map defined by the differential of the embedding. (If all the vertices of A are white, take a white vertex as the central vertex x_1.) □

If a subgraph Γ_A of Γ_i in the cocycle H is 2-loop, it satisfies one of (1), (2) or (3). Hence, it follows that all the contributions of hidden faces of H do vanish.

4.3 Construction of correction terms for anomalous faces

We now proceed to construct a correction term to cancel the contribution of the integral over the anomalous face $\widetilde{C}_{V(\Gamma)}$. If $H^*_{dR}(V_{n,j}(\mathbb{R}))$ vanishes on a certain degree, we can construct a correction term c as a form on $\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$. Otherwise we construct one on $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$, the space of long embeddings modulo immersions. Sakai in [27] follows the first approach, while Sakai and Watanabe do the second one in [29].

For the above purpose, we first describe the structure of each codimension one face $\widetilde{C}_A$. Readers can also refer to [6; 10; 27] for more detailed explanations. We only describe it when the subset $A \subset V(\Gamma) = \{1, \dots, s, s+1, \dots, s+t\}$ has at least one element of $\{1, \dots, s\}$. The anomalous face is the case where A is equal to the entire $V(\Gamma)$.

The space $\widetilde{E}_A(\mathbb{R}^j, \mathbb{R}^n)$ is defined as the pullback of the following diagram. The configuration space $\widetilde{C}_A(K, \mathbb{R}^n)$ is the fiber of the projection $\widetilde{E}_A(\mathbb{R}^j, \mathbb{R}^n) \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ at an embedding K:

$$\begin{array}{ccc} \widetilde{E}_A(\mathbb{R}^j, \mathbb{R}^n) & \longrightarrow & B_A \\ \downarrow & & \downarrow b \\ E_{V/A}(\mathbb{R}^j, \mathbb{R}^n) & \xrightarrow{D} & \mathrm{Inj}(\mathbb{R}^j, \mathbb{R}^n) \\ \downarrow \pi & & \\ \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n) & & \end{array}$$

Here, $\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$ is the space of linear injective maps from $\mathbb{R}^j$ to $\mathbb{R}^n$. $E_{V/A}(\mathbb{R}^j,\mathbb{R}^n)$ is the bundle over $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ whose fiber is the configuration space $C_{V/A}$ of $V(\Gamma)/A$. B_A is the bundle over $\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$ with the "normalized configuration space" of A as the fiber; see Notation 4.4. D is a map from $E_{V/A}(\mathbb{R}^j,\mathbb{R}^n)$ to $\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$ which assigns the differential at the collapsed point A/A.

Notation 4.4 We define the configuration spaces $C_{V/A}$ and B_A as follows. Write $V(\Gamma)/A$ and A as

$$V(\Gamma)/A=\{i_1,\dots,i_{s'},i_{s'+1},\dots,i_{s'+t'}\},$$
$$A=\{j_1,\dots,j_{s''},j_{s''+1},\dots,j_{s''+t''}\},$$

where $s'+s''=s+1$ and $t'+t''=t$. Then $C_{V/A}(K)$ for $K\in\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ and $B_A(I)$ for $I\in\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$ are defined as

$$C_{V/A}(K)=C_{s',t'}(K),\quad B_A(I)=C_{s'',t''}(I)/\sim.$$

Here, $\sim$ is a relation generated by (a) rescalings (b) translations in the I direction:

(a) $(x_{j_1},\dots,x_{j_{s''}},x_{j_{s''+1}},\dots,x_{j_{s''+t''}})\sim r(x_{j_1},\dots,x_{j_{s''}},x_{j_{s''+1}},\dots,x_{j_{s''+t''}})$, for $r\in\mathbb{R}_{>0}$;

(b) $(x_{j_1},\dots,x_{j_{s''}},x_{j_{s''+1}},\dots,x_{j_{s''+t''}})\sim v+(x_{j_1},\dots,x_{j_{s''}},x_{j_{s''+1}},\dots,x_{j_{s''+t''}})$, for $v/\!/I$.

Observe that if $A=V(\Gamma)$, (the restriction of) the form $\omega(\Gamma)$ on $\widetilde{E}_A(\mathbb{R}^j,\mathbb{R}^n)$ is described as the pullback of a form on B_A, for which we write $\omega(\Gamma)$, abusing notation.

Let $H=\sum_i\Gamma_i$ be a g-loop graph cocycle of order k. In the next proposition, we construct a correction term for anomalous faces when $H^d_{dR}(\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n))=H^d_{dR}(V_{n,j})=0$ for a certain d.

Proposition 4.5 *Assume that contributions of hidden faces of H are canceled. For each Γ of H, set $A=V(\Gamma)$ and consider the bundle $b\colon B_A\to\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$. Then the fiber integral*

$$b_*\omega(H)=\sum_i b_*\omega(\Gamma_i)\in\Omega^d_{dR}(\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)),\quad d=k(n-j-2)+(g-1)(j-1)+(j+1),$$

is closed. Moreover, when $b_*\omega(H)$ is exact, we can construct a correction term of the anomalous face by using $\mu\in\Omega^{d-1}_{dR}(\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n))$ which satisfies $d\mu=(-1)^{r+1}b_*\omega(H)$, for $r=d-j$:

$$c(H)=\pi_*D^*\mu.$$

Proof Without loss of generality, assume $H=\Gamma_1+\Gamma_2$. The first assertion follows because the integrals $b_{1*}\omega(\Gamma_1)$ and $b_{2*}\omega(\Gamma_1)$ do not have anomalous faces, while contributions of hidden and principal faces of the integrals are canceled.

Next, we show that $I(H)+c(H)$ is closed. Using Stokes' theorem (see the paragraph after Definition 2.35), we have

$$(-1)^r dI(H)=\int_{\widetilde{C}_{V(\Gamma_1)}}\omega(\Gamma_1)+\int_{\widetilde{C}_{V(\Gamma_2)}}\omega(\Gamma_2)=\pi_*D^*b_{1*}\omega(\Gamma_1)+\pi_*D^*b_{2*}\omega(\Gamma_2).$$

On the other hand,

$$(-1)^r dc(H) = (-1)^r d\pi_* D^*\mu = (-1)^r \pi_* D^* d\mu + \pi_*^{\partial} D^*\mu = -\pi_* D^* b_{1*}\omega(\Gamma_1) - \pi_* D^* b_{2*}\omega(\Gamma_2),$$

so $d(I(H) + c(H)) = 0$. □

Next, we consider the case $H^d_{dR}(\mathrm{Inj}(\mathbb{R}^j, \mathbb{R}^n)) = 0$ is not satisfied. In this case, we construct a correction term of

$$r^* I(H) \in \Omega^{k(n-j-2)+(g-1)(j-1)}\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n),$$

where $r : \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ is the natural projection defined in Definition 2.4. Define

$$\overline{D} : [0,1] \times C_1(\mathbb{R}^j) \times \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) \to \mathrm{Inj}(\mathbb{R}^j, \mathbb{R}^n)$$

by $\overline{D}(t, x, \overline{K}) = D_x(\overline{K}_t)$, where D is the differential. Note that $\overline{D}(1) = D \circ (\mathrm{id} \times r)$. Let

$$p_{23} : [0,1] \times C_1(\mathbb{R}^j) \times \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) \to C_1(\mathbb{R}^j) \times \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$$

be the projection to the second and the third factors.

Definition 4.6 We define the correction term for Γ by

$$\bar{c}(\Gamma) = (-1)^r (-1)^{d+1} \pi_* p_{23*} \overline{D}^* b_* \omega(\Gamma).$$

For a graph cocycle $H = \sum_i \Gamma_i$, set $\bar{c}(H) = \sum_i \bar{c}(\Gamma_i)$.

Proposition 4.7 *Let $H = \sum_i \Gamma_i$ be a graph cocycle such that contributions of hidden faces are canceled. Then $\bar{c}(H)$ is a correction term of $r^* I(H)$.*

Proof By using the closedness of $b_*\omega(H) = b_{1*}\omega(\Gamma_1) + b_{2*}\omega(\Gamma_2)$, we have

$$\begin{aligned}
(-1)^r d\bar{c}(H) &= (-1)^{d+1} \pi_* d(p_{23*} \overline{D}^* b_{1*}\omega(\Gamma_1)) + (-1)^{d+1} \pi_* d(p_{23*} \overline{D}^* b_{2*}\omega(\Gamma_2)) \\
&= -\pi_* (\mathrm{id} \times r)^* D^* b_{1*}\omega(\Gamma_1) - \pi_* (\mathrm{id} \times r)^* D^* b_{2*}\omega(\Gamma_2) \\
&= -r^* \pi_* D^* b_{1*}\omega(\Gamma_1) - r^* \pi_* D^* b_{2*}\omega(\Gamma_2) \\
&= -(-1)^r d(r^* I(H)).
\end{aligned}$$

See the diagram below:

$$\begin{array}{ccccc}
\mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n) & \xleftarrow[\pi]{} & C_1(\mathbb{R}^j) \times \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n) & & \\
\uparrow{\scriptstyle r} & & \uparrow{\scriptstyle \mathrm{id}\times r} & \searrow^{D} & \\
\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) & \xleftarrow[\pi]{} & C_1(\mathbb{R}^j) \times \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) & & \mathrm{Inj}(\mathbb{R}^j, \mathbb{R}^n) \xleftarrow[b]{} B_{V(\Gamma)} \\
 & & \uparrow{\scriptstyle p_{23}} & \nearrow_{\overline{D}} & \\
 & & [0,1] \times C_1(\mathbb{R}^j) \times \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n) & &
\end{array}$$

□

5 Review of Sakai and Watanabe's construction of the wheel-like cycles

Here, we review Sakai and Watanabe's construction [29, Section 4] of the wheel-like cycles

$$c_k : (S^{n-j-2})^k \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n), \quad \boldsymbol{v} \mapsto \varphi_k^{\boldsymbol{v}}$$

and

$$\tilde{c}_k : (S^{n-j-2})^k \to \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$$

for $k \geq 2$. The long embeddings $\varphi_k^{\boldsymbol{v}}$ are constructed by "perturbating" [29, 4.1.1] the original long embedding φ_k. The long embedding φ_k is constructed from a specific oriented immersed 2-disk in $\mathbb{R}^3$, called the wheel-like ribbon presentation. This embedding φ_k has specific k parts called "crossings", where we can perform the perturbation to the direction $v_i \in S^{n-j-2}$ $(i = 1, \dots, k)$. This perturbation is analogous to the perturbation for constructing $k(n-3)$-cycles in $\mathrm{Emb}(\mathbb{R}^1, \mathbb{R}^n)$. (See [10].) These crossings can be described as images of several standard disks and annuli in $\mathbb{R}^j$ [29, 4.2.2], which look like "planetary systems". See [29, Figure 4.6].

Below, we often consider $\mathbb{R}^n$ as the product $\mathbb{R}^3 \times \mathbb{R}^{n-j-2} \times \mathbb{R}^{j-1}$. In this section, the standard embedding $\iota : \mathbb{R}^j \subset \mathbb{R}^n$ is considered as $x_1 \mapsto x_1$ and $x_i \mapsto x_{n-j+i}$ for $2 \leq i \leq j$.

5.1 The wheel-like ribbon presentations

The original embedding φ_k is constructed from the wheel-like ribbon presentation $P \subset \mathbb{R}^3$ (defined below) by thickening P and taking the boundary. The embedding φ_k is realized in $\mathbb{R}^3 \times \mathbf{0} \times \mathbb{R}^{j-1}$.

Definition 5.1 [29, Definition 4.1] The wheel-like ribbon presentation $P = \mathcal{D} \cup \mathcal{B}$ of order k is a based oriented, immersed 2-disk in $\mathbb{R}^3$ such that

- $\mathcal{D} = D_0 \cup D_1 \cup \cdots \cup D_k$ are disjoint $(k+1)$ disks,
- $\mathcal{B} = B_1 \cup B_2 \cup \cdots \cup B_k$ are disjoint k bands $(B_i \approx I \times I)$,
- B_{i+1} connects D_0 with D_i so that $B_{i+1} \cap D_0 = \{0\} \times I$ and $B_{i+1} \cap D_i = \{1\} \times I$ (B_1 is considered as B_{k+1}),
- B_i intersects transversally with the interior of D_i,
- the base point $*$ is on the boundary of D_0 (but not on the boundaries of B_i).

Near the intersection of B_i and D_i, they look, using a local coordinate, like

$$B = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid |x_1| \leq \tfrac{1}{2}, |x_2| < 3, x_3 = 0\},$$
$$D = \{(x_1, x_2, x_3) \in \mathbb{R}^3 \mid |x_1|^2 + |x_3|^2 \leq 1, x_2 = 0\}.$$

See the left picture in Figure 4.

Definition 5.2 [29, Definition 4.2] We define the ribbon $(j+1)$-disk V_P by

$$V_P = \left(\mathcal{B} \times \left[-\tfrac{1}{4}, \tfrac{1}{4}\right]^{j-1}\right) \cup \left(\mathcal{D} \times \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]^{j-1}\right) \subset \mathbb{R}^3 \times \mathbf{0} \times \mathbb{R}^{j-1}.$$

Note that the thicknesses of bands and that of disks are different.

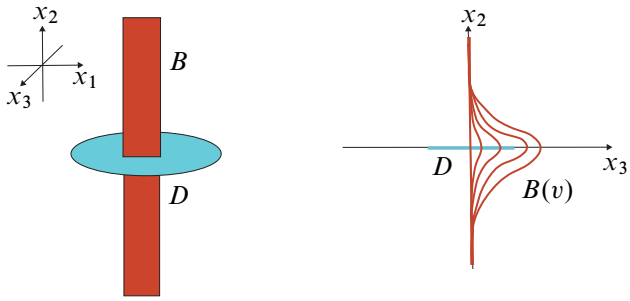


Figure 4: Perturbation of a crossing.

Definition 5.3 After smoothing corners of V_P, the embedding φ_k is defined by the connected sum $\partial V_P \# \iota(\mathbb{R}^j)$ at the base point.

5.2 Perturbation of crossings

Here, we define the perturbation $\varphi_k^{\boldsymbol{v}}$ of the original embedding φ_k. First, we define the perturbation $P_{\boldsymbol{v}}$ of the ribbon presentation P. The embedding $\varphi_k^{\boldsymbol{v}}$ is constructed from $P_{\boldsymbol{v}}$ in the same way as φ_k. Recall that the ribbon presentation P is constructed in $\mathbb{R}^3 \times \mathbf{0} \times \mathbf{0}$, and the thickened presentation V_p is constructed in $\mathbb{R}^3 \times \mathbf{0} \times \mathbb{R}^{j-1}$. The perturbation is performed using the coordinates $(x_4, \dots, x_{n-j+1})$ of $\mathbb{R}^{n-j-2}$ and the coordinate x_3 of $\mathbb{R}^3$.

Notation 5.4 We consider the $(n-j-2)$-dimensional sphere S^{n-j-2} as

$$S = S^{n-j-2} = \{(x_3, \dots, x_{n-j+1}) \in \mathbb{R}^{n-j-1} \mid (x_3-1)^2 + x_4^2 + \cdots + x_{n-j+1}^2 = 1\}.$$

Note that if $n-j-2=0$, S is the set of two points $S^0 = \{x_3 = 0, x_3 = 2\}$.

Notation 5.5 The band obtained by perturbing B to the $v \in S$ direction is written as $B(v)$. (See the right picture in Figure 4)

$$B(v) = \{(x_1, x_2, \gamma(x_2)v) \in \mathbb{R}^2 \times \mathbb{R}^{n-j-1} \mid |x_1| \le \tfrac{1}{2}, |x_2| < 3\}.$$

Here $\gamma(y)$ is a test function defined as

$$\gamma(y) = \begin{cases} \exp\bigl(-y^2/\sqrt{9-y^2}\bigr) & \text{if } |y| \le 3, \\ 0 & \text{if } |y| \ge 3. \end{cases}$$

Definition 5.6 We define $B_i(v_i)$, for $v_i \in S$, as the band in $\mathbb{R}^3 \times \mathbb{R}^{n-j-2}$ constructed from B_i by replacing $B \times \mathbf{0}$ by $B(v_i)$. More precisely, let $U_i \subset \mathbb{R}^3 \times \mathbb{R}^{n-j-2}$ be a small neighborhood of the crossing $B_i \cap D_i$, and let

$$\xi_i : U_i \to [-3,3]^{n-j+1}$$

be a local coordinate that maps B_i to $B \times \mathbf{0}$ and D_i to $D \times \mathbf{0}$. Then $B_i(v_i)$ coincides with

$$(B_i \setminus U_i) \cup \xi_i^{-1}(B(v)).$$

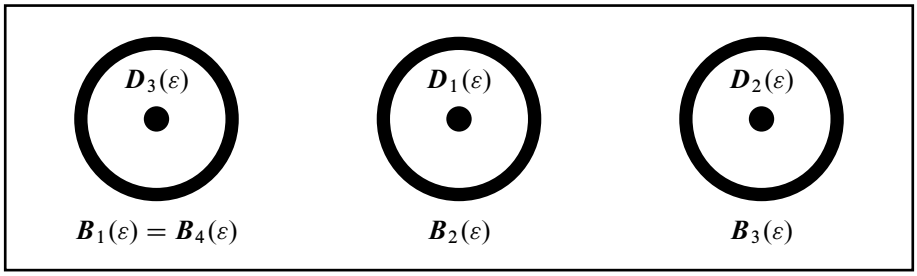


Figure 5: $\varphi_3^{\boldsymbol{v},\varepsilon}$ as embeddings from standard disks and annuli.

Notation 5.7 Let $\boldsymbol{v} = (v_1, \dots, v_k) \in (S^{n-j-2})^k$. We define $\mathcal{B}_{\boldsymbol{v}}$ by

$$\mathcal{B}_{\boldsymbol{v}} = B_1(v_1) \cup B_2(v_2) \cup \cdots \cup B_k(v_k).$$

The perturbation $P_{\boldsymbol{v}} \subset \mathbb{R}^{n-j+1}$ of the wheel-like ribbon presentation P is defined by $P_{\boldsymbol{v}} = \mathcal{D} \cup \mathcal{B}_{\boldsymbol{v}}$. Set

$$V_{P_{\boldsymbol{v}}} = \left(\mathcal{B}_{\boldsymbol{v}} \times \left[-\tfrac{1}{4}, \tfrac{1}{4}\right]^{j-1}\right) \cup \left(\mathcal{D} \times \left[-\tfrac{1}{2}, \tfrac{1}{2}\right]^{j-1}\right) \subset \mathbb{R}^n.$$

Definition 5.8 As a submanifold in $\mathbb{R}^n$, the perturbation $\varphi_k^{\boldsymbol{v}}$ of φ_k is defined by the connected sum $\partial V_{P_{\boldsymbol{v}}} \# \iota(\mathbb{R}^j)$. The parametrization of this submanifold is described in Section 5.3.

Definition 5.9 We define the wheel-like cycle c_k by

$$c_k : (S^{n-j-2})^k \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n), \quad \boldsymbol{v} \mapsto \varphi_k^{\boldsymbol{v}}.$$

Notation 5.10 [29, Definition 4.6] Let $\widehat{U}_i = U_i \times \left[-\frac{3}{4}, \frac{3}{4}\right]^{j-1}$. Inside the neighborhood $\widehat{U}_i$, there are two components $\widehat{D}_i$ and $\widehat{B}_i(v_i)$ of $\varphi_k^{\boldsymbol{v}}$. $\widehat{D}_i$ is homeomorphic to the punctured j-disk, while $\widehat{B}_i(v_i)$ is homeomorphic to the annulus $I \times S^{j-1}$. $\widehat{D}_i$ and $\widehat{B}_i(v_i)$ correspond to D_i and $B_i(v_i)$ respectively. The pair $(\widehat{U}_i, \widehat{D}_i \cup \widehat{B}_i(v_i))$ is called the i^{th} crossing of $\varphi_k^{\boldsymbol{v}}$.

5.3 The path to the trivial immersion

In the previous subsections, we have given a continuous $(S^{n-j-2})^{\times k}$-family of embedded $\mathbb{R}^j$ in $\mathbb{R}^n$. To give a parametrization to this submanifold $\varphi_k^{\boldsymbol{v}}$ for each $\boldsymbol{v}$, we consider sequential resolutions of crossings. (See [29, Remark 4,3].) These resolutions are possible in the space of immersions. More precisely, we consider a path from $\varphi_k^{\boldsymbol{v}}$ to the trivially immersed $\mathbb{R}^j$, associated with the following moves.

(m1) Pull the disk D_i of each crossing to the x_1-direction so that D_i and $B_i(v_i)$ do not intersect.

(m2) Pull back D_i and $B_i(v_i)$ to the based disk D_0.

Using this path, we can equip each point of the submanifold $\varphi_k^{\boldsymbol{v}}$ with a coordinate. This parametrization is continuous with respect to $\boldsymbol{v}$. Moreover, this path gives a lift $\tilde{\varphi}_k^{\boldsymbol{v}}$ of $\varphi_k^{\boldsymbol{v}}$ to $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$. We define the cycle $\tilde{c}_k : (S^{n-j-2})^k \to \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ by $\boldsymbol{v} \mapsto \tilde{\varphi}_k^{\boldsymbol{v}}$.

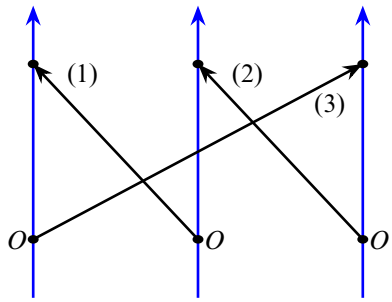


Figure 6: The cycle c_3 as a chord diagram on three directed lines.

5.4 Crossings as embeddings from standard disks and annuli

In [29], Sakai and Watanabe deformed $\varphi_k^{\boldsymbol{v}}$ by isotopy to give embeddings $\varphi_k^{\boldsymbol{v},\varepsilon}$. In this deformation, we "shrink" $\widehat{D}_i$ and $\widehat{B}_i(v_i)$ by $\varphi_k^{\boldsymbol{v}}$. See [29, 4.2.1] for the precise definition. This shrinking aims to make pairing arguments with cocycles easier. We write $\widehat{D}_i(\varepsilon)$ (resp. $\widehat{B}_i(v_i,\varepsilon)$) for the image of $\widehat{D}_i$ (resp. $\widehat{B}_i(v_i)$) by this deformation.

The i^{th} crossing $(\widehat{U}_i, \widehat{D}_i(\varepsilon)\cup\widehat{B}_i(v_i,\varepsilon))$ of $\varphi_k^{\boldsymbol{v},\varepsilon}$ is described as an image of a standard disk and annulus $\boldsymbol{D}_i(\varepsilon)\cup\boldsymbol{B}_i(\varepsilon)$:

$$\boldsymbol{D}_i(\varepsilon) = \{(x_1,\dots,x_j)\in\mathbb{R}^j \mid (x_1-p_i)^2+x_2^2+\cdots+x_j^2\le(\varepsilon^2)^2\} \qquad (1\le i\le k),$$
$$\boldsymbol{B}_i(\varepsilon) = \{(x_1,\dots,x_j)\in\mathbb{R}^j \mid (3\varepsilon/4)^2\le(x_1-p_{i-1})^2+x_2^2+\cdots+x_j^2\le\varepsilon^2\} \quad (1\le i\le k+1),$$

where $p_i = i/k$ $(1\le i\le k-1)$ and $p_0 = p_k = 0$. Recall that $\boldsymbol{B}_1(\epsilon) = \boldsymbol{B}_{k+1}(\epsilon)$.

The subset $\bigcup_i \boldsymbol{D}_i\cup\bigcup_i \boldsymbol{B}_i(\varepsilon)$ looks like "planetary systems" on $\mathbb{R}^j$. (See Figure 5.)

Remark 5.11 In Section 6, the cycle $c_k\colon \boldsymbol{v}\mapsto\varphi_k^{\boldsymbol{v},\varepsilon}$ is described as the chord diagram on k directed lines as in Figure 6.

6 Construction of the cycles

In this section, we construct a cycle of the space of long embeddings from what we call "a set of planetary systems" on $\mathbb{R}^j$. Furthermore, we will see that such a set of planetary systems is described by a chord diagram on directed lines. The sets of planetary systems corresponding to Sakai and Watanabe's wheel-like cycles are the ones in which each planetary system has exactly one planetary orbit. (See Figure 5.) In terms of ribbon presentations, we use ribbon presentations with more than one node, while wheel-like ribbon presentations have exactly one node, the based disk D_0.

6.1 Planetary systems and chord diagrams on directed lines

Using the terminology "planetary system" for configuration spaces appears in Sinha's work [31].

Definition 6.1 (a set of planetary systems on $\mathbb{R}^j$) A set of planetary systems $\mathcal{S}$ on $\mathbb{R}^j$ of order k consists of the following:

(1) points a_i at $(i, 0, \dots, 0) \in \mathbb{R}^j$ $(i = 1, \dots, s)$,

(2) spheres $S(L_i^l, a_i)$ $(l = 1, \dots, t_i)$ with radius L_i^j centered at a_i $(i = 1, \dots, s)$.

Here, s and t_i $(i = 1, \dots, s)$ are nonnegative integers such that the total number of points and spheres is $2k$.[4] We also assume L_i^l is sufficiently smaller than 1, and assume L_i^l is sufficiently smaller than $L_i^{l'}$ if $l < l'$. For example, set $\delta = 1/10^N$ and $L_i^l = \delta^{(t_i+1-l)}$.

Below we call points of (1) fixed stars. We call spheres of (2) planetary orbits. Each fixed star forms one planetary system which consists of t_i planetary orbits.

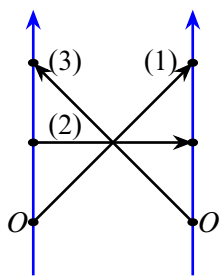
Definition 6.2 An ordered pairing $\{p_i\}_{i=1,\dots,k}$ for $\mathcal{S}$ is a pairing among the set of fixed stars and planetary orbits. Fixed stars must be paired with planetary orbits. Planetary orbits may be paired with other planetary orbits (orbit-orbit pairing). The order of two elements of each pair is taken so that the fixed star (if it exists) is the first.

A set of planetary systems and its ordered pairing can be described by a chord diagram on directed lines.

Definition 6.3 (chord diagram on directed lines) A chord diagram on s directed lines of order k consists of the following data.

- Integers $t_i \geq 0$ $(i = 1, \dots, s)$ such that
$$\sum_{i=1}^{s} (t_i + 1) = 2k.$$
- An ordered pairing $\{p_i\}_{i=1,\dots,k}$ among $2k$ points $V = \{(i, l) \in \mathbb{Z}^2 | 1 \leq i \leq s, 0 \leq l \leq t_i\}$.

Such a chord diagram is drawn as:



The ordered pairing must be taken so that it satisfies:

- At least one of the two points of each pair is not on the x-axis (that is, its second coordinate is not zero).
- Any point $v = (i, 0)$ on the x-axis must be the first of some ordered pair.

[4] s and t_i here are not related to s and t of $C_{s,t}$.

A set of planetary systems and its ordered pairing determines a chord diagram on directed lines in an expected way. Namely, points on the x-axis are fixed stars. Other points are planetary orbits. Conversely, a chord diagram on directed lines determines a set of planetary systems (up to rescalings) and its ordered pairing.

Let C be a chord diagram on directed lines. Remember that the set of chords (pairings) $E(C)$ is ordered, and each chord is oriented.

Definition 6.4 (induced ordering of $V(C)$) We order the set of vertices $V(C)$ in a canonical way using the above ordering and orientations. That is, the initial vertex of the i^{th} chord is labeled by $(2i-1)$. The end vertex of the i^{th} chord is labeled by $2i$.

Using the ordering of $V(C)$, let the image of $\tau\colon\{1,\dots,g-1\}\to\{1,\dots,2k\}=V(C)$ be the first vertices of orbit-orbit pairing; $g=k-s$. We assume that τ preserves the orders.

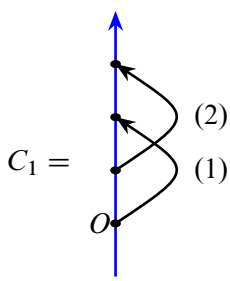
6.2 Construction of the cycles

Once we fix a set of planetary systems and its ordered pairing, namely a chord diagram on directed lines, we can construct a "generalized ribbon cycle". These generalized cycles have more parameters than the wheel-like cycles in Section 5.

Definition 6.5 (generalized ribbon cycles) Let $\mathcal{S}$ be a set of planetary systems of order k. Let $\{p_i\}_{i=1,\dots,k}$ be an ordered pairing. Let $(g-1)$ be the number of orbit-orbit pairings. From this data, we construct a $k(n-j-2)+(g-1)(j-1)$ cycle $\psi=\psi(\mathcal{S},\{p_i\}_{i=1,\dots,k})$ in the following. We call this cycle the generalized ribbon cycle associated with $(\mathcal{S},\{p_i\}_{i=1,\dots,k})$. For simplicity, we only construct for the following case:

- There is only one fixed star a_1 (and hence one planetary system).
- There are three planetary orbits $S(L_l,a_1)$ for $l=1,2,3$.
- The pairing is taken as $(a_1,S(L_2))$, $(S(L_1),S(L_3))$.

The corresponding chord diagram is



The cycles corresponding to other $(\mathcal{S},\{p_i\}_{i=1,\dots,k})$ are constructed similarly.

Below we construct a cycle in $\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$,

$$\phi\colon S^{j-1}\times S^{n-j-2}\times S^{n-j-2}\to\mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n).$$

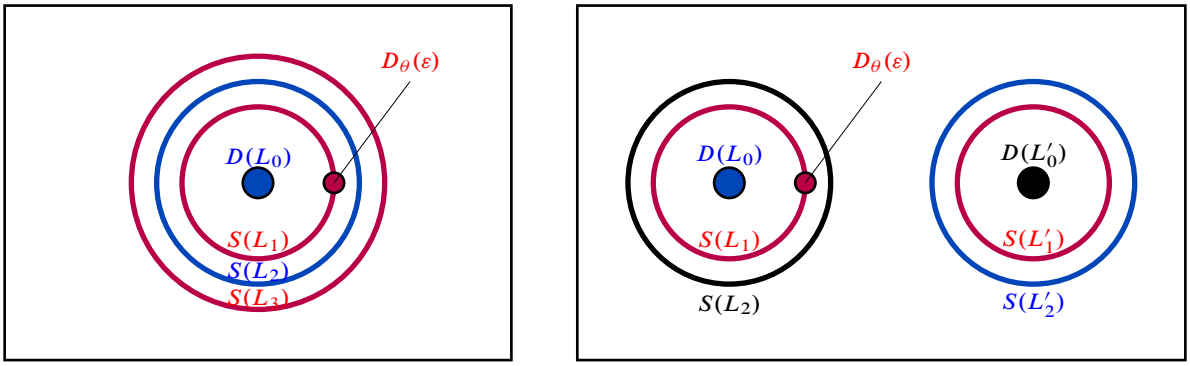


Figure 7: Preimage on $\mathbb{R}^j$ of ϕ (left) and ψ (right).

Let $S(L)$ and $D(L)$ denote the $(j-1)$-sphere and j-ball respectively, of radius L centered at O in $\mathbb{R}^j$. And let $L_0, L_1, L_2, L_3 \in \mathbb{R}$ satisfy

$$0 < L_0 < L_1 < L_2 < L_3 < L_4 = 1,$$

where L_i is sufficiently smaller than L_{i+1} (for example, set $L_i = \delta^{4-i}$). For each $\theta \in S^{j-1}$, we write $D_\theta(\varepsilon)$ for the ball of radius ε centered at $\theta \in S(L_1)$. We assume ε is sufficiently smaller than L_1. (See Figure 7.) Steps 1 and 2 below are to construct a cycle $\phi_\theta \colon S^{n-j-2} \times S^{n-j-2} \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n)$ for each parameter $\theta \in S^{j-1}$. Let $0 < h_4 < h_3 < h_2 < h_1 < h_0$ be coordinates of the x_{j+1} axis.

Step 1 Deform $D(L_3)$ by isotopy in $\mathbb{R}^j \times \mathbb{R} \subset \mathbb{R}^n$ to the x_{j+1} direction, so that $\mathbb{R}^j$ is embedded in $\mathbb{R}^{j+1}$ as

$$S(r) \times (h_4, h_0)$$

in the range $h_4 < x_{j+1} < h_0$. Here $S(L_3)$, $S(L_2)$ and $S(L_1)$ are mapped to $S(r) \times \{h_3\}$, $S(r) \times \{h_2\}$ and $S(r) \times \{h_1\}$ respectively. We assume r is sufficiently small compared to $\min\{|h_i - h_{i+1}|\}_i$.

Step 2 For $L = L_2, L_3$, let $S(L) \times I$ be the tubular neighborhood of $S(L)$ in $\mathbb{R}^j$ with width 2ε. Construct a ribbon crossing of $S(L_3) \times I$ and $D_\theta(\varepsilon)$ near $S(r) \times h_3$, and a ribbon crossing of $S(L_2) \times I$ and $D(L_0)$ near $S(r) \times h_2$. Here, a ribbon crossing is the crossing given by the perturbation of a crossing of a ribbon presentation as in Section 5.2. See Figure 8. Then we obtain a cycle

$$\phi_\theta \colon S^{n-j-2} \times S^{n-j-2} \to \mathrm{Emb}(\mathbb{R}^j, \mathbb{R}^n).$$

We describe a ribbon presentation corresponding the cycle ϕ_θ in Section 6.3. By choosing a ribbon presentation suitably, we can construct this cycle in the unknot component.

Step 3 Move $D_\theta(\varepsilon)$ on $S(L_1)$. In $\mathbb{R}^{j+1}$, we change the behavior only near h_1. At the height h_1, we turn the branch tube around the stem tube. See Figure 8. Then we get the desired cycle ϕ. The parametrization of ϕ is given in Section 6.4.

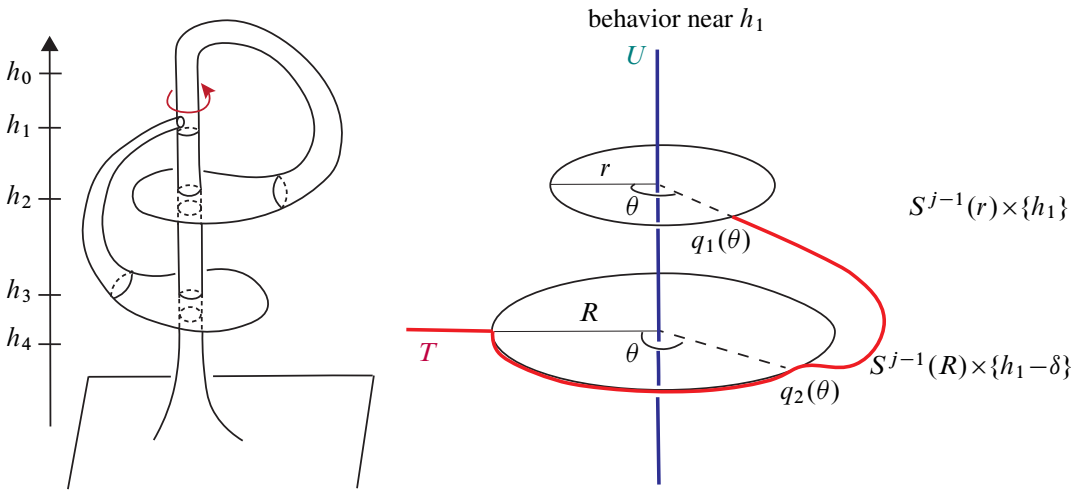


Figure 8: Image on $\mathbb{R}^{j+1}$.

We describe Step 3 more precisely. We write U for the stem tube $S^{j-1}(r)\times(h_4,h_0)$. In Figure 8, we draw the core c_U of U in blue. For $\theta\in S^{j-1}$, let $\partial D_\theta(\varepsilon)$ be mapped to the θ side of U. For the initial point $O=(1,0,\dots,0)\in S^{j-1}$, we construct, near h_1, a branch tube $T=T(O)$ with small radius which starts from U and goes to the x_1-direction. We draw the core $c=c(O)$ of T in red.

For $\theta\in S^{j-1}$, we replace the core c with $c(\theta)$ defined as follows.

- $c(\theta)$ starts from the point $q_1(\theta)=(r\theta,h_1)\in S^{j-1}(r)\times\{h_1\}$ and goes straight to the point $(R\theta,h_1)\in S^{j-1}(R)\times\{h_1\}(R>r)$.
- Next, $c(\theta)$ runs down to $S^{j-1}(R)\times\{h_1-\delta\}$ preserving $\theta\in S^{j-1}(R)$ and reaches the point $q_2(\theta)=(R\theta,h_1-\delta)$.
- Then $c(\theta)$ connects $q_2(\theta)$ and $q_2(O)$ through the shortest geodesic of $S^{j-1}(R)$.

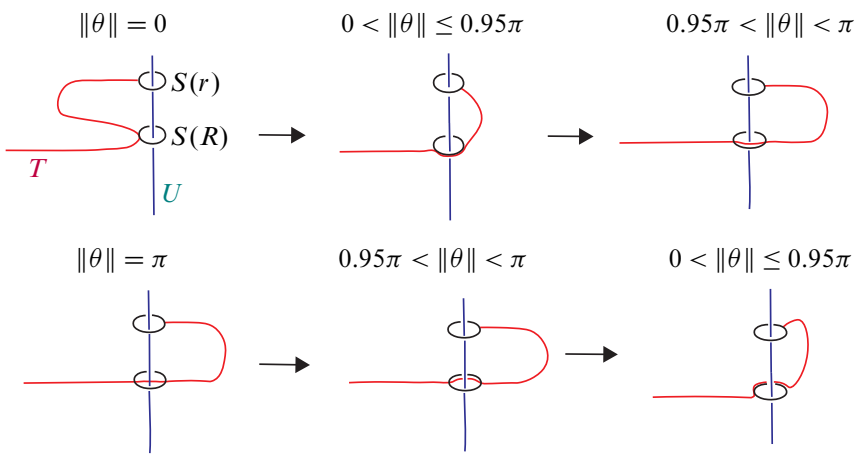


Figure 9: The replaced core $c(\theta)$ near h_1 ($a=0.05\pi$).

However, there is one problem. When θ is the antipode $A \in S^{j-1}$ of the base point O, there are infinitely many geodesics from $q_2(A)$ to $q_2(O)$. To avoid this, $c(\theta)$ takes a short-cut from $q_2(\theta)$ to $q_2(O)$ inside the ball $D^j(R)$, when θ is in a small neighborhood of A. At the same time, when θ is near A, we perturb $c(\theta)$ to the x_{j+2}-direction.

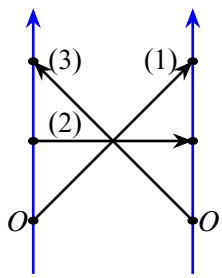
More precisely, let

$$\|\cdot\|\colon S^{j-1} \to [0, \pi]$$

be the function that assigns the geodesic distance from the base point O. Then in the range $\|\theta\| > \pi - a$ for a sufficiently small a, we assume that the core $c(\theta)$ approaches the core c_U of U, and assume that they intersect with each other in $\mathbb{R}^{j+1}$ when $\|\theta\| = \pi$. At this time, we perturb $c(\theta)$ in the positive direction of the x_{j+2} axis. See Figure 9.

Then we take a tubular neighborhood $N(\theta) \in \mathbb{R}^{j+1}$, which is perturbed to the x_{j+2} direction when $\|\theta\| > \pi - a$. The replaced tube $T(\theta)$ is defined as the boundary of $N(\theta)$.

As a candidate of a dual cycle of H, we construct the generalized ribbon cycle in Figure 10 from the diagram



The ribbon presentation corresponding to the cycle ψ is given in Section 6.3. The cycle ψ is also constructed in the unknot component. In general, cycles have parameters $\theta_i' \in S^{j-1}$ $(i = 1, \dots, g-1)$ and $y_j \in S^{n-j-2}$ $(j = 1, \dots, k)$.[5]

6.3 Ribbon presentations with more than one node

Here, we give ribbon presentations corresponding to the (cycles of) long embeddings in Section 6.2. (We review the notation of ribbon presentations at the beginning of this subsection.) These ribbon presentations have a node other than the based disk D_0, and this node is connected (by two bands) to two leaves. Using moves of ribbon presentations, we show that the cycles of Section 6.2 are in the path component of the trivial long embedding. The key point is the way the two leaves intersect with bands.

Definition 6.6 A ribbon presentation $P = \mathcal{D} \cup \mathcal{B}$ of order k is a based oriented, immersed 2-disk in $\mathbb{R}^3$ such that

- $\mathcal{D} = D_0 \cup D_1 \cup \cdots \cup D_k$ are disjoint $(k+1)$ disks,

[5]We change the notation for the parameter moving on S^{j-1}, from "θ" to "θ'". In Section 7, we use the letter "θ" for a different parameter.

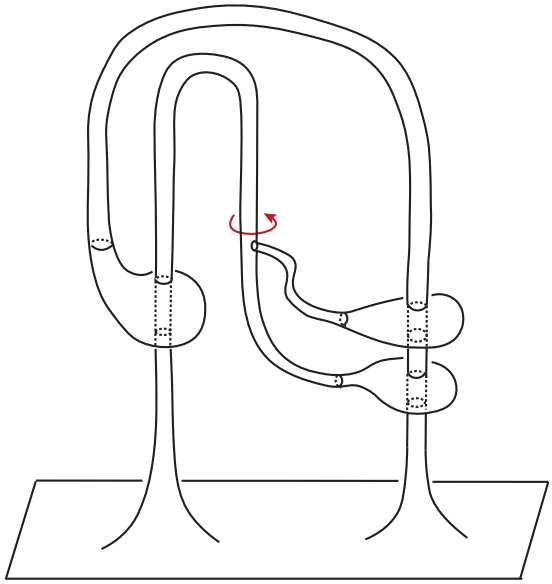


Figure 10: The cycle ψ.

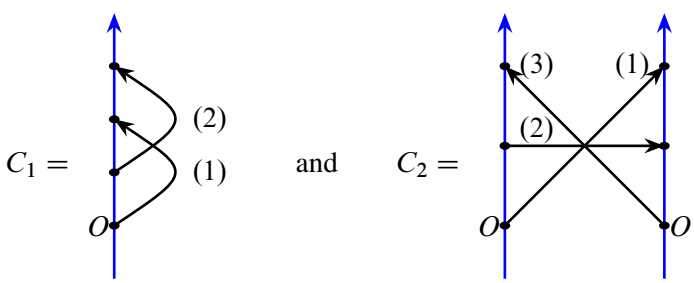
- $\mathcal{B} = B_1 \cup B_2 \cup \cdots \cup B_k$ are disjoint k bands ($B_i \approx I \times I$),
- each band connects two different disks,
- B_i can intersect transversally with the interior of disks,
- the base point $*$ is on the boundary of D_0 (but not on the boundaries of B_i).

Notation 6.7 [17, Definition 3.1] A disk without intersecting bands is called a node. A disk is called a leaf if it has at least one intersecting band and it is incident to exactly one band. We draw a gray disk for a node except for the based disk D_0. We call a ribbon presentation simple if each leaf intersects with exactly one band.

In [17], Habiro and Shima introduced moves of ribbon presentations that do not change the isotopy class of the ribbon 2-knot represented by the presentation. The following are examples of the moves.

Notation 6.8 We define $S1$ and $S4$ moves as in Figure 11.

To the chord diagrams



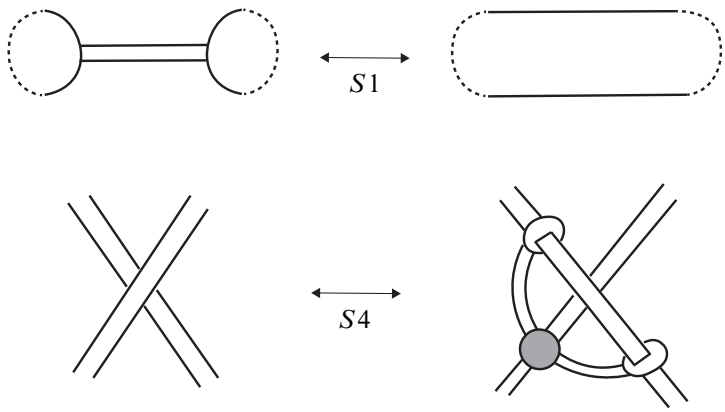


Figure 11: $S1$ and $S4$ moves.

we assign the simple ribbon presentations P_1 in Figure 12 and P_2 in Figure 13. Each directed line corresponds to a sequence of bands and nodes that connects the based disk D_0 and a leaf. The origin of each directed line corresponds to this leaf. Each chord corresponds to a crossing. The initial point of each chord corresponds to the disk (leaf) of the crossing. If this initial point is not an origin, this leaf is connected to a node by a band. The target point of each chord corresponds to a small segment of the band of the crossing.

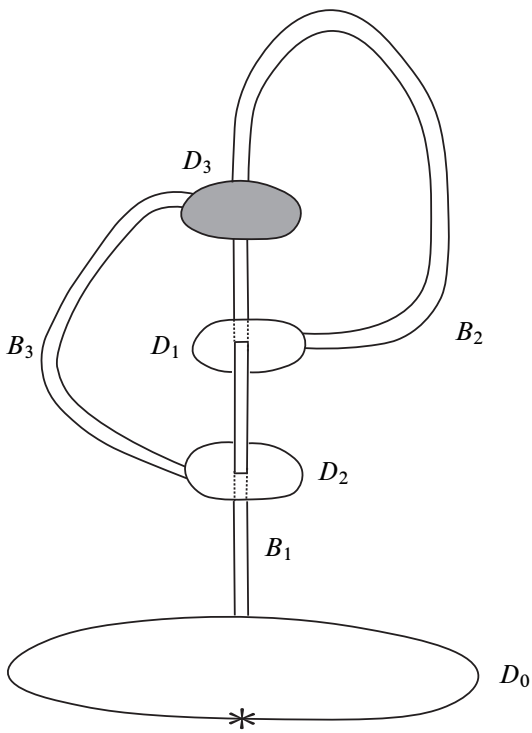


Figure 12: Ribbon presentation P_1 corresponding to C_1.

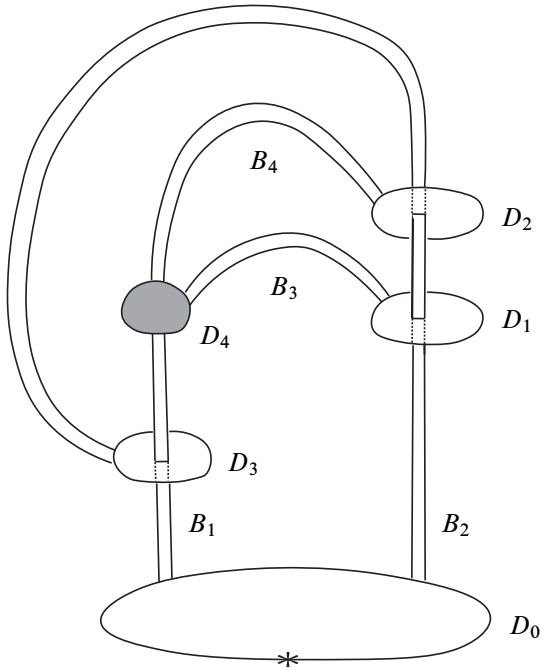
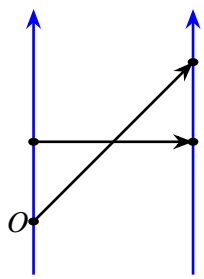
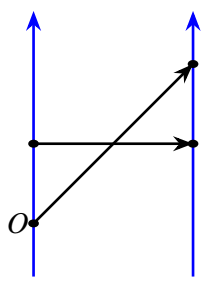


Figure 13: Ribbon presentation P_2 corresponding to C_2.

Observe that our chord diagrams C_1 and C_2 have a subdiagram of the form



Recall that we have constructed a crossing for each chord. There are two ways to construct a crossing, depending on which orientation the band intersects with the disk. In the part



we choose the left presentation of Figure 14. That is, we assume that two crossings have opposite orientations. We can write this crossing information on chord diagrams by assigning a sign to each chord.

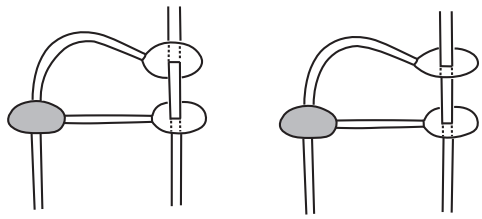
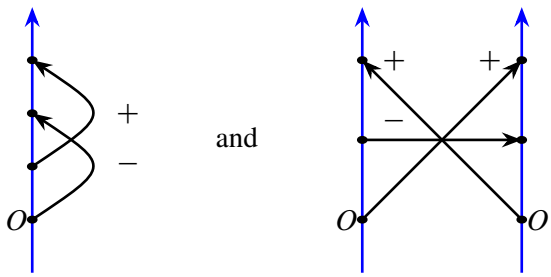


Figure 14: Suitable ribbon presentation (left).

For C_1 and C_2, we choose



Notation 6.9 Let ϕ and ψ be the cycles constructed from the ribbon presentations P_1 and P_2 in Section 6.2. If $n-j=2$, define the cycle

$$\psi(0,0,0)\colon S^{j-1} \to \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^{j+2})$$

as the $(0,0,0) \in S^0 \times S^0 \times S^0$ component of ψ, that is, the component where no crossing is resolved. Define $\phi(0,0)$ similarly.

Proposition 6.10 *The ribbon presentations P_1 and P_2 are equivalent to the trivial presentation. Thus, when $n-j \geq 3$, both ϕ and ψ are in the unknot component. When $n-j=2$, both $\phi(0,0)$ and $\psi(0,0,0)$ are in the unknot component.*

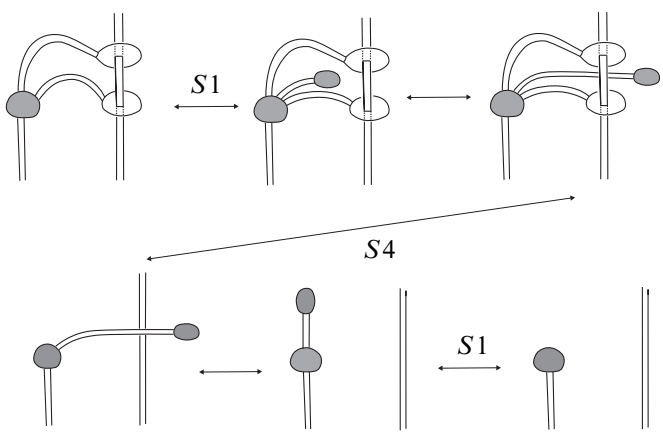


Figure 15: Resolving the pair of two crossings.

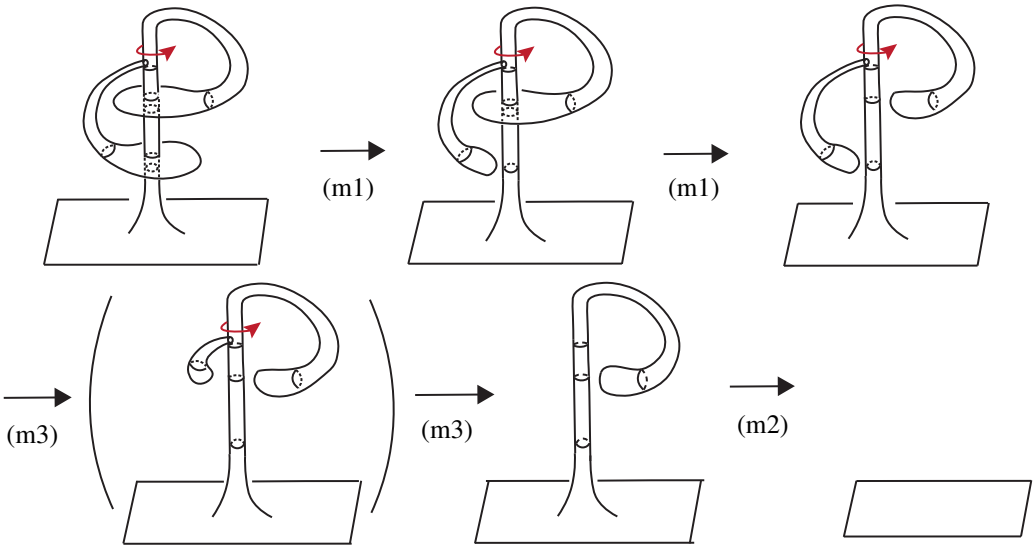


Figure 16: Lift of ϕ to $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$.

Proof By using moves of ribbon presentations including $S4$, we can resolve the pair of two crossings of opposite orientations as in Figure 15. Then, we can show the resulting ribbon presentation is equivalent to the trivial presentation. □

6.4 Lifting the cycles to $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$

Finally, we give a parametrization to the cycle ϕ and the cycle ψ respectively. We give this parametrization by using a path to the trivially immersed $\mathbb{R}^j$ as in Section 5.3. Moreover, this path gives a lift of ϕ and ψ to $\overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$.

The path is constructed as follows. Let $\widehat{D}_i$ be the ball of crossings corresponding to $D_\theta(\varepsilon)$. In the ribbon presentation, the disk D_i is connected to the node. First, we resolve each crossing, by the move (m1) of Section 5.3. Then

(m3) pull back $\widehat{D}_i$ to the stem tube (see Figure 16).

Finally, we deform the immersion to the trivial immersion by the move (m2) of Section 5.3.

7 Some lemmas for computing cocycle-cycle pairing

In this section, we state preliminary lemmas which are used to calculate pairings between cocycles and cycles of the space of long embeddings, as pairings between graphs and diagrams (see Theorem 7.14). Readers who are familiar with configuration space integrals may skip until Lemma 7.7. Graphs in this section are admissible (see Definition 2.19) graphs of defect 0.

7.1 Preliminary lemmas

Lemma 7.1 (symmetry lemma) *Let a labeled BCR graph Γ have an even symmetry ι (a change of labels of even sign that produces the same labeled graph). Consider the involution on $C_{s,t}$*

$$\sigma : (x_1, x_2, \dots, x_s, x_{s+1}, \dots, x_{s+t}) \mapsto (x_{\iota^{-1}(1)}, x_{\iota^{-1}(2)}, x_{\iota^{-1}(s)}, x_{\iota^{-1}(s+1)}, \dots, x_{\iota^{-1}(s+t)}).$$

Let the two disjoint submanifolds X and Y of $C_{s,t}$ satisfy $\sigma(X) = Y$. Then $\int_X I(\Gamma) = \int_Y I(\Gamma)$.

Proof We observe that

$$\int_Y I(\Gamma) = \int_{\sigma(X)} (\sigma^{-1})^* I(\Gamma),$$

since the changes of signs are canceled. □

Lemma 7.2 (localizing lemma [29, Lemma 4.7]) *Let Γ be a BCR graph of order k, $\#B(\Gamma) = s$ and $\#W(\Gamma) = t$. Let ψ be a generalized ribbon cycle with crossings $p_1, \dots, p_k$. Recall that each crossing consists of the image of a ball D_{p_i} (or $D_{p_i}(\theta'^{-1}_{\tau^{-1}(2i-1)})$) and the image of an annulus $I \times S_{p_i}$. Let C_0 be the subspace of $C_{s,t}(\psi)$ which consists of configurations $(x_1, x_2, \dots, x_s, x_{s+1}, \dots, x_{s+t})$ satisfying the following:*

- *There exists i such that either D_{p_i} or $I \times S_{p_i}$ has no black vertex.*

Then $\int_{C_0} I(\Gamma) = 0$.

Proof Let p_i be such a crossing. Then the contribution of the integral vanishes by dimensional reasons (or is canceled) with respect to the parameter y_i of the crossing p_i. □

Set

$$X_\alpha = \begin{cases} D_{p_i} \text{ or } D_{p_i}(\theta'^{-1}_{\tau^{-1}(2i-1)}) & \text{if } \alpha = 2i-1, \\ I \times S_{p_i} & \text{if } \alpha = 2i. \end{cases}$$

By the localizing lemma, we only consider the graph with $s = 2k$ and $t = 0$, and configurations that there are some black vertex $x_{\sigma^{-1}(\alpha)}$ in X_α. We write the set of these configurations as C_σ, $\sigma \in \mathfrak{S}_{2\alpha}$. If $\sigma \neq \tau$, C_σ and C_τ have no intersection.

Lemma 7.3 (pairing lemma [29, Lemma 4.9]) *If $x_{\sigma^{-1}(2i-1)} \in D_{p_i}$ and $x_{\sigma^{-1}(2i)} \in I \times S_{p_i}$ are not connected by dashed edges, $\int_{C_\sigma} I(\Gamma) = 0$.*

Proof This can be shown by a similar dimensional reason to the proof of Lemma 7.2. □

Lemma 7.4 (same system lemma [29, Lemma 4.8]) *Let Γ be a BCR graph of order k (and defect 0), $\#B(\Gamma) = 2k$ and $\#W(\Gamma) = 0$. Let ψ be a generalized ribbon cycle with crossings $p_1, \dots, p_k$. If X_α and X_β are in different planetary systems, and $x_{\sigma^{-1}(\alpha)}$ and $x_{\sigma^{-1}(\beta)}$ are connected by solid edges (see Figure 17), $\int_{C_\sigma} I(\Gamma) = 0$.*

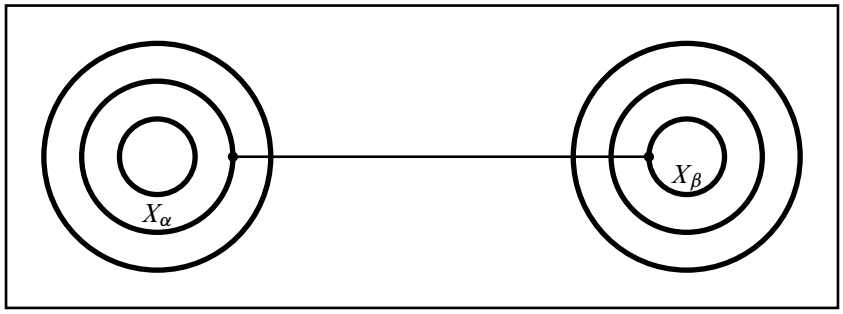


Figure 17: An edge connecting different systems.

Proof This is because the direction

$$\frac{x_{\sigma^{-1}(\beta)} - x_{\sigma^{-1}(\alpha)}}{\|x_{\sigma^{-1}(\beta)} - x_{\sigma^{-1}(\alpha)}\|} \in S^{j-1}$$

varies in a small neighborhood of a point in S^{j-1}. □

Remark 7.5 More precisely, $\int_{C_\sigma} I(\Gamma) = 0$ above means $\int_{C_\sigma} I(\Gamma)$ approaches 0 as the radii of the two planetary systems approach 0.

Lemma 7.6 (ingoing lemma) *Let Γ be a BCR graph of order k, $\#B(\Gamma) = 2k$ and $\#W(\Gamma) = 0$. Let ψ be a generalized ribbon cycle with crossings (pairings) $p_1, \dots, p_k$. Let X_α be a planetary orbit in a planetary system. Assume that all solid edges adjacent to $x_{\sigma^{-1}(\alpha)}$ are outgoing. That is, if $x_{\sigma^{-1}(\alpha)}$ and $x_{\sigma^{-1}(\beta)}$ are connected by solid edges, X_β is outside X_α (see Figure 18). Then $\int_{C_\sigma} I(\Gamma) = 0$.*

Proof Since the integrand form does not depend on the parameter $x_{\sigma^{-1}(\alpha)}$, it factors through

$$C_\sigma \to C_\sigma / X_\alpha.$$

□

Lemma 7.7 (decomposing lemma)

- *Let $e = \{\sigma^{-1}(\alpha), \sigma^{-1}(\beta)\}$ be a solid edge. Let X_α and X_β be in the same planetary system and X_β be outside. Then if $X_\beta = I \times S_p$, $\phi_e^* \omega_{S^{j-1}}$ depends only on the variable $\theta_{\sigma^{-1}(\beta)} \in S_p$ of $x_{\sigma^{-1}(\beta)} = (r_{\sigma^{-1}(\beta)}, \theta_{\sigma^{-1}(\beta)}) \in X_\beta$. If $X_\beta = D_p(\theta')$, it depends only on θ'.*
- *Let $e = \{\sigma^{-1}(\alpha), \sigma^{-1}(\beta)\}$ be a dashed edge. Let $X_\alpha = D_p$ (or $D_p(\theta')$) and $X_\beta = I \times S_p$. Then $\phi_e^* \omega_{S^{n-1}}$ depends on only on the variable $x_{\sigma^{-1}(\alpha)} \in X_\alpha$ and on the first factor $r_{\sigma^{-1}(\beta)} \in I$ of $x_{\sigma^{-1}(\beta)} = (r_{\sigma^{-1}(\beta)}, \theta_{\sigma^{-1}(\beta)}) \in X_\beta$.*

Proof This follows from the way we choose the scaling of planetary orbits. □

Definition 7.8 (graph-chord pairing) Let C be a chord diagram on directed lines of order k. Let Γ be a labeled BCR graph of order k. Then the pairing $\langle \Gamma, C \rangle$ is defined by counting as follows. First, we only count when

(I) $\#B(\Gamma) = 2k (= \#V(C))$.

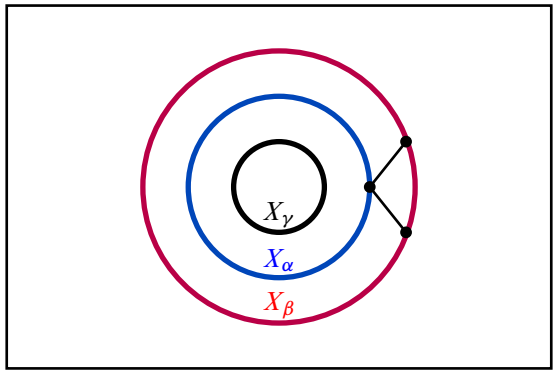


Figure 18: Outgoing edges.

Let $\#B(\Gamma) = \#V(C) = 2k$. We count permutations $\sigma : \#B(\Gamma) \to \#V(C)$ which satisfy (I) and all of the following:

(II) σ induces the map $\bar{\sigma} : E(\Gamma) \to E(C)$.

(III) If two black vertices v_1 and v_2 are in the same solid component, $\sigma(v_1)$ and $\sigma(v_2)$ are on the same directed line.

(IV) If a vertex w of $V(C)$ is not on the x-axis, w has an ingoing solid edge: there exists a vertex w' lower on the same directed line such that $\sigma^{-1}(w)$ and $\sigma^{-1}(w')$ are connected by some solid edge of Γ.

The sign of the counting is $+1$ if and only if the orientation of Γ coincides with the induced orientation determined by the induced label from σ. (See Definition 7.9.)

Definition 7.9 (induced label on BCR diagrams) Let $\sigma : B(\Gamma) \to V(C)$ satisfy conditions (I)–(IV) of Definition 7.8. Then σ defines an induced label on the underlying graph $\overline{\Gamma}$ in a natural way: We order black vertices using σ and the ordering of $V(C)$. (See Definition 6.4.) We orient solid edges so that the orientations are compatible with directed lines. We orient dashed edges so that the orientations are compatible with those of chords. We order edges as follows:

(1) Ingoing solid edge (if it exists) from the initial vertex of the i^{th} chord ($i = 1, \dots, k$). There are $(g-1)$ solid edges ordered by this first step.

(2) The $\bar{\sigma}^{-1}(i)^{\text{th}}$ dashed edge and the ingoing solid edge from the target point of the i^{th} chord ($i = 1, \dots, k$).

Remark 7.10 We do not use the induced ordering of edges when $(n, j) = (\text{odd}, \text{odd})$.

Theorem 7.11 (refer to [29, Lemma 4.12]) *Let C be a chord diagram of order k associated with a generalized ribbon cycle ψ. Assume $r(c)$ chords have the negative sign. Let Γ be a BCR graph of order k. Then the value of the integral $I(\Gamma)$ on ψ is equal to the pairing $(-1)^{r(c)}\langle \Gamma, C \rangle$.*

Proof By the localizing lemma, Lemma 7.2, the pairing lemma, Lemma 7.3, the same system lemma, Lemma 7.4, and the ingoing lemma, Lemma 7.6, respectively, the integral survives only on C_σ for σ which satisfies (I)–(IV) of Definition 7.8. By the decomposing lemma, Lemma 7.7, the integral on C_σ can be written as

$$\pm(-1)^{r(c)}\left(\int_{\theta'_j\in S}\bigwedge_{j=1}^{g-1}\eta_{\tau(j)}\right)\left(\bigwedge_i\int_{x_{\sigma^{-1}(2i-1)}\in D_{p_i}}\int_{r_{\sigma^{-1}(2i)}\in I}\omega_{p_i}\int_{\theta_{\sigma^{-1}(2i)}\in S_{p_i}}\eta_{p_i}\right).$$

Here ω_{p_i} is the form associated with the dashed edge $\{x_{\sigma^{-1}(2i-1)}, x_{\sigma^{-1}(2i)}\}$. η_{p_i} is the form associated with the solid edge that goes down from $x_{\sigma^{-1}(2i)}$. The form $\eta_{\tau(j)}$ is associated with the solid edge that goes down from $x_{\tau(j)}$. The sign at the head is $+$ if and only if the orientation of the configuration space C_σ coincides with the orientation induced by the coordinate $(x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \dots, x_{\sigma^{-1}(2k)})$, and $\omega(\Gamma)$ is exactly $\bigwedge_{j=1}^{g-1}\eta_{\tau(j)}\wedge\bigwedge(\omega_{p_i}\eta_{p_i})$.

Recall that the image of the annulus $I\times S_{p_i}$ is perturbed to the direction $y_i\in S^{n-j-2}$. Then, in the part $\int_{x\in D}\int_{r\in I}\theta$, the linking number of D and $\bigcup_{y_i\in S^{n-j-2}} I(y_i)$ is computed. Since this link is the Hopf link $S^{j-1}\cup\Sigma S^{n-j-2}\hookrightarrow\mathbb{R}^n$, the linking number is $+1$ or -1 depending on the sign of the chord. □

Notation 7.12 Let $C=(\{t_i\}_{i=1,\dots,s},\{p_i\}_{i=1,\dots,k})$ be a chord diagram on directed lines. Let $G(C)$ be the set of BCR graphs (without a label) mapped on C satisfying all the conditions of Definition 7.8.

Lemma 7.13 *$G(C)$ consists of $\sum_{i=1}^{s}2^{t_i-1}$ graphs.*

Proof Since dashed edges must be mapped on chords, the remaining choices are solid edges. The top vertex v on each directed line must be one of the ends of some solid component. The solid component first goes down from v to O through some of the vertices on the directed line. After reaching O, the solid component goes up though the remaining vertices. We can choose which of (t_i-1) vertices between v and O the solid component passes when it goes down. □

7.2 The counting formula of configuration space integrals

Theorem 7.14 (counting formula) *Let $H=\sum w(\Gamma_i)\Gamma_i$ be a graph cocycle that consists of g-loop graphs of order k. Let C be a chord diagram on directed lines of order k, with $r(C)$ chords of negative sign. Let ψ be a generalized ribbon cycle constructed from C. Then*

$$I(H)(\psi)=(-1)^{r(C)}\sum_{\overline{\Gamma}\in G(C)}w(\Gamma)\#\operatorname{Aut}(\overline{\Gamma})s(\Gamma,\overline{\Gamma}).$$

Here Γ is a labeled graph in H such that the underlying graph of Γ is $\overline{\Gamma}$. The sign $s(\Gamma,\overline{\Gamma})\in\{+1,-1\}$ is 1 if and only if the orientation of Γ coincides with that of $\overline{\Gamma}$ with the induced label.

Proof This follows from Theorem 7.11 and the symmetry lemma, Lemma 7.1. □

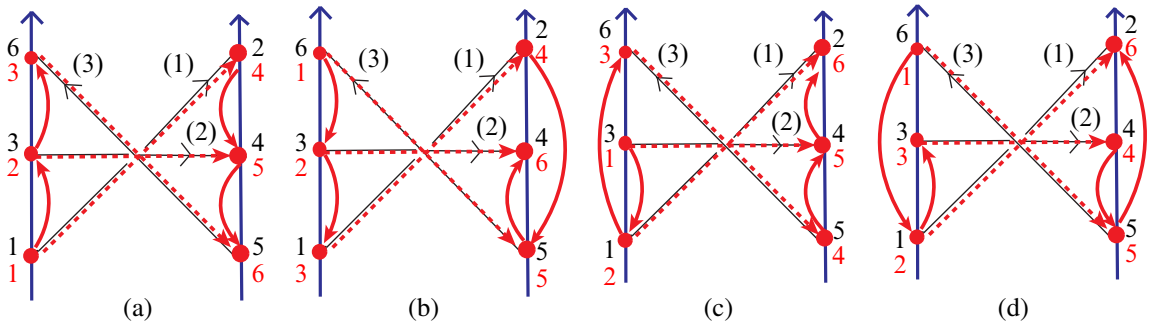
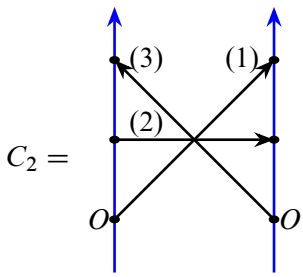
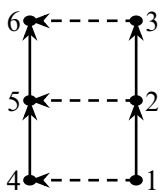


Figure 19: BCR graphs of $G(C_2)$.

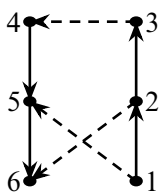
Example 7.15 For the chord diagram



$G(C_2)$ consists of the four graphs drawn in red in Figure 19. The first BCR graph (a) is:



The other three graphs, (b), (c) and (d), are:



The induced label is written in black. The sign $s(\Gamma, \overline{\Gamma})$ is (a) -1, (b) $+1$, (c) $+1$ and (d) -1, respectively.

7.3 Cancellation of correction terms

Finally, we see lemmas for canceling the contribution of the collection term $\bar{c}(H)$.

Lemma 7.16 [29, Lemma 4.13] *Let $\bar{\psi}: X \to \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ be a generalized ribbon cycle of order k, and let $p_1, p_2, \ldots, p_k$ be its crossings. Then there exists a cycle $\bar{\psi}_0: X \to \overline{\mathrm{Emb}}(\mathbb{R}^j, \mathbb{R}^n)$ satisfying the following*:

- *$\bar{\psi}_0$ is obtained from $\bar{\psi}$ by resolving all crossings.*
- *The two differentials $D\bar{\psi}_0, D\bar{\psi}\colon [0,1]\times C_1(\mathbb{R}^j)\times X \to \mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$ are homotopic through an isotopy of $C_1(\mathbb{R}^j)$ that has a support in small neighborhoods of balls of crossings.*

Lemma 7.17 *The cycle $\bar{\psi}_0$ satisfies the following.*

(1) $\bar{c}(H)(\bar{\psi}-\bar{\psi}_0)=0$.

(2) $I(H)(r_*\bar{\psi}_0)=0$, *where* $r\colon \overline{\mathrm{Emb}}(\mathbb{R}^j,\mathbb{R}^n)\to \mathrm{Emb}(\mathbb{R}^j,\mathbb{R}^n)$ *is the natural projection.*

(3) $\bar{\psi}_0$ *is a null-homotopic cycle.*

Proof For (1), recall that $\bar{c}(H)$ depends only on the differential

$$\overline{D}\colon [0,1]\times C_1(\mathbb{R}^j)\times \overline{\mathrm{Emb}}(\mathbb{R}^j,\mathbb{R}^n)\to \mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n),$$

and $\bar{c}(H)$ is defined as the fiber integral of the pullback of some closed form ω on $\mathrm{Inj}(\mathbb{R}^j,\mathbb{R}^n)$. By the second property of Lemma 7.16, the two compositions $D\bar{\psi}_0=\overline{D}\circ(\mathrm{id}_{[0,1]}\times \mathrm{id}_{C_1(\mathbb{R}^j)}\times\bar{\psi}_0)$ and $D\bar{\psi}=\overline{D}\circ(\mathrm{id}_{[0,1]}\times \mathrm{id}_{C_1(\mathbb{R}^j)}\times\bar{\psi})$ differ only in small neighborhoods of balls of crossings. Moreover, observe that neither $D\bar{\psi}_0$ nor $D\bar{\psi}$ depend on the parameter space X, in these small neighborhoods of balls. Hence by a dimensional reason, the integral vanishes on these neighborhoods. So we have (1). Parts (2) and (3) are due to the first property of Lemma 7.16. □

Corollary 7.18 $$(r^*I(H)+\bar{c}(H))(\bar{\psi})=I(H)(\psi).$$

Proof By Lemma 7.17(1), (2) and (3), we have

$$\begin{aligned}(r^*I(H)+\bar{c}(H))(\bar{\psi}) &= (r^*I(H)+\bar{c}(H))(\bar{\psi}-\bar{\psi}_0) && \text{(by (3))}\\ &= (r^*I(H))(\bar{\psi}-\bar{\psi}_0) && \text{(by (1))}\\ &= I(H)(\psi) && \text{(by (2)).}\end{aligned}$$ □

Remark 7.19 In particular, we have $\bar{c}(H)(\bar{\psi})=0$. We can show this directly without using $\bar{\psi}_0$, if we make the argument of the proof of (1) more precise: the parameter that runs on S^{n-j-2} only affects a small neighborhood of the annulus of exactly one crossing; so if the cycle $\bar{\psi}$ has at least two crossings, the integral $\bar{c}(H)(\bar{\psi})$ vanishes by a dimensional reason.

8 Proof of the main result

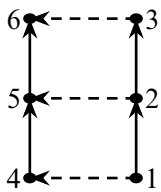
Finally, we perform the pairing between the cocycle given in Section 4 and the cycle given in Section 6, and prove Theorem 3.1 and Corollary 3.4. The following completes the proof of Theorem 3.1.

Theorem 8.1 $$\bar{z}_2^3(\bar{\psi})=1.$$

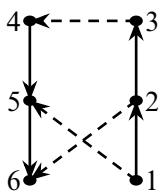
Proof By Corollary 7.18,

$$\bar{z}_2^3(\bar{\psi}) = I(H)(\psi).$$

Observe that the underlining graph of



has 4 automorphisms, and the underlining graph of



has 2 automorphisms. In the graph cocycle H,

$$w\left(\begin{array}{c}\text{6}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{3}\\ \text{5}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{2}\\ \text{4}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{1}\end{array}\right) = -\frac{1}{4} \qquad \text{and} \qquad w\left(\begin{array}{c}\text{4}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{3}\\ \text{5}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{2}\\ \text{6}\bullet\!\!-\!\!-\!\!-\!\!-\bullet\text{1}\end{array}\right) = -1.$$

Hence by the counting formula, Theorem 7.14, and Example 7.15,

$$I(H)(\psi) = -\sum_{\overline{\Gamma}\in G(C_2)} w(\Gamma)\#\operatorname{Aut}(\overline{\Gamma})s(\Gamma,\overline{\Gamma}) = -\left(-\tfrac{1}{4}\times 4\times(-1)+(-1)\times 2\times(+1+1-1)\right) = 1. \quad \square$$

Remark 8.2 We can give a simpler proof of Theorem 8.1 using graph homology [29]. We give a sketch of the proof here. Let $\mathcal{A}(k,g)$ be the graph homology generated by g-loop BCR graphs of order k. Let $\overline{w}\colon \mathcal{A}(k,g)\to\mathbb{R}$ be a weight system on it. Then

$$H = \sum_{\substack{\overline{\Gamma}\\ k(\overline{\Gamma})=k, g(\overline{\Gamma})=g}} \frac{\overline{w}(\Gamma)}{\#\operatorname{Aut}(\overline{\Gamma})}\Gamma$$

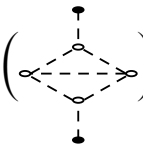
is a graph cocycle. Set

$$w(\Gamma) = \frac{\overline{w}(\Gamma)}{\#\operatorname{Aut}(\overline{\Gamma})}.$$

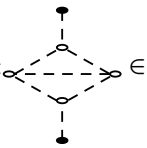
Then the counting formula, Theorem 7.14, says

$$I(H)(\psi) = (-1)^{r(C)} \sum_{\overline{\Gamma}\in G(C)} \overline{w}(\Gamma)s(\Gamma,\overline{\Gamma}).$$

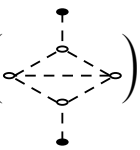
Consider the weight system $\overline{w}\colon \mathcal{A}(3,2) \to \mathbb{R}$ such that

$\overline{w}\Big($  $\Big) = \pm 1.$

Observe that the linear combination of four graphs $\sum_{\overline{\Gamma}\in G(C_2)} s(\Gamma,\overline{\Gamma})\Gamma$ gives the same class as the hairy graph

$\pm$  $\in \mathcal{A}(3,2).$

(To see this, use STU relations.) Then we have

$$I(H)(\psi) = (-1)^{r(C_2)}\overline{w}\Big(\sum_{\overline{\Gamma}\in G(C_2)} s(\Gamma,\overline{\Gamma})\Gamma\Big) = \pm\overline{w}\Big($$  $$\Big) \neq 0.$$

Finally, we prove Corollary 3.4, the case $n-j-2=0$. Let

$$\bar{\psi}(0,0,0)\colon S^{j-1} \to \overline{\mathrm{Emb}}(\mathbb{R}^j,\mathbb{R}^{j+2})$$

be the $(0,0,0)\in S^0\times S^0\times S^0$ component of $\bar{\psi}$, that is, the component where no crossing is resolved. By Proposition 6.10, $\bar{\psi}(0,0,0)$ is in the unknot component (the path component of the trivial path of the trivial immersion).

Corollary 8.3 *When $n-j=2$, $\bar{\psi}$ and $\bar{\psi}(0,0,0)$ give the same homology class. Hence, when j is odd, $\bar{\psi}$ gives an nontrivial element of*

$$\pi_*(\overline{\mathrm{Emb}}(\mathbb{R}^j,\mathbb{R}^{j+2})_\iota)\otimes\mathbb{Q}.$$

Proof Since $S_0=\{x_3=0,x_3=2\}$, we write

$$\bar{\psi} = \sum_{\epsilon_i\in\{0,2\}} \bar{\psi}(\varepsilon_1,\varepsilon_2,\varepsilon_3).$$

Observe that if one of ε_i is 2, $\bar{\psi}(\varepsilon_1,\varepsilon_2,\varepsilon_3)$ is a degenerate cycle. □

References

[1] **V I Arnold**, *The cohomology ring of the colored braid group*, Mat. Zametki 5 (1969) 227–231 MR Zbl In Russian; translated in Math. Notes 5 (1969) 138–140

[2] **G Arone**, **V Turchin**, *On the rational homology of high-dimensional analogues of spaces of long knots*, Geom. Topol. 18 (2014) 1261–1322 MR Zbl

[3] **G Arone**, **V Turchin**, *Graph-complexes computing the rational homotopy of high dimensional analogues of spaces of long knots*, Ann. Inst. Fourier (Grenoble) 65 (2015) 1–62 MR Zbl

[4] **S Axelrod**, **I M Singer**, *Chern–Simons perturbation theory, II*, J. Differential Geom. 39 (1994) 173–213 MR Zbl

[5] **D Bar-Natan**, *On the Vassiliev knot invariants*, Topology 34 (1995) 423–472 MR Zbl

[6] **R Bott**, *Configuration spaces and imbedding invariants*, Turkish J. Math. 20 (1996) 1–17 MR Zbl

[7] **R Bott**, **C Taubes**, *On the self-linking of knots*, J. Math. Phys. 35 (1994) 5247–5287 MR Zbl

[8] **R Budney**, *A family of embedding spaces*, from "Groups, homotopy and configuration spaces", Geom. Topol. Monogr. 13, Geom. Topol. Publ., Coventry (2008) 41–83 MR Zbl

[9] **R Budney**, **D Gabai**, *Knotted 3-balls in S^4*, preprint (2019) arXiv 1912.09029

[10] **A S Cattaneo**, **P Cotta-Ramusino**, **R Longoni**, *Configuration spaces and Vassiliev classes in any dimension*, Algebr. Geom. Topol. 2 (2002) 949–1000 MR Zbl

[11] **A S Cattaneo**, **C A Rossi**, *Wilson surfaces and higher dimensional knot invariants*, Comm. Math. Phys. 256 (2005) 513–537 MR Zbl

[12] **J Conant**, **J Costello**, **V Turchin**, **P Weed**, *Two-loop part of the rational homotopy of spaces of long embeddings*, J. Knot Theory Ramifications 23 (2014) art. id. 1450018 MR Zbl

[13] **B Fresse**, **V Turchin**, **T Willwacher**, *The rational homotopy of mapping spaces of E_n operads*, preprint (2017) arXiv 1703.06123

[14] **T G Goodwillie**, **J R Klein**, **M S Weiss**, *Spaces of smooth embeddings, disjunction and surgery*, from "Surveys on surgery theory, II", Ann. of Math. Stud. 149, Princeton Univ. Press (2001) 221–284 MR Zbl

[15] **T G Goodwillie**, **M Weiss**, *Embeddings from the point of view of immersion theory, II*, Geom. Topol. 3 (1999) 103–118 MR Zbl

[16] **K Habiro**, **T Kanenobu**, **A Shima**, *Finite type invariants of ribbon 2-knots*, from "Low-dimensional topology", Contemp. Math. 233, Amer. Math. Soc., Providence, RI (1999) 187–196 MR Zbl

[17] **K Habiro**, **A Shima**, *Finite type invariants of ribbon 2-knots, II*, Topology Appl. 111 (2001) 265–287 MR Zbl

[18] **M Kontsevich**, *Feynman diagrams and low-dimensional topology*, from "First European Congress of Mathematics, II", Progr. Math. 120, Birkhäuser, Basel (1994) 97–121 MR Zbl

[19] **G Kuperberg**, **D P Thurston**, *Perturbative 3-manifold invariants by cut-and-paste topology*, preprint (1999) arXiv math/9912167

[20] **C Lescop**, *On the Kontsevich–Kuperberg–Thurston construction of a configuration-space invariant for rational homology 3-spheres*, preprint (2004) arXiv math/0411088

[21] **C Lescop**, *Invariants of links and 3-manifolds from graph configurations*, preprint (2020) arXiv 2001.09929

[22] **D Leturcq**, *Generalized Bott–Cattaneo–Rossi invariants of high-dimensional long knots*, J. Math. Soc. Japan 73 (2021) 815–860 MR Zbl

[23] **R Longoni**, *Nontrivial classes in $H^*(\mathrm{Imb}(S^1, \mathbb{R}^n))$ from nontrivalent graph cocycles*, Int. J. Geom. Methods Mod. Phys. 1 (2004) 639–650 MR Zbl

[24] **M Mimura**, **H Toda**, *Topology of Lie groups: I, II*, Transl. Math. Monogr. 91, Amer. Math. Soc., Providence, RI (1991) MR Zbl

[25] **K E Pelatt**, **D P Sinha**, *A geometric homology representative in the space of knots*, from "Manifolds and K–theory", Contemp. Math. 682, Amer. Math. Soc., Providence, RI (2017) 167–188 MR Zbl

[26] **K Sakai**, *Nontrivalent graph cocycle and cohomology of the long knot space*, Algebr. Geom. Topol. 8 (2008) 1499–1522 MR Zbl

[27] **K Sakai**, *Configuration space integrals for embedding spaces and the Haefliger invariant*, J. Knot Theory Ramifications 19 (2010) 1597–1644 MR Zbl

[28] **K Sakai**, *An integral expression of the first nontrivial one-cocycle of the space of long knots in* $\mathbb{R}^3$, Pacific J. Math. 250 (2011) 407–419 MR Zbl

[29] **K Sakai**, **T Watanabe**, *1-loop graphs and configuration space integral for embedding spaces*, Math. Proc. Cambridge Philos. Soc. 152 (2012) 497–533 MR Zbl

[30] **D P Sinha**, *Manifold-theoretic compactifications of configuration spaces*, Selecta Math. 10 (2004) 391–428 MR Zbl

[31] **D P Sinha**, *The (non-equivariant) homology of the little disks operad*, from "OPERADS 2009", Sémin. Congr. 26, Soc. Math. France, Paris (2013) 253–279 MR Zbl

[32] **T Watanabe**, *Configuration space integral for long n-knots and the Alexander polynomial*, Algebr. Geom. Topol. 7 (2007) 47–92 MR Zbl

[33] **T Watanabe**, *On Kontsevich's characteristic classes for higher dimensional sphere bundles, I: The simplest class*, Math. Z. 262 (2009) 683–712 MR Zbl

[34] **T Watanabe**, *On Kontsevich's characteristic classes for higher-dimensional sphere bundles, II: Higher classes*, J. Topol. 2 (2009) 624–660 MR Zbl

[35] **T Watanabe**, *Higher order generalization of Fukaya's Morse homotopy invariant of 3-manifolds, I: Invariants of homology 3-spheres*, Asian J. Math. 22 (2018) 111–180 MR Zbl

[36] **T Watanabe**, *Some exotic nontrivial elements of the rational homotopy groups of* $\mathrm{Diff}(S^4)$, preprint (2018) arXiv 1812.02448

[37] **T Watanabe**, *Theta-graph and diffeomorphisms of some 4-manifolds*, preprint (2020) arXiv 2005.09545

[38] **M Weiss**, *Embeddings from the point of view of immersion theory, I*, Geom. Topol. 3 (1999) 67–101 MR Zbl

[39] **E Witten**, *Quantum field theory and the Jones polynomial*, Comm. Math. Phys. 121 (1989) 351–399 MR Zbl

Graduate School of Mathematical Sciences, The University of Tokyo
Tokyo, Japan

yoshioka@ms.u-tokyo.ac.jp

Received: 28 February 2023 Revised: 10 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3429–3501
DOI: 10.2140/agt.2025.25.3429
Published: 1 October 2025

Asymptotic cones of snowflake groups and the strong shortcut property

CHRISTOPHER H CASHEN
NIMA HODA
DANIEL J WOODHOUSE

We exhibit an infinite family of snowflake groups all of whose asymptotic cones are simply connected. Our groups have neither polynomial growth nor quadratic Dehn function, the two usual sources of this phenomenon. We further show that each of our groups has an asymptotic cone containing an isometrically embedded circle or, equivalently, has a Cayley graph that is not strongly shortcut. These are the first examples of groups whose asymptotic cones contain "metrically nontrivial" loops but no topologically nontrivial ones.

20F65, 20F69, 51F30

1. Introduction 3429
2. Preliminaries 3433
3. Snowflake groups 3436
4. The snowflake Cayley graph is not strongly shortcut 3448
5. The idea of the proof that asymptotic cones are simply connected 3451
6. Reduction to bi-Lipschitz cycles 3455
7. Bi-Lipschitz cycles in snowflake groups 3457
8. Constructing fillings 3483
9. Proof of the main theorem and closing remarks 3498
References 3499

1 Introduction

The strong shortcut property was introduced by the second named author in order to explore commonalities between the many theories of nonpositively curved groups which have been developed in recent decades [25]. A graph Γ is *strongly shortcut* if, for some $K > 1$, there is a bound on the lengths of the K-bi-Lipschitz cycles of Γ. This property has a natural generalization to rough geodesic metric spaces [26]. A group G is *strongly shortcut* if it satisfies one of the following three equivalent conditions [26]:

(1) G has a strongly shortcut Cayley graph.

(2) G acts properly and cocompactly on a strongly shortcut graph.

(3) G acts metrically properly and coboundedly on a strongly shortcut rough geodesic metric space.

© 2025 The Authors, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Strongly shortcut groups are finitely presented and have polynomial isoperimetric functions [25].

The class of strongly shortcut spaces is vast. It includes:

- Asymptotically CAT(0) spaces [26; 28]; in particular, CAT(0) spaces [12; 20], Gromov-hyperbolic spaces [20] and $\widetilde{\mathrm{SL}(2,\mathbb{R})}$ with the Sasaki metric [28].
- 1-skeletons of systolic [15; 23; 27; 41], quadric [3; 24] and finite dimensional CAT(0) cubical [2; 20; 33] complexes [25].
- Standard Cayley graphs of Coxeter groups [25; 34].
- Coarsely injective spaces of uniformly bounded geometry [22]; in particular, Helly graphs [14] of bounded degree.
- Discrete Heisenberg groups of all dimensions; follows from the asymptotic cone characterization of the strong shortcut property (Theorem A of [26]) and the main result of [30] which implies that there are no isometric circles in the asymptotic cones of real Hesienberg groups.

In particular, all Thurston geometries except Sol are strongly shortcut. The class of strongly shortcut groups is correspondingly vast, including CAT(0) groups, hyperbolic groups, Helly groups, systolic groups, quadric groups, the discrete Heisenberg groups and hierarchically hyperbolic groups [5; 6; 22]. In particular, finitely presented small cancellation groups [14; 24; 36; 42] and mapping class groups [31; 32] of surfaces are strongly shortcut.

There is a characterization of strongly shortcut spaces in terms of asymptotic cones [26] (see Section 2.3 for definitions): a space is strongly shortcut if and only if none of its asymptotic cones contains an isometrically embedded copy of a unit length circle with its length metric. One might wonder whether, for Cayley graphs of groups, this condition is equivalent to all asymptotic cones being simply connected. We provide the first counterexamples.

We consider an infinite family of *snowflake groups*,

$$G_L := \langle a, x, y, s, t \mid sas^{-1} = x, tat^{-1} = y, a^L = xy, [a, x] = [x, y] = [y, a] = 1 \rangle,$$

where $L \geq 6$ is an even integer. These are special cases[1] of the groups introduced by Brady and Bridson [9], who show that the Dehn function for such a group is $n^{2\alpha}$ for $\alpha := \log_2 L$.

Our main results are:

Theorem 9.0.1 *For all even $L \geq 6$, every asymptotic cone of G_L is simply connected.*

Corollary 4.0.3 *For all even $L \geq 6$, some asymptotic cone of G_L contains an isometrically embedded circle.*

[1] Specifically, $G_{p,1}$ of Brady and Bridson is G_{2p} in our notation, which is why the parameter L is even. The restriction to $L \geq 6$ is explained in Remark 3.1.2.

We find an isometrically embedded circle in an asymptotic cone of the Cayley graph of G_L corresponding to the generating set $\{a, s, t\}$. This means that this particular Cayley graph is not a shortcut graph. We do not know if G_L is a strongly shortcut group — it is possible that it admits an action on some other graph that is strongly shortcut. If G_L is a strongly shortcut group then this is the first example of a group for which the strong shortcut property depends on the choice of generating set. If G_L is not a strongly shortcut group then this is the first example of a group for which every asymptotic cone being simply connected does not imply the strong shortcut property.

The structures of asymptotic cones of snowflake groups are interesting independent of the strong shortcut property. An application of having all cones simply connected is due to Riley [39, Theorem C]: the isodiametric and filling length functions of G_L are both linear. It is also interesting just to have more diverse examples of groups with all cones simply connected. As far as we are aware, the only existing examples are groups of polynomial growth, groups with at most quadratic Dehn function, and certain combinations of these; see Section 2.3. Our examples are not in any of these classes. We also do not know any other examples of groups having all of their asymptotic cones simply connected, but with at least some of them containing isometrically embedded circles. Conjecturally there are no such groups of polynomial growth. It is an open question whether there exist such groups with quadratic Dehn function.

The idea of the proof that asymptotic cones are simply connected There is a criterion, the "loop subdivision property", which we recall in Theorem 2.3.2, for a group to have all of its asymptotic cones simply connected. This says that every asymptotic cone is simply connected if and only if every word in the generators representing the trivial element of the group can be filled by a Van Kampen diagram of uniformly bounded area, subject to the constraint that the boundary lengths of 2-cells of the diagram are bounded above by a fixed fraction of the length of the word to be filled.

To construct such diagrams, we establish a fat/thin dichotomy. The "fat" words are those such that the corresponding loop in the Cayley graph is a K-bi-Lipschitz loop for K close to 1. Thin words we cut into a bounded number of small pieces and at most one fat one using the "circle tightening lemma" of [26]. This reduces the problem to constructing diagrams for K-bi-Lipschitz loops for K arbitrarily close to 1. This reduction step occurs in Section 6.

For $K = 1$, ie geodesic loops, there are some natural example words to consider: the "snowflake words" that give the groups their names. We describe how to fill loops made from snowflake words in Section 5.1.

In the general case our thesis is that, for K sufficiently close to 1, all K-bi-Lipschitz loops "ought to look like" a snowflake loop, and admit a filling diagram using a similar construction. In retrospect, this is more a declaration of naïveté than adroitness. It turns out to be true, but it was much harder than we had initially expected to make rigorous.

Why is it hard? Consider these two facts:

Fact 1.0.1 (reverse Hölder inequality) For any $\alpha > 1$ and any $r_1, r_2, \dots, r_n \geq 0$,

$$\left(\frac{1}{n}\right)^{1-1/\alpha}(r_1^{1/\alpha} + \cdots + r_n^{1/\alpha}) \leq (r_1 + \cdots + r_n)^{1/\alpha} \leq r_1^{1/\alpha} + \cdots + r_n^{1/\alpha}.$$

Fact 1.0.2 (power-law distortion, cf [9] or Theorem 3.2.9) Given even $L \geq 6$ and setting $\alpha := \log_2 L$, there exist C_{lo} and C_{hi} such that for $g \in \{a, x, y\}$ and all $|m| > 0$,

$$C_{\text{lo}} \leq \frac{|g^m|}{|m|^{1/\alpha}} < C_{\text{hi}}.$$

We will show in Section 3.2 that Fact 1.0.2 is about as good as can be expected; the relationship between $|m|$ and $|g^m|$ is not monotone.

Taken together, these two facts suggest that to travel between two points of a distorted subgroup like $\langle a \rangle$, it is more efficient to make one long hop rather than several short ones; the distortion is nonlinear, so the long hop can achieve proportionally more savings. For example, consider for $m, n > 0$ a path between 1 and a^{m+n} consisting of a geodesic from 1 to a^m, followed by a geodesic from a^m to a^{m+n}. We expect

$$\frac{|a^m| + |a^n|}{|a^{m+n}|} \sim \frac{m^{1/\alpha} + n^{1/\alpha}}{(m+n)^{1/\alpha}} = \frac{1 + (n/m)^{1/\alpha}}{(1 + n/m)^{1/\alpha}}.$$

This ratio goes to 1 as n/m goes to 0 or infinity, but equals $2^{1-1/\alpha}$ when $m = n$. So the expectation should be that the path is very efficient, in the sense that the ratio of its length to the distance between its endpoints is very close to 1, if and only if one of the two subpaths dominates the other, in the sense that n/m is either very large or very small. If this were true it would give us good control over bi-Lipschitz paths when the bi-Lipschitz constant is close to 1; they would, more or less, have to be dominated by a long segment that makes good use of the distortion, possibly with some relatively very short additional segments at the beginning and end.

However, Fact 1.0.2 is too rough an estimate to confirm this intuition. It can only tell us, for instance, that

$$\frac{|a^m| + |a^m|}{|a^{2m}|} > \frac{C_{\text{lo}}}{C_{\text{hi}}} \cdot 2^{1-1/\alpha},$$

but $(C_{\text{lo}}/C_{\text{hi}}) \cdot 2^{1-1/\alpha}$ is less than 1 when $L \geq 10$ (see Proposition 3.2.10), so this bound is useless. The deviation of the distortion function from being a power-law on the nose overwhelms the efficiency estimates coming from reverse Hölder. We must make finer estimates. The bulk of the paper is a detailed analysis of bi-Lipschitz paths in snowflake groups.

The plan of the paper Section 2 is background material. In Section 3 we analyze geodesics between elements in the vertex group of the snowflake group. In Section 4 we show that snowflake loops are geodesic loops, and conclude that our Cayley graph is not strongly shortcut.

In Section 5 we describe how snowflake loops satisfy the loop subdivision property and give a more detailed explanation of how we will adapt this construction to the general case. The first step, in Section 6,

is to reduce to the case of bi-Lipschitz loops with bi-Lipschitz constant sufficiently close to 1. We analyze such loops in Section 7, and construct fillings of them in Section 8. We tie things up and make some closing remarks in Section 9.

Acknowledgements

C Cashen is supported by the Austrian Science Fund (FWF): P30487-N35 and P34214-N, and visited the University of Arkansas during part of this work. N Hoda is supported by the ERC grant GroIsRan.

2 Preliminaries

2.1 Basics

If S is a finite set, let $F(S)$ denote the free group generated by S.

If G is a group generated by a finite set S, then the *word metric* on G with respect to S is defined by taking $|g|_S$ to be the minimal length of a word $w \in F(S)$ such that $w =_G g$. The *Cayley graph* $\mathrm{Cay}(G, S)$ of G with respect to S is the labelled directed graph whose vertices are in bijection with G, and such that there is an edge from g to h labelled s when $h = gs$. We say that a word in $F(S)$ labels an edge path, or is read on an edge path, if it is the concatenation of labels of the edges, with the convention that an s edge traversed against its orientation contributes the letter s^{-1}. Given a word $w \in F(S)$ and a basepoint $g \in G$ there is a unique edge path starting from g labelled by w.

Paths in $\mathrm{Cay}(G, S)$ will always start and end at vertices and will always be parameterized by arclength. If γ is a path, $\bar{\gamma}$ is the path with opposite parameterization. We use $+$ to denote concatenation. We frequently conflate paths with their labels, but there will be only one reasonable interpretation. For example, if γ is a path in $\mathrm{Cay}(G, S)$ from g to h, with $g, h \in G$, δ is another path in $\mathrm{Cay}(G, S)$ whose endpoints differ by $k \in G$, and $s, t \in S$ then $s.\gamma + \bar{\delta} + st^{-1}$ is a path from sg to $shk^{-1}st^{-1}$ while $s + \gamma + \bar{\delta} + st^{-1}$ is a path from 1 to $sg^{-1}hk^{-1}st^{-1}$.

Let $\langle S \mid \mathcal{R}\rangle$ be a finite presentation of a group G. If $w =_G 1$ then w can be written in $F(S)$ as $w = \prod_{i=1}^{n} u_i r_i u_i^{-1}$ for some u_i and r_i such that either r_i or r_i^{-1} belongs to $\mathcal{R}$. Define $\mathrm{Area}(w)$ to be the minimal such n. The *Dehn function* or *isoperimetric function* of the presentation is the function $n \mapsto \max_{|w|_S \le n,\, w =_G 1} \mathrm{Area}(w)$. While the precise function depends on the choice of presentation, it turns out that up to a certain notion of equivalence of functions all of the finite presentations yield equivalent functions, so we define the *Dehn function of the group* to be this equivalence class. The equivalence relation preserves, in particular, degrees of polynomial growth, so it makes sense to say that a group has a Dehn function of a particular polynomial degree.

2.2 Van Kampen diagrams

Definition 2.2.1 Let $\mathcal{A}$ be an alphabet. A *Van Kampen diagram* for a word $w \in \mathcal{A}^*$ over a subset $\mathcal{R} \subset \mathcal{A}^*$ of cyclically reduced words is a compact, connected, simply connected 2-dimensional cell complex D

embedded in the plane such that edges are directed and labelled by elements of $\mathscr{A}$ in such a way that starting from some base boundary vertex the word read around ∂D is w and the word read around the boundary of each 2-cell is freely and cyclically reduced and, up to cyclic permutation and inversion, belongs to $\mathscr{R}$.

A 2-cell of D is also called a *region*.

A Van Kampen diagram is *reduced* if there does not exist a pair of 2-cells sharing an edge such that the boundary labels starting from a shared edge are inverses.

The *area* of a Van Kampen diagram is the number of 2-cells.

A Van Kampen diagram is *minimal* or has *minimal area* if it has the fewest number of 2-cells among all Van Kampen diagrams for the given boundary word. Minimal diagrams are necessarily reduced.

A *disk diagram* is a Van Kampen diagram that is topologically a disk.

An *arc* is a maximal segment in the 1-skeleton of D such that every interior vertex has valence 2. For the purpose of drawing pictures it is often convenient to replace each arc by a single directed edge labelled by the word in $\mathscr{A}^*$ read along the arc. In this case the length of the new edge is defined to be the length of its label, that is, the length of the original arc. This is called *forgetting the valence* 2 *vertices*.

The *mesh* of a diagram is the maximum length of the boundary of a 2-cell.

Given a specified base vertex in D, there is a unique label and orientation preserving map $D^{(1)} \to \mathrm{Cay}(G, \mathscr{A})$ sending the base vertex to 1, where $G = \langle \mathscr{A} \mid \mathscr{R} \rangle$. Thus, we sometimes say that D gives a *filling* for the word read around ∂D, which is a word $w \in F(\mathscr{A})$ such that $w =_G 1$.

The groups we are interested in are double HNN extensions over $\mathbb{Z}$; that is, groups G having a presentation of the form $\langle a_1, \dots, a_k, s, t \mid \mathscr{R}, sus^{-1}v^{-1}, twt^{-1}x^{-1} \rangle$, where $H := \langle a_1, \dots, a_k \mid \mathscr{R} \rangle$ is the *vertex group*, s and t are called *stable letters*, and u, v, w, and x are infinite order elements of H. It is a standard result of Bass–Serre theory that H is a subgroup of the resulting group G.

Let $\mathscr{R}' := \mathscr{R} \cup \{sus^{-1}v^{-1}, twt^{-1}x^{-1}\}$. Since each stable letter appears in only one relator, and appears as an inverse pair, we can say more about the structure of reduced diagrams over $\mathscr{R}'$. Suppose D is a reduced diagram and some region has an s-edge in its boundary. Either this edge belongs to ∂D, or there is an adjacent region that shares the s-edge. However, there are only two possible occurrences of $s^{\pm}$ in $\mathscr{R}$, and one of the possible choices would create an unreduced diagram, so there is a unique way that the boundary of the adjacent region can be labelled. That region has one more edge labelled with s or s^{-1}, which is either on ∂D or else is adjacent to another region whose boundary label is determined, etc. Define an equivalence relation on 1-cells of D labelled $s^{\pm}$ such that two such edges are equivalent if they belong to a common 2-cell. For each equivalence class of s-edge, define the *s-corridor* containing it to be the union of cells containing one of those edges in its closure. We define t-corridors similarly, and just say *corridor* if we do not care which of the stable letters the corridor corresponds to. It is immediate from the definitions that distinct corridors have disjoint interiors.

A priori there are three possible types of s-corridor: a single s-edge, a closed loop of 2-cells, or a string of 2-cells ending at distinct edges of ∂D. The first type cannot occur if D is a disk diagram. This is true in particular when ∂D embeds in X. The second type cannot occur in a minimal diagram because if it did the subdiagram inside the closed corridor would be a diagram filling a power of u, v, w, or x, but these were assumed to be infinite order elements. The same argument shows that not only does a corridor have distinct edges of ∂D, these two edges are disjoint.

Define an *HNN diagram* (with respect to the above presentation as a multiple HNN extension) as a minimal diagram over the set of relations

$$\{g \in F(\{a_i\}_{i=1}^k) \mid g =_G 1\} \cup \{su^n s^{-1} v^{-n} \mid n \in \mathbb{N}\} \cup \{tw^n t^{-1} x^{-n} \mid n \in \mathbb{N}\}.$$

Regions with boundary labelled by an element of the latter two sets correspond to corridors in a diagram over $\mathcal{R}$, and we again call them *corridors*. Regions labelled by element of the first set correspond to vertices in the tree dual to the corridors in a diagram over $\mathcal{R}$. We call such a region a *vertex region*.

2.3 Asymptotic cones

We briefly recall the definition and some history of asymptotic cones. See [10; 17; 39] for more.

Definition 2.3.1 Let (X, d) be a metric space, ω a nonprincipal ultrafilter, $\boldsymbol{s} = (s_n) \to \infty$ a "scaling sequence" of positive numbers, and $\boldsymbol{o} = (o_n)$ a sequence of "observation points" in X.

Define a pseudometric on sequences $\boldsymbol{x} = (x_n)$ of points in X as follows.

$$d_{\omega,\boldsymbol{s}}(\boldsymbol{x}, \boldsymbol{y}) := \lim_{\omega} \frac{d(x_n, y_n)}{s_n}$$

Define the asymptotic cone $\mathrm{Cone}_\omega(X, \boldsymbol{o}, \boldsymbol{s})$ of X with respect to ω, $\boldsymbol{s}$, and $\boldsymbol{o}$ to be the quotient of the space of sequences $\boldsymbol{x}$ at finite $d_{\omega,\boldsymbol{s}}$-distance from $\boldsymbol{o}$ by identifying sequences at $d_{\omega,\boldsymbol{s}}$-distance 0 from one another.

It turns out that asymptotic cones are complete metric spaces.

We are interested in the case that X is the Cayley graph of a group, in which case X is homogeneous. In this case the choice of observation points does not matter, and we can take $\boldsymbol{o} = 1$ to be the constant sequence at the identity vertex.

Gromov [19] used a process of passing to a convergent subsequence of rescaled metric spaces in his proof of the polynomial growth theorem. Outside of the polynomial growth case, convergence fails, but Van den Dries and Wilkie [16] recognized that this could be fixed by the imposition of an ultrafilter.

Pansu [37] proved nilpotent groups have simply connected asymptotic cones. Gromov [21, $5F_1''$] showed that if every asymptotic cone of G is simply connected, then G is finitely presented and has polynomial Dehn function. Papasoglu [38] proved the converse for groups with quadratic Dehn function, but it is not true in general [11; 40; 35], even for groups with cubic Dehn function.

Druțu and Sapir characterized relatively hyperbolic groups as having asymptotic cones that are tree-graded with respect to the asymptotic cones of their peripheral subgroups [18] and used this notion to give quasi-isometric rigidity results. Behrstock [4] showed that the asymptotic cone of the mapping class group is tree graded, although it is not relatively hyperbolic. The tree graded structure of the asymptotic cones of the mapping class group were further studied by Behrstock and Minsky [8] and Behrstock, Kleiner, Minsky and Mosher [7] as part of proving quasi-isometric rigidity.

Kent [29] defines a *prairie group* to be one for which every asymptotic cone is simply connected. As far as we are aware, the only groups currently known to be prairie groups are either virtually nilpotent or have quadratic Dehn function or are built by combining such pieces: direct products of prairie groups are prairie groups, and groups that are hyperbolic relative to prairie groups are prairie groups [18].

Snowflake groups are not virtually nilpotent, have superquadratic Dehn functions, do not split as direct products, and are not hyperbolic relative to any collection of subgroups. We show that snowflake groups are prairie groups using a criterion suggested by Gromov and explained in detail by Papasoglu [38, Section 1]. As expressed by Riley [10] (cf [39]):

Theorem 2.3.2 (loop subdivision property [10, Theorem II.4.3.1]) *Let G be a group with finite generating set S. Fix a nonprincipal ultrafilter ω. The following are equivalent*:

(1) *The asymptotic cone $\mathrm{Cone}_\omega(\mathrm{Cay}(G,S),1,\boldsymbol{s})$ is simply connected for all scaling sequences $\boldsymbol{s}\to\infty$.*

(2) *Given $\lambda\in(0,1)$ there exist $N, J\in\mathbb{N}$ such that for all null-homotopic words w there is an equality*

$$w=\prod_{i=1}^{N} u_i w_i u_i^{-1}$$

in the free group $F(S)$ for some words u_i and w_i such that the w_i are null homotopic and have length $\ell(w_i)\leq\lambda\ell(w)+J$.

(3) *Given $\lambda\in(0,1)$ there exist $A, M\in\mathbb{N}$ such that for all words $w\in F(S)$ with $w=_G 1$ there is a Van Kampen diagram for w over $\mathcal{R}=\{g\in F(S)\mid g=_G 1\}$ with area at most A and mesh at most $\lambda|w|+M$.*

Remark 2.3.3 Change of finite generating set gives a bi-Lipschitz map between Cayley graphs, which in turn induces a bi-Lipschitz homeomorphism between asymptotic cones for a common ultrafilter and scaling sequence, so if all asymptotic cones of $\mathrm{Cay}(G,S)$ are simply connected, then the same is true with respect to any finite generating set of G.

3 Snowflake groups

We consider the following family of groups for $L\geq 6$ and even:

(3.0.1) $$G_L:=\langle a,x,y,s,t\mid sas^{-1}=x, tat^{-1}=y, a^L=xy, [a,x]=[x,y]=[y,a]=1\rangle.$$

Such a group is isomorphic, via $x = a^p b$ and $y = a^p b^{-1}$, to the Brady–Bridson snowflake group [9] $G_{p,q}$ with $p = L/2$ and $q = 1$ defined by

$$\textbf{(3.0.2)} \qquad G_{p,q} := \langle a, b, s, t \mid [a, b] = 1, sa^q s^{-1} = a^p b, ta^q t^{-1} = a^p b^{-1} \rangle.$$

The group $G = G_L$ is a double HNN extension over $\mathbb{Z}$ of the subgroup $H := \langle a, x, y \rangle \cong \mathbb{Z}^2$. Note that $\langle a \rangle$, $\langle x \rangle$, and $\langle y \rangle$ pairwise have trivial intersections since $H \cong \mathbb{Z}^2$, and that $H = \coprod_{i=0}^{L-1} a^i \langle x, y \rangle$. Let $X = X_L$ denote the Cayley graph of G_L with respect to the generating set $\{a, x, y, s, t\}$, with the following modification: we declare that the a, s and t edges have length 1 as usual, but assign length L to the x and y edges.[2] This follows the convention of [1] and leads to a simpler description of geodesics. This is not a significant deviation from the standard conventions, because the induced metric on vertices is still a word metric: it is the one corresponding to the generating set $\{a, s, t\}$. Metrically the x and y edges are extraneous. We include them to preserve the correspondence between the group structure as an HNN extension and the topological structure of the universal cover of the corresponding graph of spaces, which is a torus with two annuli glued on. In this correspondence we have, for instance, that cosets of $\langle a, x, y \rangle$ correspond to the planes that are the connected components of the preimage of the torus in the universal cover, so we would like to think of cosets of $\langle a, x, y \rangle$ as connected subsets of X, even if they are not convex.

Brady and Bridson introduced snowflake groups to furnish examples of groups with interesting power-law Dehn functions: the Dehn function of $G_{p,q}$ is $n^{2\alpha}$ where $\alpha := \log_2(2p/q)$.

For each $n \geq 1$ we inductively construct a pair of paths $\sigma_{n,s}$ and $\sigma_{n,t}$ from 1 to a^{L^n} by

$$\begin{aligned} \sigma_{1,s} &:= sas^{-1}tat^{-1}, \\ \sigma_{1,t} &:= tat^{-1}sas^{-1}, \\ \sigma_{n+1,s} &:= s + \sigma_{n,s} + s^{-1} + t + \sigma_{n,s} + t^{-1}, \\ \sigma_{n+1,t} &:= t + \sigma_{n,t} + t^{-1} + s + \sigma_{n,t} + s^{-1}. \end{aligned}$$

We call $\sigma_{n,s}$ and $\sigma_{n,t}$ *snowflake paths* and call $\sigma_{n,s} + \bar{\sigma}_{n,t}$ a *snowflake loop*. The snowflake loops are precisely the loops that Brady and Bridson showed were hard to fill, witnessing the Dehn function. An HNN diagram for $\sigma_{7,s} + \bar{\sigma}_{7,t}$ is illustrated in Figure 1. (The central diamond and all triangular regions are vertex regions, and the trapezoidal regions are corridors.)

We will show in Theorem 4.0.1 that the snowflake loops are geodesic. In particular, we can then compute lengths

$$\textbf{(3.0.3)} \qquad |a^{L^n}| = 5 \cdot 2^n - 4 \quad \text{if } n > 0.$$

[2] For the purposes of this paper one may choose any length at least 6 for the length of the x and y edges, since this is enough to guarantee that no 2-bi-Lipschitz path contains such an edge. We choose L because it is the length induced by the maximally symmetric metric for this group as in [13].

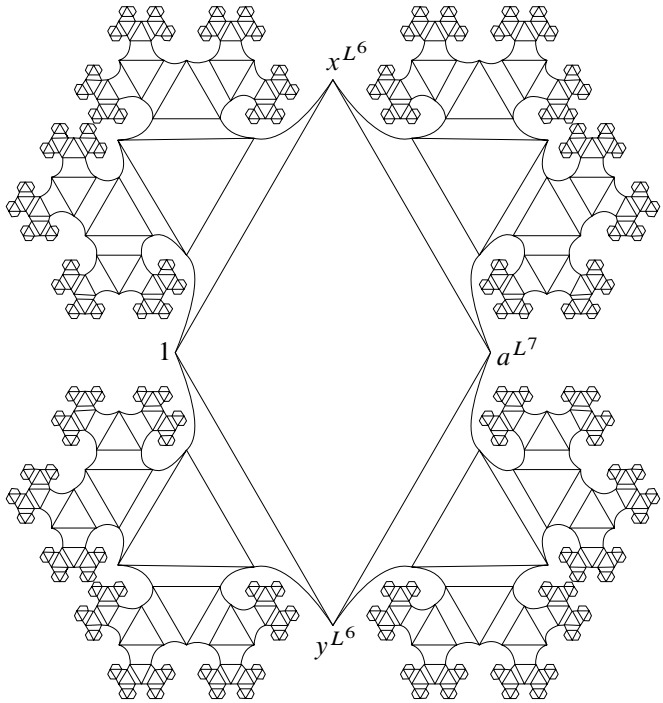


Figure 1: A snowflake for a^{L^7}.

3.1 A first result about geodesics

Recall that $H = \langle a, x, y\rangle \cong \mathbb{Z}^2$ is the vertex group of G. With respect to a fixed coset gH of H, an *escape* is a nontrivial path $\gamma\colon P \to X$ that intersects gH precisely at its endpoints. The difference between the endpoints of such a γ is a power of either a, x, or y, and we call γ an *a-escape*, an *x-escape* or a *y-escape*, respectively. Note that there are two types of a-escape: paths of the form $s^{-1}+\gamma+s$ or of the form $t^{-1}+\gamma+t$. On the other hand, every x-escape is of the form $s+\gamma+s^{-1}$, and every y-escape is of the form $t+\gamma+t^{-1}$.

A *toral path*, with respect to gH, is an edge path contained in gH. By our choice of metric on X, a toral path whose bi-Lipschitz constant is less than 2 does not contain any x or y edges, so it is just an a-path.

If γ is a single escape, then its endpoints differ by g^n for some $g \in \{a, x, y\}$ and some $n \in \mathbb{Z}$. Define the *trace* of γ to be the toral path between the endpoints of γ consisting of n-many g-edges. Define the *exponent* of γ to be n.

Given a path γ with endpoints on gH we can decompose it into a concatenation $\gamma = \gamma_1 + \cdots + \gamma_n$ of escapes and toral subpaths. We then define the trace of γ (in gH) by replacing each escape by its trace.

Since H is abelian, for any permutation σ of $\{1, \dots, n\}$ there are translations $g_i \in H$ that give a new path $\gamma' = g_1.\gamma_{\sigma(1)} + \cdots + g_n.\gamma_{\sigma(n)}$ with the same length and the same endpoints as γ. *Consolidating toral subpaths*, *x-escapes*, *etc* means choosing the permutation σ so that in γ' all of the toral subpaths,

x-escapes, etc are adjacent. Thus, by consolidating toral subpaths we may replace γ by a path γ' with the same endpoints that is a concatenation of escapes followed by a single toral subpath, for instance.

Notice that our presentation (3.0.1) for G has a symmetry that exchanges x and y and fixes $\langle a\rangle$. This induces an isometry on X, which we refer to as an *x/y-symmetry*.

Proposition 3.1.1 *For all $h \in H = \langle a, x, y\rangle$, every geodesic from 1 to h consists of a concatenation, in some order, of at most one x-escape, at most one y-escape, and some number of toral subpaths that are all of the form a^{ℓ_i}, where all of the ℓ_i have the same sign. If $L > 6$ or if $L = 6$ and $h \notin \{a^{-10}, a^{-9}, \dots, a^{10}\}$ then $|\sum_i \ell_i| < L$. If $L = 6$ and $1 \neq h \in \{a^{-10}, a^{-9}, \dots, a^{10}\}$ then the a-path from 1 to h is a geodesic but there also exists a geodesic consisting of at most one x-escape, at most one y-escape, and a toral path that is an a-path of exponent strictly less than L.*

Consequently, there is a geodesic from 1 to h of the form $\gamma_x + \gamma_y + \gamma_a$ (or any permutation of the order of these three paths) where γ_x is trivial or a geodesic x-escape, γ_y is trivial or a geodesic y-escape, and γ_a is a geodesic a-path of length less than L.

In the special case of $h = x^m y^n \neq 1$ every geodesic from 1 to h is of the form $\gamma_x + \gamma_y$ or $\gamma_y + \gamma_x$ where γ_x has exponent m and γ_y has exponent n, except in the case that $L = 6$ and $m = n = \pm 1$ so that $h = a^{\pm L}$, in which case there is such a geodesic, but the a-path from 1 to h is also geodesic.

Remark 3.1.2 One sees in Proposition 3.1.1 that there is a "corner case" of low L and small powers of a that requires extra attention when describing geodesics. The situation is even worse when $L = 4$, so we have chosen to omit G_4 completely rather than deal with additional special cases. We suspect that the main results of the paper are still true for G_4. The group G_2 has quadratic Dehn function, so we already know its asymptotic cones are simply connected by Papasoglu's theorem.

Proof A geodesic γ from 1 to h can be subdivided into toral subpaths and escapes, each of which we may assume is geodesic, since otherwise we would find a shortening of γ.

If there is more than one x-escape then by consolidating x-escapes we can replace γ by a path of the same length with the same endpoints that has adjacent x escapes. However, since x escapes are of the form $s + \gamma' + s^{-1}$, if there are two adjacent then we get an s-cancellation, so there is a shorter path from 1 to h, contradicting the fact that we started with a geodesic. By the same reasoning, there is at most one y-escape.

Suppose that there exist nontrivial geodesics that are not a single x-escape and whose endpoints differ by a power of x. Consider a shortest such geodesic, γ. By what we have already argued, γ consists of at most one x-escape, at most one y-escape, and paths, either toral path or a-escapes, whose endpoints differ by powers of a. By the "shortest" hypothesis, γ does not contain an x-escape, because if it did we could remove it and get a shorter path whose endpoints still differ by a power of x. Let x^m be the difference between the endpoints of γ, and y^n be the difference between endpoints of the y-escape.

Since the remainder of γ is paths whose endpoints differ by powers of a, we get that $x^m = a^\ell y^n$. Now, $x^m y^m = a^{mL}$, so $y^{m+n} = a^{mL-\ell}$. Since $\langle a\rangle \cap \langle y\rangle = 1$, this is only possible when $m+n = mL-\ell = 0$. However, this gives us

$$|x^m| = |\gamma| = |y^m| + |a^{mL}|.$$

By x/y-symmetry, $|x^m| = |y^m|$, which implies $mL = 0$, a contradiction. Thus, every geodesic whose endpoints differ by a power of x is a single x-escape. The corresponding statement for y follows from x/y-symmetry.

From this it follows that there are no geodesic a-escapes, for if there were we would have a geodesic γ such that, up to x/y-symmetry, $\gamma = s^{-1} + \gamma' + s$, where γ' is a geodesic whose endpoints differ by a power of x. But we have just shown that such a geodesic is a single x-escape, so $\gamma' = s + \gamma'' + s^{-1}$, which contradicts that γ is geodesic.

By consolidating toral subpaths we may replace γ by a path with the same length and the same endpoints in which all of the toral subpaths are adjacent. If any of them have different signs then there is an a-cancellation, which contradicts γ being a shortest path from 1 to h. Thus, we may suppose γ has a single toral subpath of the form a^ℓ for $|\sum_i \ell_i| = \ell \geq 0$. The toral subpath must be geodesic, since otherwise we could shorten γ. We have $|a^L| = |xy| = |sas^{-1}tat^{-1}| \leq 6$, so if $L > 6$ then no a-path of exponent at least L is geodesic. If $L = 6$ then a similar argument shows $|a^{12}| \leq 8$, so no a-path of exponent at least 11 is geodesic. Furthermore, if $\ell \geq 6$ then we can replace the toral subpath by the path $sas^{-1}tat^{-1}a^{\ell-6}$ to get a geodesic γ' with the same endpoints as γ. If $h \notin \{a^{-10}, a^{-9}, \dots, a^{10}\}$ then γ could not have been purely toral, because a geodesic toral path has endpoints that differ by a power of a with exponent less than 11 when $L = 6$. Since a-escapes are not geodesic, γ must have already had at least one x-escape or y-escape. Since γ' has one more x-escape and one more y-escape than γ, it has at least two of some flavor of escape, which contradicts the fact that γ' is geodesic. This leaves for our consideration the case that $h = a^\ell$ for $0 < \ell \leq 10$. Suppose there is a geodesic γ from 1 to h that is not toral. Then it contains either an x or y-escape, but since the endpoints are in $\langle a\rangle$ it must contain both with equal exponents. The minimum length of an escape is 3, so $|\gamma| \geq 6$, with equality only if $\ell = 6$. Thus, if $\ell < 6$ then the a-path from 1 to a^ℓ, which has exponent strictly less than L, is the unique geodesic. If $6 \leq \ell \leq 10$ and the escapes of γ have minimal length then, up to reordering, $\gamma = sas^{-1}tat^{-1}a^{\ell-6}$, which has length ℓ, so is the same length as the a-path. If the escapes are not minimal length then $\ell \geq |\gamma| \geq 8 + |a^{\ell-6p}|$ for some $p \geq 2$. The only solution, since $0 < \ell \leq 10$, is when $\ell = 10$ and $p = 2$, in which case the paths labelled a^{10}, $sas^{-1}tat^{-1}a^4$, and $sa^2s^{-1}ta^2t^{-1}a^{-2}$ all have the same length. We have shown that when $L = 6$ and $0 < \ell \leq 10$ it is not possible to find a shorter path from 1 to a^ℓ than the a-path, so $|a^\ell| = \ell$. We have also exhibited paths, either a^ℓ when $\ell < 6$ or $sas^{-1}tat^{-1}a^{\ell-6}$ when $6 \leq \ell \leq 10$, of the desired length, hence geodesic, and of the desired form: at most one x-escape, at most one y-escape, and a toral subpath of exponent strictly less than L.

Finally, consider $h = x^m y^n$ with $m \neq 0 \neq n$. Suppose a geodesic γ from 1 to h is not just an x-escape and a y-escape. By commutativity we may assume that the x and y-escapes, if any, come first, and end

at a point $g \in \langle x, y\rangle$. Then $1 \neq g^{-1}h \in \langle a\rangle \cap \langle x, y\rangle = \langle a^L\rangle$. Thus, the geodesic contains a toral subpath whose exponent is a nontrivial multiple of L. As we have seen above, this is impossible except when $L = 6$ and $h = a^{\pm 6}$. In this case there exists a geodesic that is a single x-escape and a single y-escape, but there is also one that is a toral path.

Now suppose the x-escape has exponent p and the y-escape has exponent q. Then $h = x^m y^n = x^p y^q$ implies $x^{m-p} = y^{q-n}$, but $\langle x\rangle \cap \langle y\rangle = 1$, so $p = m$ and $q = n$. □

3.2 Geodesics between elements of $\langle a\rangle$

Lemma 3.2.1 *If $0 \leq r < L$ and $q \in \mathbb{Z}$ then there exists a geodesic from 1 to a^{qL+r} that consists of a geodesic from 1 to either a^{qL} or $a^{(q+1)L}$, followed by a geodesic a-path.*

Proof By Proposition 3.1.1, there is a geodesic from 1 to a^{qL+r} of the form $\gamma_x + \gamma_y + a^p$ with $|p| < L$. Suppose the endpoints of γ_x differ by x^m and the endpoints of γ_y differ by y^n. Then $x^m y^n = a^{qL+r-p}$, but since $\langle x, y\rangle \cap \langle a\rangle = \langle a^L\rangle$, this implies $m = n$ and $qL + r - p = mL$. Since $r - p = (m-q)L$ and $-L < r - p < 2L$, we have $m = q$ or $m = q + 1$. Thus, $\gamma_x + \gamma_y$ ends at either a^{qL} or $a^{(q+1)L}$. □

Corollary 3.2.2 (1) $|a^\ell| = \ell$ *for* $0 < \ell \leq 3 + L/2$.

(2) $|a^\ell| = 6 + L - \ell$ *for* $3 + L/2 \leq \ell \leq L$.

(3) $|a^{qL}| = 4 + 2|a^q|$ *for all* $q \neq 0$.

(4) $||a^\ell| - |a^{\ell+1}|| = 1$ *for all* $\ell \in \mathbb{Z}$.

(5) $||a^{qL}| - |a^{(q+1)L}|| = 2$ *for all* $q > 0$.

Proof If $q \neq 0$ then $a^{qL} = x^q y^q$ and by Proposition 3.1.1 there is a geodesic from 1 to a^{qL} consisting of a single x-escape and a single y-escape, both of exponent q. A geodesic x-escape of exponent q is of the form $s + \underline{a}^q + s^{-1}$, for some geodesic $\underline{a}^q$ whose endpoints differ by a^q, and similarly for the y-escape, so item (3) follows.

By Lemma 3.2.1, if $0 < \ell < L$ then either there is a geodesic from 1 to a^ℓ that is an a-path of length ℓ, or there is a geodesic from 1 to a^ℓ that consists of a geodesic from 1 to a^L followed by an a-path of length $L - \ell$. The length of the latter is $6 + L - \ell$. When $\ell < 3 + L/2$ the former is shorter, when $\ell > 3 + L/2$ the latter is shorter, and when $\ell = 3 + L/2$ they have the same length. This proves items (1) and (2).

Clearly $||a^\ell| - |a^{\ell+1}|| \leq 1$, so to prove item (4) we only need to show it is not 0. It cannot be 0 because the relators all have even length, so two neighboring vertices cannot be equidistant to a third vertex.

Item (5) follows from items (3) and (4). □

We can use Lemma 3.2.1 and its corollary to construct geodesics from 1 to powers of a with nonzero exponent inductively: Suppose that we know a geodesic from 1 to $a^{\ell'}$ for all $0 < \ell' < \ell$. By Corollary 3.2.2, this is true for $\ell \leq L + 1$, so we may assume $\ell \geq L + 1$. Write $\ell = qL + r$ for $0 \leq r < L$, so $q > 0$.

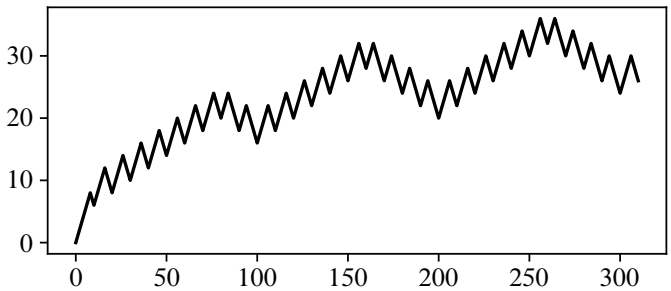


Figure 2: $|a^\ell|$ versus ℓ for $L = 10$. Local minima occur when $L|\ell$, second order minima occur when $L^2|\ell$, etc.

By the induction hypothesis, we know a geodesic γ_{qL} from 1 to a^{qL} and a geodesic γ_{q+1} from 1 to a^{q+1}. By Lemma 3.2.1, there is a geodesic from 1 to a^ℓ that consists of a geodesic through either a^{qL} or $a^{(q+1)L}$, followed by an a-path. By Proposition 3.1.1 there exists a geodesic from 1 to $a^{(q+1)L}$ consisting of an x-escape whose endpoints differ by x^{q+1} and a y-escape whose endpoints differ by y^{q+1}. The former has the form $s + \underline{a}^{q+1} + s^{-1}$ for some geodesic $\underline{a}^{q+1}$ whose endpoints differ by a^{q+1}, and similarly for the latter. But since we know one such geodesic, γ_{q+1}, we know a geodesic $\gamma_{(q+1)L} := s + \gamma_{q+1} + s^{-1} + t + \gamma_{q+1} + t^{-1}$ from 1 to $a^{(q+1)L}$. Therefore, the shorter of $\gamma_{qL} + a^r$ and $\gamma_{(q+1)L} + a^{L-r}$ is a geodesic from 1 to a^{qL+r}.

We wrote a computer program to compute $|a^\ell|$ inductively as described. The results for $L = 10$ are shown in Figure 2. The main takeaway from this experiment is that the function $\ell \mapsto |a^\ell|$ is not monotonic, and in fact monotonicity fails at arbitrarily large scales. This essentially comes down to the fact that $|a^\ell|$ is relatively shorter when ℓ is close to being a power of L, and is somewhat larger if ℓ can only be expressed as a combination of several different powers of L. In the rest of this subsection we make the relationship between $|\ell|$ and $|a^\ell|$ as precise as possible.

Consider an expression $0 < m = \sum_{i=0}^{j} m_i L^i$. Inductively associate to such an expression a path $\alpha + \beta + \gamma$ where γ is m_0-many a-edges, either forward or backward corresponding to the sign of m_0, and $\alpha = s + \delta + s^{-1}$ and $\beta = t + \delta + t^{-1}$ where δ is the path corresponding to $\sum_{i=0}^{j-1} m_{i+1} L^i$. This path has length

$$\sum_{i=0}^{j} |m_i| \cdot 2^i + 4(2^j - 1). \tag{3.2.3}$$

Call $m = \sum_{i=0}^{j} m_i L^i$ a *geodesic expression* of m if the corresponding path is a geodesic. By our inductive argument, every integer admits a geodesic expression. We derive constraints on the coefficients:

Lemma 3.2.4
- *For $0 \le \ell \le 2 + L/2$ there is a geodesic from 1 to a^ℓ consisting of an a-path of length at most $2 + L/2$.*
- *For $\ell > 2 + L/2$ there exists a geodesic from 1 to a^ℓ consisting of one x-escape, one y-escape, and an a-path of length at most $L/2$.*

Proof The first item and the second item for $L|\ell$ follow immediately from Corollary 3.2.2 items (1) and (3), respectively. The second item for $3 + L/2 \le \ell \le L$ follows from the proof of Corollary 3.2.2(2).

Now consider a^ℓ for $\ell \in (qL, (q+1)L)$ with $q \ge 1$. By Lemma 3.2.1, there is a geodesic from 1 to a^ℓ passing through either a^{qL} or $a^{(q+1)L}$. By Corollary 3.2.2, $||a^{qL}| - |a^{(q+1)L}|| = 2$. So if $|a^{qL}| < |a^{(q+1)L}|$ then for $\ell = qL + r$ with $r \le L/2$ there is a geodesic from 1 to a^ℓ passing through a^{qL} and then proceeding to a^ℓ via an a-path of length $r \le L/2$, while for $r \ge 1 + L/2$ there is a geodesic from 1 to a^ℓ passing through $a^{(q+1)L}$ and then proceeding to a^ℓ via an a-path of length $L - r \le L/2 - 1$. The argument for $|a^{qL}| > |a^{(q+1)L}|$ is similar. □

Corollary 3.2.5 *For all $m > 0$ there is a geodesic expression $m = \sum_{i=0}^{j} m_i L^i$ satisfying*

$$0 < m_j \le L/2 + 2 \quad \text{and} \quad |m_i| \le L/2 \quad \text{for } i < j. \tag{3.2.6}$$

Note the leading coefficient has to be positive when m is because $\sum_{i=0}^{j-1} \frac{1}{2} L \cdot L^i < L^j$. The conditions of (3.2.6) are not sufficient to conclude that the corresponding expression is geodesic, but they give useful general constraints.

Lemma 3.2.7 *If $0 < m = \sum_{i=0}^{j} m_i L^i$ is a geodesic expression satisfying (3.2.6) with $j \ge 1$ then $|j - \log_L m| < 1$.*

Proof Observe that

$$\left|\sum_{i=0}^{j-1} m_i L^i\right| \le \sum_{i=0}^{j-1} \frac{L}{2} L^i = \frac{1}{2}\frac{L}{L-1}(L^j - 1) < L^j \left(\frac{L}{2(L-1)}\right) \le \frac{3}{5} L^j.$$

So,

$$L^{j-1} < \tfrac{2}{5} L^j < m < L^j\left(\tfrac{1}{2}L + 2 + \tfrac{3}{5}\right) < L^{j+1}.$$ □

Lemma 3.2.8 *If $0 < m = \sum_{i=0}^{j} m_i L^i$ is a geodesic expression satisfying (3.2.6) with $j \ge 1$ then $5 \cdot 2^j - 4 \le |a^m| \le (L+6) \cdot 2^j - 4$.*

Proof By (3.2.3), $|a^m| = \sum_{i=0}^{j} |m_i| \cdot 2^i + 4(2^j - 1)$. The lower bound is achieved by taking $m_j = 1$ and $m_i = 0$ for $0 \le i < j$. The upper bound is achieved by taking $m_j = L/2 + 2$ and $m_i = L/2$ for all $0 \le i < j$. □

We can now show that the distortion of the subgroup $\langle a \rangle$ (and therefore also of the conjugate subgroups $\langle x \rangle$ and $\langle y \rangle$) is a power function:

Theorem 3.2.9 (power-law distortion) *There is a constant C depending on L such that for $g \in \{a, x, y\}$, $|m| > 0$, and $\alpha := \log_2 L$ we have*

$$1 \le \frac{|g^m|}{|m|^{1/\alpha}} < C.$$

Moreover, the first inequality is strict except in the case $g = a$ and $|m| = 1$.

Proof Suppose $g = a$. By Corollary 3.2.2(1), if $0 < m < L/2 + 3$ then $|a^m| = m \geq m^{1/\alpha} \geq 1$, with equality only if $m = 1$. If $L/2+3 \leq m < L^{3/2}$ then by Lemma 3.2.4, there is a geodesic from 1 to a^m that contains an x-escape and a y-escape, so has length at least 6. However, $m^{1/\alpha} < (L^{3/2})^{1/\alpha} = 2^{3/2} < 3$, so $|a^m|/m^{1/\alpha} > 2$. In the other direction, $|a^m|/m^{1/\alpha} \leq m^{1-1/\alpha}$, so the upper bound for $0 < m < L^{3/2}$ is satisfied if $C \geq (L^{3/2})^{1-1/\alpha} = (L/2)^{3/2}$.

Consider the case $m \geq L^{3/2}$. There is a geodesic expression $m = \sum_{i=0}^{j} m_i L^i$ for some $j \geq 1$ satisfying (3.2.6). By Lemma 3.2.7, $L^{j-1} < m < L^{j+1}$, so $2^{j-1} < m^{1/\alpha} < 2^{j+1}$.

Lemma 3.2.8 gives

$$m^{1/\alpha} + \tfrac{3}{2}m^{1/\alpha} - 4 = \tfrac{5}{2}m^{1/\alpha} - 4 < 5 \cdot 2^j - 4 \leq |a^m| \leq (L+6) \cdot 2^j - 4 < 2(L+6)m^{1/\alpha}.$$

But $m \geq L^{3/2}$ implies $\frac{3}{2}m^{1/\alpha} - 4 > 0$, so

$$m^{1/\alpha} < |a^m| < 2(L+6)m^{1/\alpha}.$$

If $g \in \{x, y\}$ and $m \neq 0$ then $|g^m| = 2 + |a^m|$ and the result follows from the result for a by taking $C := 2 + \max\{2(L+6), (L/2)^{3/2}\}$. □

In light of the claim in the introduction that the constant of Theorem 3.2.9 overwhelms the constant of the reverse Hölder inequality, and that this necessitates the labor of this paper, one might wonder whether this phenomenon is real or if it is an artifact of a suboptimal statement, or proof, of Theorem 3.2.9. Proposition 3.2.10 shows that the phenomenon is real; even if we pass to asymptotic bounds, the gap between C_{lo} and C_{hi} grows with L.

Proposition 3.2.10 *For even $L \geq 6$ consider $a \in G_L$; let $\alpha(L) := \log_2 L$,*

$$C_{\text{lo}}(L) := \liminf_{m\to\infty} \frac{|a^m|}{m^{1/\alpha(L)}} \quad \text{and} \quad C_{\text{hi}}(L) := \limsup_{m\to\infty} \frac{|a^m|}{m^{1/\alpha(L)}}.$$

Then $C_{\text{hi}}(L)/C_{\text{lo}}(L) > (L+6)/10$.

When $L \geq 10$, $(C_{\text{lo}}(L)/C_{\text{hi}}(L)) \cdot 2^{1-1/\alpha(L)} < 1$.

Proof For fixed $L \geq 6$ let $\alpha := \alpha(L)$ and $M := (L-2)/2$. For $n \geq 0$ let $m_n := ML^n - M\sum_{i=0}^{n-1} L^i$. By Corollary 3.2.2(1), $|a^{m_0}| = m_0 = M$. Notice that $m_{n+1} = Lm_n - M$. It follows from Lemma 3.2.1 and Corollary 3.2.2 parts (5) and (3) that $|a^{m_{n+1}}| = M + 4 + 2|a^{m_n}|$. Induction then shows

$$|a^{m_n}| = (2^{n+1} - 1)M + 2^{n+2} - 4.$$

Thus,

$$\begin{aligned}\frac{|a^{m_n}|}{m_n^{1/\alpha}} &= \frac{(2^{n+1}-1)M + 2^{n+2} - 4}{\left(M\left(\frac{L^{n+1}-2L^n+1}{L-1}\right)\right)^{1/\alpha}} \\ &= \frac{(2^{n+1}-1)\left(\frac{L-2}{2}\right) + 2^{n+2} - 4}{\left(\left(\frac{L-2}{2}\right)\left(\frac{L^{n+1}-2L^n+1}{L-1}\right)\right)^{1/\alpha}}\end{aligned}$$

$$\begin{aligned}
&= \frac{2^n\big((1-2^{-n-1})(L-2)+4-2^{2-n}\big)}{L^{n/\alpha}\big(\frac{1}{2}\cdot\frac{L-2}{L-1}\cdot(L-2+L^{-n})\big)^{1/\alpha}} \\
&= \frac{(1-2^{-n-1})(L-2)+4-2^{2-n}}{\big(\frac{1}{2}\cdot\frac{L-2}{L-1}\cdot(L-2+L^{-n})\big)^{1/\alpha}} \\
&\xrightarrow{n\to\infty} \frac{L+2}{\big(\frac{1}{2}\cdot\frac{(L-2)^2}{L-1}\big)^{1/\alpha}} \\
&= \frac{L+2}{\big(\frac{1}{2}\cdot\frac{(L-2)^2}{L-1}\big)^{\log_L 2}} \qquad \textbf{(3.2.11)}
\end{aligned}$$

As a function of L, this result is greater than $L/2+3$ on $[6,\infty)$, so

$$C_{\mathrm{hi}}(L) \geq \lim_{n\to\infty} \frac{|a^{m_n}|}{m_n^{1/\alpha}} > L/2+3.$$

In the other direction, induction with Corollary 3.2.2(3) gives

$$C_{\mathrm{lo}}(L) \leq \lim_{n\to\infty} \frac{|a^{L^n}|}{(L^n)^{1/\alpha}} = \lim_{n\to\infty} \frac{5\cdot 2^n-4}{2^n} = 5.$$

This proves the first claim. The second claim is easy to see for $L \geq 14$ using the bounds

$$\frac{C_{\mathrm{hi}}(L)}{C_{\mathrm{lo}}(L)} > \frac{(L+6)}{10}$$

and $2^{-1/\alpha} < 1$. It turns out also to be true for $L \in \{10, 12\}$ using (3.2.11) and explicit computation. □

Remark 3.2.12 For the rest of the paper L, and consequently, α and C are fixed, and these names are only used for the constants of Theorem 3.2.9.

3.3 Geodesics between elements of H

Recall that x and y edges in X have length L, so geodesics consist only of a, s, and t edges.

Lemma 3.3.1 *Let $g \in \{x, y\}$. Define $\mathrm{sign}(q)$ to be -1, 0, or $+1$ if q is negative, zero, or positive, respectively.*

(1) *For $m \in \mathbb{N}$, $|g^m| = 2 + |a^m|$.*

(2) *For $m \in \mathbb{N}$, $||g^m| - |g^{m+1}|| = 1$.*

(3) *If $\ell = qL + r$ with $0 \leq |r| < L$ and $m, n \in \mathbb{Z}$ then*

$$|a^\ell x^m y^n| = \min\{|r| + |x^{m+q}| + |y^{n+q}|,\ L - |r| + |x^{m+q+\mathrm{sign}(r)}| + |y^{n+q+\mathrm{sign}(r)}|\}.$$

(4) *There exists a $K > 1$ such that for all ℓ and m, $K|a^\ell g^m| \geq |a^\ell| + |g^m|$.*

Proof Item (1) follows immediately from Proposition 3.1.1, since a geodesic from 1 to x^m is a single x-escape. Item (2) follows from this and Corollary 3.2.2(4). For item (3) the $\ell = 0$ case follows immediately

from Proposition 3.1.1, while the general case is the analogue of Lemma 3.2.1: there exists a geodesic from 1 to $a^\ell x^m y^n$ passing though either $x^{m+q}y^{n+q}$ or $x^{m+q+\mathrm{sign}(r)}y^{n+q+\mathrm{sign}(r)}$, and then proceeding to $a^\ell x^m y^n$ via a geodesic a-path.

To prove item (4), it suffices to consider $\ell \geq 0$ and $g = x$.

If $\ell = 0$ or $m = 0$ then any $K > 1$ will suffice, so assume not.

Applying item (3),

$$|a^\ell x^m| = \min\{|r| + |x^{m+q}| + |y^q|,\ L - |r| + |x^{m+q+\mathrm{sign}(r)}| + |y^{q+\mathrm{sign}(r)}|\},$$
$$|a^\ell| + |x^m| = |x^m| + \min\{|r| + |x^q| + |y^q|,\ L - |r| + |x^{q+\mathrm{sign}(r)}| + |y^{q+\mathrm{sign}(r)}|\}.$$

Thus, there are four possibilities for $K|a^\ell x^m| - (|a^\ell| + |x^m|)$, depending on which terms realize the two minima. Rewrite $K|a^\ell x^m| - (|a^\ell| + |x^m|)$ as $T' + T$ where T' contains the $|r|$ and L terms, and T contains everything else. If $r = 0$ the minima are clear and $T' = (K-1)|r| = 0$. In general for $r \neq 0$,

$$T' \in \{(K-1)|r|,\ (K+1)|r| - L,\ KL - (K+1)|r|,\ (K-1)(L-|r|)\}.$$

All of these are nonnegative once $K \geq L - 1$, so we assume this lower bound on K and then it suffices to find K such that the remaining part T is nonnegative. The inequality $T \geq 0$ is equivalent to

$$K \geq \frac{2|x^{q''}| + |x^m|}{|x^{q'}| + |x^{m+q'}|}$$

for some $q', q'' \in \{q, q + \mathrm{sign}(r)\}$. Notice $|q' - q''| \leq 1$.

If $q' = 0$ then $|q''| \leq 1$, so $|x^{q''}| \leq 3$, and it suffices to take $K \geq 7$.

If $0 \neq q'$ then by item (2) we have $|x^{q''}|/|x^{q'}| \leq (|x^{q'}| + 1)/|x^{q'}| \leq 4/3$, so

$$K' := \frac{3|x^{q'}| + |x^m|}{|x^{q'}| + |x^{m+q'}|} > \frac{2|x^{q''}| + |x^m|}{|x^{q'}| + |x^{m+q'}|}.$$

Thus, it suffices to take $K \geq K'$, in the case that $\ell > 0$, $m \neq 0$, and $q' \neq 0$. We deduce upper bounds on K', hence sufficient lower bounds on K, according to several subcases.

When $-q' = m$ we have $|x^m| = |x^{q'}|$, so $K' = 4$.

If $q' = -1$ and $m \neq 0, 1$ then

$$K' = \frac{9 + |x^m|}{3 + |x^{m-1}|} \leq \frac{9 + |x^m|}{3 + |x^m| - 1} \leq \frac{12}{5}.$$

Since $\ell > 0$, we have $q' \geq -1$, so it remains to consider cases where $q' > 0$ and $m + q' \neq 0$. Suppose that $m < -q' < 0$. Then

$$\frac{|x^{q'}|}{|x^m|} < \frac{C(q')^{1/\alpha}}{|m|^{1/\alpha}} = C\left|\frac{q'}{m}\right|^{1/\alpha} < C,$$

and, using reverse Hölder and Theorem 3.2.9,

$$|x^{m+q'}| > |m+q'|^{1/\alpha} = (-m-q')^{1/\alpha} \geq (-m)^{1/\alpha} - (q')^{1/\alpha} > \frac{1}{C}|x^m| - |x^{q'}|.$$

Thus,

$$K' < \frac{3|x^{q'}| + |x^m|}{|x^{q'}| + \frac{1}{C}|x^m| - |x^{q'}|} = 3C\frac{|x^{q'}|}{|x^m|} + C < C(3C+1).$$

Now suppose $-q' < m < 0$, so $|x^m|/|x^{q'}| < C$ and

$$K' = \frac{3 + |x^m|/|x^{q'}|}{1 + |x^{m+q'}|/|x^{q'}|} < \frac{3+C}{1}.$$

Finally, we consider $m > 0$. In this case,

$$|x^{m+q'}| > (m+q')^{1/\alpha} > 2^{-1+1/\alpha}(m^{1/\alpha} + (q')^{1/\alpha}) > C^{-1}\cdot 2^{-1+1/\alpha}(|x^m| + |x^{q'}|).$$

The corresponding estimate for K' is then

$$\begin{aligned} K' &< \frac{3|x^{q'}| + |x^m|}{|x^{q'}| + C^{-1}\cdot 2^{-1+1/\alpha}(|x^m| + |x^{q'}|)} \\ &= \frac{3|x^{q'}| + |x^m|}{(1 + C^{-1}\cdot 2^{-1+1/\alpha})|x^{q'}| + C^{-1}\cdot 2^{-1+1/\alpha}|x^m|} \\ &< \frac{3}{(1 + C^{-1}\cdot 2^{-1+1/\alpha})} + \frac{1}{C^{-1}\cdot 2^{-1+1/\alpha}} \\ &= C\cdot 2^{1-1/\alpha}\left(\frac{3}{1 + C\cdot 2^{1-1/\alpha}} + 1\right). \end{aligned}$$

□

Corollary 3.3.2 *If $\ell = qL + r$ with $0 \leq |r| < L$ and $m, n \in \mathbb{Z}$ then*

$$|a^\ell x^m y^n| \geq |m+q|^{1/\alpha} + |n+q|^{1/\alpha} - 5.$$

Proof By Lemma 3.3.1(3),

$$|a^\ell x^m y^n| = \min\{|r| + |x^{m+q}| + |y^{n+q}|,\ L - |r| + |x^{m+q+\mathrm{sign}(r)}| + |y^{n+q+\mathrm{sign}(r)}|\}.$$

By Theorem 3.2.9, $|m+q|^{1/\alpha} + |n+q|^{1/\alpha} - 5 \leq |x^{m+q}| + |y^{n+q}| - 5$, which is clearly less than the first term of the minimum, so it suffices to prove it is also a lower bound for the second. But $L - |r| \geq 1$ and, by Lemma 3.3.1(2) and the fact that $|x| = |y| = 3$, we have $|g^{k\pm 1}| \geq |g^k| - 3$ for any $k \in \mathbb{Z}$ and $g \in \{x, y\}$. □

3.4 Lines, intersection, projections

The following lemmas are needed in Section 8.

Lemma 3.4.1 *For every $h \in H - \langle a\rangle$ there are exactly two closest points of $\langle a\rangle$. They are the unique element of $\langle a\rangle \cap h\langle x\rangle$ and the unique element of $\langle a\rangle \cap h\langle y\rangle$.*

Proof The isometric action by a stabilizes $\langle a\rangle$, and $H = \coprod_{i=0}^{L-1} a^i \langle x, y\rangle$, so it suffices to assume $h = x^m y^n \in \langle x, y\rangle - \langle a\rangle$. Consider a closest point a^p of $\langle a\rangle$ to h. By Proposition 3.1.1, there is a geodesic from h to a^p of the form $\gamma_x + \gamma_y + \gamma_a$, but γ_a is clearly trivial because $\gamma_x + \gamma_y$ ends at a point of $\langle a\rangle$. Suppose γ_x and γ_y are both nontrivial. By x/y-symmetry there is a geodesic x-escape γ_x' with the same exponent as γ_y such that $\gamma' = \gamma_x + \gamma_x'$ ends on $\langle a\rangle$. Then $|\gamma'| = |\gamma_x| + |\gamma_y|$, but γ' consists of consecutive x-escapes, so can be shortened, contradicting the fact that $\gamma_x + \gamma_y$ was a shortest path to $\langle a\rangle$. Thus, either $a^p = hx^{n-m} \in H\langle x\rangle$ or $a^p = hy^{m-n} \in h\langle y\rangle$, and by x/y-symmetry, both of these choices are the same distance from h. □

Lemma 3.4.2 *For all $h \in H$ there exists $\ell \in \mathbb{Z}$ with $|\ell| \le L/2$ such that $\langle x\rangle \cap ha^\ell\langle y\rangle \neq \varnothing$, in which case the intersection is a unique element.*

Proof It suffices to consider the case $h = a^p$. Since $\langle x, y\rangle \cap \langle a\rangle = \langle a^L\rangle$, we have $\langle x\rangle \cap a^{\ell+p}\langle y\rangle \neq \varnothing$ if and only if L divides $\ell + p$, in which case the intersection is the element $x^{(\ell+p)/L} = a^{\ell+p}y^{-(\ell+p)/L}$. Choose ℓ to be an integer of smallest absolute value for which $L|(\ell + p)$. □

Lemma 3.4.3 *For all $n \in \mathbb{Z}$, the unique closest point of $y^n\langle x\rangle$ to 1 is y^n.*

Proof The claim is trivial if $n = 0$, so suppose not. Suppose $y^n x^m$ is a closest point of $y^n\langle x\rangle$ to 1. If $m \neq 0$ then by Proposition 3.1.1 there is a geodesic from 1 to $x^m y^n$ consisting of a nontrivial y-escape of exponent n followed by a nontrivial x-escape. This is contradictory, since such a geodesic contains an interior point in $y^n\langle x\rangle$, so $m = 0$. □

Lemma 3.4.4 *If $L|p$ then $x^{p/L}$ is the unique closest point of $\langle x\rangle$ to a^p.*

Proof If $p = 0$ the claim is trivial, so suppose not. Let x^m be a closest point of $\langle x\rangle$ to a^p. By Proposition 3.1.1 there is a geodesic from a^p to x^m of the form $\gamma_a + \gamma_y + \gamma_x$.

First, notice γ_x is trivial, because otherwise $\gamma_a + \gamma_y$ is a shorter path from a^p to $\langle x\rangle$.

Next, observe that if γ_a has exponent r then γ_y is a y-escape with endpoints a^{p+r} and x^m, so $a^{p+r}y^n = x^m$ for some n. Since $\langle x, y\rangle \cap \langle a\rangle = \langle a^L\rangle$, this implies $L|(p+r)$. By hypothesis, $L|p$, so $L|r$, but $|r| < L$, so $r = 0$. This leaves us with $a^p = x^m y^{-n}$, which is true if and only if $m = p/L = -n$. □

4 The snowflake Cayley graph is not strongly shortcut

In this section we prove the following theorem.

Theorem 4.0.1 *The snowflake graph X_L contains arbitrarily large geodesic loops.*

Corollary 4.0.2 *The snowflake graph X_L is not shortcut or strongly shortcut.*

Corollary 4.0.3 *There is an asymptotic cone of X_L that contains an isometrically embedded unit length circle.*

Proof of Corollary 4.0.3 Choose the observation point to be fixed at 1 and choose the scaling sequence to match the lengths of the sequence of geodesic loops of Theorem 4.0.1. □

The existence of an asymptotic cone containing an isometrically embedded unit length circle is in fact equivalent to not being strongly shortcut [26, Theorem A].

Proof of Theorem 4.0.1 Recall for each $p \geq 1$ we inductively construct snowflake paths

$$\begin{aligned}
\sigma_{1,s} &:= sas^{-1}tat^{-1},\\
\sigma_{1,t} &:= tat^{-1}sas^{-1},\\
\sigma_{p+1,s} &:= s + \sigma_{p,s} + s^{-1} + t + \sigma_{p,s} + t^{-1},\\
\sigma_{p+1,t} &:= t + \sigma_{p,t} + t^{-1} + s + \sigma_{p,t} + s^{-1}.
\end{aligned}$$

We aim to show snowflake loops are geodesic; that is, for all $p \geq 1$, for every pair of antipodal points on the loop $\sigma_{p,s} + \bar{\sigma}_{p,t}$ the two complementary segments of the loop are geodesics. It suffices to show that a geodesic between two antipodal points has length at least $|\sigma_{p,s}| = |\sigma_{p,t}| = |\sigma_{p,s} + \bar{\sigma}_{p,t}|/2$. Furthermore, we show that every geodesic between a pair of antipodal points on the loop passes through either 1 or a^{L^p}.

By Corollary 3.2.2(3) and induction on p, we have $|a^{L^p}| = 5 \cdot 2^p - 4 = |\sigma_{p,s}| = |\sigma_{p,t}|$, so the snowflake paths themselves are geodesic.

Next, consider the points $u = x^{L^{p-1}}$ and $v = y^{L^{p-1}}$, which are the only pair of antipodal points other than $\{1, a^{L^p}\}$ that belong to H. By Proposition 3.1.1, every geodesic γ between u and v consists of a single geodesic x-escape of exponent L^{p-1} and a single geodesic y-escape of exponent L^{p-1}, so $|u^{-1}v| = |x^{L^{p-1}}| + |y^{L^{p-1}}| = |\sigma_{p,s}|$.

For the general case, by symmetry it suffices to consider antipodal points $u, v \notin H$ such that u lies on the first half of $\sigma_{p,s}$, which is an x-escape with trace in $\langle x \rangle$, and v lies on the second half of $\sigma_{p,t}$, which is an x-escape with trace in $a^{L^p}\langle x \rangle = y^{L^{p-1}}\langle x \rangle$. Let γ be a geodesic from u to v. Decompose γ as $\gamma = \gamma_1 + \gamma_2 + \gamma_3$, where γ_1 is the shortest subpath of γ connecting u to $\langle x \rangle$, and γ_3 is the shortest subpath of γ connecting v to $y^{L^{p-1}}\langle x \rangle$. Note since $u, v \notin H$, $\gamma_1 = \gamma_1' + s^{-1}$ and $\gamma_3 = s + \gamma_3'$.

As γ_2 is a geodesic between elements of H, by Proposition 3.1.1 we can rearrange escapes and toral subpaths of γ_2 to obtain a geodesic γ_2' with the same endpoints that consists of at most one x-escape, at most one y-escape, and at most one toral subpath of exponent less than L, in that order. First, we observe that γ_2' cannot contain an x-escape as it would start with an s-edge, which would given an s-cancellation with the final edge of γ_1, contradicting the fact that γ is geodesic.

Next, observe that there is no toral subpath in γ_2', because γ_2' is a path between $\langle x \rangle$ and $a^{L^p}\langle x \rangle$, so if the toral path has exponent ℓ and the y-escape has exponent n then we would have $a^\ell y^n = a^{L^p} x^m$ for

some m. This implies $a^{L^p-\ell} = x^{-m}y^n$, but $\langle a\rangle \cap \langle x, y\rangle = \langle a^L\rangle$, so ℓ is divisible by L. But $0 \le \ell < L$, so $\ell = 0$. Thus, γ_2' is a single geodesic y-escape of exponent L^{p-1}, which implies that $\gamma_2 = \gamma_2'$ and $|\gamma_2| = |y^{L^{p-1}}| = 2 + |a^{L^{p-1}}|$, by Lemma 3.3.1(1).

We proceed by induction on p.

In the case $p = 1$ the snowflake paths are $\sigma_{1,s} = sas^{-1}tat^{-1}$ and $\sigma_{1,t} = tat^{-1}sas^{-1}$, so the pairs of antipodal points we are considering are $\{u = s, v = tat^{-1}sa = a^L s\}$ and $\{u = sa = xs, v = tat^{-1}s = ys\}$. Observe that in both cases the points us^{-1} and vs^{-1} do not differ by a power of y, so it is not the case that both γ_1' and γ_3' are trivial. Thus $|\gamma_2| = |\gamma| - |\gamma_1'| - 1 - |\gamma_3'| - 1 \le 6 - 3 = 3$. The only length 3 y-escapes from $\langle x\rangle$ to $a^L\langle x\rangle$ are the ones labelled tat^{-1}, so we have $\gamma = \gamma_1' + s^{-1}tat^{-1}s + \gamma_3'$. Since $|\gamma| \le 6$, one of γ_1' or γ_3' is trivial and the other consists of a single edge. If $\{u, v\} = \{s, a^L s\}$ then we can choose either $\gamma_1' = 1$ and $\gamma_3' = a$ to get γ to be the path from u labelled $s^{-1}tat^{-1}sa$, which goes through 1, or $\gamma_1' = a$ and $\gamma_3' = 1$ to get γ to be the path from u labelled $as^{-1}tat^{-1}s$, which goes through a^L. These are precisely the two subsegments of the snowflake loop between u and v. The calculation is similar for $\{u, v\} = \{xs, ys\}$.

Now we suppose that the claims are true up to some power p, and extend to the case of $\sigma_{p+1,s} + \bar\sigma_{p+1,t}$. With the setup as above, γ_1' begins at u and ends at sa^m, for some m, so $s^{-1}.\gamma_1'$ is a path from $s^{-1}u$ to a^m. Since γ_2 is a single y-escape of exponent L^p, $\bar\gamma_3'$ begins at

$$sa^m s^{-1} y^{L^p} s = x^m y^{L^p} s = y^{L^p} x^m s = y^{L^p} sa^m$$

and ends at v; hence $s^{-1}y^{-L^p}.\bar\gamma_3'$ begins at a^m and ends at $s^{-1}y^{-L^p}v$. Thus $s^{-1}.\gamma_1' + s^{-1}y^{-L^p}.\bar\gamma_3'$ is a path from $s^{-1}u \in \sigma_{p,s}$ to $s^{-1}y^{-L^p}v \in \sigma_{p,t}$. Furthermore, since u and v were antipodal, we claim that $s^{-1}u$ and $s^{-1}y^{-L^p}v$ are antipodal in $\sigma_{p,s} + \bar\sigma_{p,t}$. To see this, let α be the segment of $\sigma_{p+1,s}$ from u to $a^{L^{p+1}}$, and let β be the segment of $\bar\sigma_{p+1,t}$ from $a^{L^{p+1}}$ to v, so that $|\alpha| + |\beta| = |\sigma_{p+1,s}|$. The path α decomposes as $\alpha' + s^{-1} + t + \sigma_{p,s} + t^{-1}$, where $\alpha' + s^{-1}$ is the subsegment of $\sigma_{p+1,s}$ from u to x^{L^p}. We have that $s^{-1}.\alpha'$ is the subsegment of $\sigma_{p,s}$ from $s^{-1}u$ to $s^{-1}x^{L^p}s = a^{L^p}$. On the other hand, $\beta = s + \beta'$, where β' is the subsegment of $\bar\sigma_{p+1,t}$ from $a^{L^{p+1}}s$ to v, and $s^{-1}y^{-L^p}.\beta'$ is the subsegment of $\bar\sigma_{p,t}$ from $s^{-1}y^{-L^p}a^{L^{p+1}}s = a^{L^p}$ to $s^{-1}y^{-L^p}v$. Thus, $s^{-1}.\alpha' + s^{-1}y^{-L^p}.\beta'$ is the subsegment of $\sigma_{p,s} + \bar\sigma_{p,t}$ from $s^{-1}u$ to $s^{-1}y^{-L^p}v$ going through a^{L^p}, and it has length

$$\begin{aligned}|\alpha'| + |\beta'| &= |\alpha| - 3 - |\sigma_{p,s}| + |\beta| - 1 = |\sigma_{p+1,s}| - |\sigma_{p,s}| - 4\\ &= 5\cdot 2^{p+1} - 4 - (5\cdot 2^p - 4) - 4 = 5\cdot 2^p - 4 = |\sigma_{p,s}|.\end{aligned}$$

Since $s^{-1}u$ and $s^{-1}y^{-L^p}v$ are antipodal in $\sigma_{p,s} + \bar\sigma_{p,t}$, the induction hypothesis says that the subsegment δ of $\sigma_{p,s} + \bar\sigma_{p,t}$ from $s^{-1}u$ to $s^{-1}y^{-L^p}v$ passing through a^{L^p} is geodesic. Split δ at a^{L^p} as a concatenation $\delta = \delta_1 + \delta_3$, and define $\delta' := s.\delta_1 + s^{-1} + t + \sigma_{p,s} + t^{-1} + s + \bar\delta_3$. Then δ' is the subsegment of $\sigma_{p+1,s} + \bar\sigma_{p+1,t}$ from u to v that goes through $a^{L^{p+1}}$. We must show δ' is geodesic. We have

$$\begin{aligned}|\delta'| - |\gamma| &= |\delta_1| + |\delta_3| + 4 + |\sigma_{p,s}| - \big(|\gamma_1'| + |\gamma_3'| + 2 + |\gamma_2|\big)\\ &= |\delta_1| + |\delta_3| + 4 + |a^{L^p}| - \big(|\gamma_1'| + |\gamma_3'| + 4 + |a^{L^p}|\big) = |\delta_1| + |\delta_3| - |\gamma_1'| - |\gamma_3'|.\end{aligned}$$

On the one hand this quantity is nonnegative, because δ' is a path from u to v and γ is a geodesic with the same endpoints. On the other hand it is nonpositive, since $s^{-1}.\gamma_1' + s^{-1}y^{-L^p}.\gamma_3'$ is a path from $s^{-1}u$ to $s^{-1}y^{-L^p}v$ and $\delta = \delta_1 + \delta_3$ is a geodesic with the same endpoints. Thus, $|\delta'| = |\gamma|$, so δ' is geodesic. Since the complement to δ in $\sigma_{p+1,s} + \bar{\sigma}_{p+1,t}$ has the same length and the same endpoints, it is geodesic too. We also get that $s^{-1}.\gamma_1' + s^{-1}y^{-L^p}.\gamma_3'$ is a geodesic from $s^{-1}u$ to $s^{-1}y^{-L^p}v$, which are antipodal points on $\sigma_{p,s} + \bar{\sigma}_{p,t}$, so by the induction hypothesis $s^{-1}.\gamma_1' + s^{-1}y^{-L^p}.\gamma_3'$ passes through either 1 or a^{L^p}, necessarily at the concatenation point. This means that either γ_1 ends at 1 or γ_3 begins at $a^{L^{p+1}}$, so γ passes through either 1 or $a^{L^{p+1}}$. That completes the induction. □

5 The idea of the proof that asymptotic cones are simply connected

5.1 The snowflake case

Let us sketch the argument that snowflake loops can be subdivided in a way that satisfies Theorem 2.3.2. The idea in the general case will be that, for K sufficiently close to 1, K-bi-Lipschitz loops look sufficiently similar to snowflake loops that the same strategy works.

Consider a snowflake loop γ for a^{L^p} with HNN diagram as in Figure 1. We will subdivide each region and replace it by regions with relatively short boundaries. The diagram has a central diamond, the sides of which are labelled $x^{L^{p-1}}$ and $y^{L^{p-1}}$. From (3.0.3), the length of γ is $2(5\cdot 2^p - 4)$. We choose a subdivision constant Λ, which in this case we can take to be $\Lambda = L$. Subdivide the central diamond into Λ^2 small diamonds whose sides are labelled $x^{L^{p-2}}$ and $y^{L^{p-2}}$. If we take one of these small diamonds and replace its sides by geodesics we get a small snowflake loop whose length is $2(5\cdot 2^{p-1} - 4) < \frac{1}{2}|\gamma|$. Propagate the subdivision into the branches of the snowflakes to get subdivisions of the neighboring triangular regions, as in Figure 3. This gives us $\sim \Lambda^2$ more short loops.

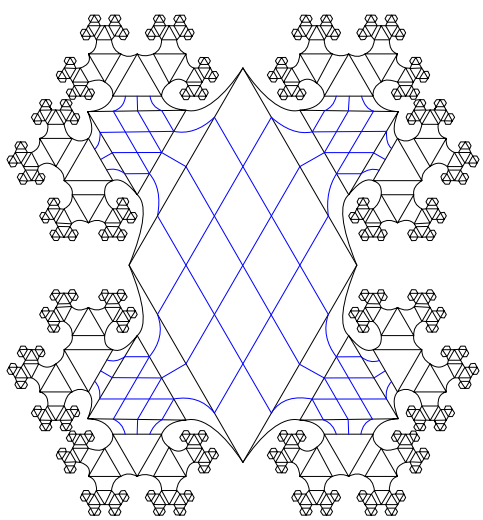


Figure 3: Subdivided snowflake to depth 1.

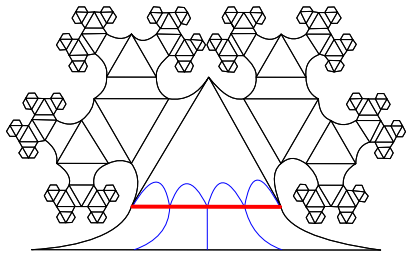


Figure 4: Capping off a branch.

Obviously, we cannot continue this propagation indefinitely, since the depth of the snowflake depends on the boundary word, but to apply Theorem 2.3.2 we need the number of short loops to be uniform. Consider the situation depicted in Figure 4 — we have subdivided some number of levels deeply into the snowflake, and are faced with the remaining branch. The incoming corridor has been subdivided into Λ-many pieces. These pieces have been replaced by short geodesic segments. The idea is to make one short loop consisting of these Λ-many geodesics, plus the outer boundary of the branch. The incoming edge of the triangle, the red edge in Figure 4, is labelled $a^{L^{p-m}}$ if we are at depth m. Thus, the length of the outer boundary of the branch is $|a^{L^{p-m}}|$, while the length of each of the Λ-many short geodesics is $|a^{L^{p-m-1}}|$, since $\Lambda = L$. Given our formula (3.0.3), we can solve to see that $|a^{L^{p-m}}| + \Lambda|a^{L^{p-m-1}}| \leq \frac{1}{2}|\gamma|$ once $m \sim \log_2 L$. So at some depth m, *independent of* p, the paths have become short enough that we can cap off a branch with a single short loop.

This argument decomposes the snowflake loop γ into $\sim L^3$ loops of length at most $\frac{1}{2}|\gamma|$. If a similar bound were to hold for bi-Lipschitz loops, then, by [10, Theorem 4.3.3], this would imply that the Dehn function of G_L is bounded above by $n^{3\log_2 L}$. This is consistent with the true Dehn function, which is known to be $n^{2\log_2 L}$. We have made no effort to optimize the decomposition, and do not know if it is possible to subdivide into only $\sim L^2$-many short loops.

5.2 The general case

We must show that for all $\lambda \in (0, 1)$ there exist M and A such that every loop γ in X can be filled by a diagram with area at most A and mesh at most $\lambda|\gamma| + M$.

Step 1: reduce to bi-Lipschitz loops In Section 6 we show that if γ fails to be bi-Lipschitz at a large scale, in a sense to be made precise, then it is easy to fill: introduce a geodesic segment between two far apart points where the bi-Lipschitz condition fails and take an area 2 filling of the Θ graph with the 1-skeleton mapping to the union of γ and this geodesic segment.

Otherwise, if γ fails to be bi-Lipschitz only at smaller scales then Theorem 6.0.1 says it can be tightened to a K-bi-Lipschitz loop without changing it very much, where we are free to choose $K > 1$ as small as we like. Corollary 6.0.3 says if we can find the desired filling of the bi-Lipschitz loop then we can extend it to the original loop without changing the area or the mesh too much. Thus, we have reduced the problem to finding some $K > 1$ for which we can fill all K-bi-Lipschitz loops.

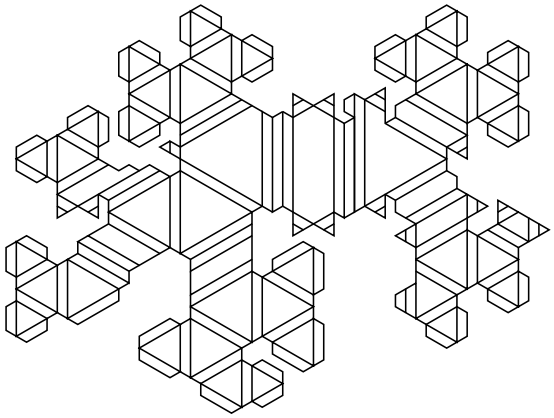


Figure 5: A bi-Lipschitz snowflake.

Interlude: how different are bi-Lipschitz loops from snowflakes? Figure 5 shows a schematic of an HNN diagram for a bi-Lipschitz loop. We point out some differences between Figure 5 and Figure 1:

(1) Figure 5 has no central diamond.

(2) There can be adjacent corridors with no vertex region in between.

(3) Not all of the noncentral vertex regions are triangles.

We claim these differences are not severe. It turns out there is a way to choose a central region. Moreover, although regions may have any number of sides, we will show that most of them are negligibly small and are easy to account for in our filling diagram without affecting the mesh very much. More specifically, the central region has at most four long sides, and the other regions have either two or three long sides. It will take some work to make these statements precise.

Branching of the diagram played an important role in the snowflake case, so the presence of bigon regions may be concerning. We assemble maximal chains of corridor and bigon regions into subdiagrams that we call *enfilades* in Section 7.3.

Step 2: all diagrams have a central "region" Consider an HNN diagram $D \to X$ for an embedded loop γ in X. Since γ is embedded, D is a disk. If D is a single region then call that one the central region. If not, D contains corridors. Recall that corridors have disjoint interiors and bisect D, and contain two disjoint edges in ∂D. Furthermore, corridors are 2-cells with boundary label of the form $sa^n s^{-1} x^{-n}$ or $ta^n t^{-1} y^{-n}$. We parameterize such a 2-cell as a rectangle $[0,n]\times[0,1]$ and define a map $\phi\colon D \to |T|$ from D to the geometric realization $|T|$ of the dual tree T to the corridor decomposition of D, such that ϕ restricted to a corridor parameterized $[0,n]\times[0,1]$ is projection to the second coordinate followed by a linear isometry to the geometric realization of an edge.

Since corridors contain disjoint edges on ∂D, for every $x \in |T|$, $\phi^{-1}(x)$ contains at least two points of ∂D, so $\gamma - \phi^{-1}(x)$ consists of at least two components. Define a function f on $|T|$ by

$$x \mapsto (\text{maximum length of a component of } \gamma - \phi^{-1}(x)) - \tfrac{1}{2}|\gamma|.$$

The function f is typically not continuous. We claim that there is a unique point at which f achieves a nonpositive value.

First, suppose that f achieves a nonpositive value at x. Then $\gamma - \phi^{-1}(x)$ consists of finitely many segments γ_i, each of length at most $|\gamma|/2$. For any other $y \in |T|$, there is a component γ' of $\gamma - \phi^{-1}(y)$ and an i_0 such that $\gamma - \gamma'$ is contained in γ_{i_0}. Furthermore, γ' crosses corridors corresponding to the edges of T on the geodesic between x and y. Thus, $|\gamma'| \geq |\gamma| - |\gamma_{i_0}| + 2d_T(x, y) > |\gamma| - |\gamma_{i_0}| \geq |\gamma|/2$, so $f(y) \geq |\gamma'| - |\gamma|/2 > 0$. So if f achieves a nonpositive value, it does so at a unique point of $|T|$.

Next, suppose that v is a vertex of T with $f(v) > 0$. Then there is a component γ' of $\gamma - \phi^{-1}(v)$ of length strictly greater than $|\gamma|/2$. The closure $\mathrm{cl}(\gamma')$ of γ' maps via ϕ to a loop in T at v that begins and ends by crossing some edge e incident to v. Suppose this edge is parameterized $[0, 1]$ with 0 at v. Then for $x \in (0, 1)$ there is a component of $\gamma - \phi^{-1}(x)$ whose closure is a subsegment of γ' of length $|\gamma'| - 2x$. So either $f(v) > 1$ and f is linearly decreasing and positive on $[0, 1)$, or $f(v) = 1$ and f achieves value 0 at the midpoint of e. In both cases, $f(v) > 0$ implies that $f(v)$ is not a local minimum.

Finally, suppose that f is positive and decreasing on some edge parameterized $[0, 1]$. Then for each $x \in (0, 1)$ there is a component γ_x of $\gamma - \phi^{-1}(x)$ such that $|\gamma_x| > |\gamma|/2$ and as $x \to 1$ the closed segments $\mathrm{cl}(\gamma_x)$ form a decreasing nested family all containing $\phi^{-1}(1) \cap \partial D$. The limit γ' of this family as $x \to 1$ is a closed subsegment of γ of length at least $|\gamma|/2$ containing $\phi^{-1}(1) \cap \partial D$. There are two possibilities: either $\phi^{-1}(1) \cap \partial D$ is exactly two points and $f(1) = |\gamma'| - |\gamma|/2$, so $f|_{(0,1]}$ is continuous at 1, or $\phi^{-1}(1) \cap \partial D$ has more than two points and the components of $\gamma - \phi^{-1}(1)$ are either $\gamma - \gamma'$ or proper subsegments of γ'. In this case $f(1) \leq |\gamma'| - |\gamma|/2$, so $f(1)$ is a strict lower bound for f on $(0, 1)$.

Since there are finitely many vertices, we conclude that f achieves a minimum value. If $f(x) > 0$ then x is not a local minimum, so the minimum value achieved by f is nonpositive.

If f achieves value 0 on the midpoint of an edge then we call the corresponding corridor the central region R of D. If f achieves its minimum value at a vertex v then we define $R = \phi^{-1}(v)$ to be the central region of D. This includes a degenerate possibility that R is not really a region; it can be that $\phi^{-1}(v)$ is a tree in the 1-skeleton of D with leaves on ∂D. The important point is that the map $D \to X$ sends $\phi^{-1}(v)$ into a single coset of H.

Remark 5.2.1 There is a natural map from the dual tree in the above argument into the Bass–Serre tree of G, so to every vertex $v \in T$ there is associated a coset $g_v H$. However, we cannot assume this map is injective. In particular, if v is a vertex of T, we cannot conclude that the closures of components of $\gamma - \phi^{-1}(v)$ are *escapes* from $g_v H$. One of these components could return to $g_v H$ in its interior. We need a more general definition to describe these paths.

Definition 5.2.2 Let γ be a path in X that intersects a coset gH. Define a *compound escape* of γ from gH to be a subpath γ' of γ with both endpoints in gH such that one of the following holds:

- γ' begins with s, ends with s^{-1}, and has endpoints that differ by a power of x.

- γ' begins with t, ends with t^{-1}, and has endpoints that differ by a power of y.
- γ' begins with s^{-1}, ends with s, and has endpoints that differ by a power of a.
- γ' begins with t^{-1}, ends with t, and has endpoints that differ by a power of a.

The first two cases we call compound x- and y-escapes, respectively, and the last two compound a-escapes.

So our usual, simple, escapes are also compound escapes, but a compound escape may return to gH in its interior.

In the case where γ is a K-bi-Lipschitz (and hence embedded) loop, we gain two things by identifying the central region:

- If R is the central region, each component of $\gamma - R$ is a K-bi-Lipschitz path with endpoints in a coset of H.
- If R is a noncentral region then at most one component γ' of $\gamma - R$ can fail to be K-bi-Lipschitz, and $\gamma - \gamma'$ is a single K-bi-Lipschitz path with endpoints in a coset of H.

We analyzed *geodesics* between elements of H in Section 3.3, and now we must:

Step 3: analyze bi-Lipschitz paths between elements of H This is Section 7.1. The underlying idea is that for such a path to be efficient it should make as much use of the distortion of H as possible, and the way that happens is when the path is almost entirely either one long escape or two long escapes that are an x-escape and a y-escape.

Step 4: apply the previous step to analyze central and noncentral regions of the diagram This is done in Sections 7.2 and 7.3. The conclusions, as alluded to above, are that the central region has at most four long sides, and they are configured in one of 6 possible ways, and every other region has at most 3 long sides, and in the case of 3 they approximately form a triangle of a-, x-, and y-escapes just like in the snowflake case. Furthermore, we show that when we lump together adjacent bigon regions and corridors into enfilades, the enfilades are metrically thin.

Step 5: construct a new filling diagram We have established that the structure of the diagram is a central region with at most 4 adjacent enfilades, each of which leads to a triangular region. Each of these has two outgoing sides with adjacent enfilades leading to more triangular regions, etc. The strategy now will be the same as in the snowflake case: choose a way to subdivide the central region and propagate this subdivision out across the original diagram in a way that the mesh is controlled and such that after a uniform number of steps we can cap off the remaining branches with single 2-cells. This is done in Section 8.

6 Reduction to bi-Lipschitz cycles

The following theorem is a specialized form of a result of Hoda [26].

Theorem 6.0.1 (circle tightening lemma [26, Theorem G]) *Let $N > 1$ and $K > 1$. Let $\theta > 1$ be small enough (depending on N and K) and let $M > 0$ be large enough (depending on N, K, and θ). Let $\alpha\colon S \to X$ be a 1-Lipschitz map from a Riemannian circle S to a geodesic metric space X. If*

$$d_X(\alpha(p), \alpha(\bar{p})) \geq \frac{1}{\theta} d_S(p, \bar{p})$$

for every antipodal pair $p, \bar{p} \in S$ and $|S| > M$ then there exists a countable collection $\{Q_i\}_i$ of pairwise disjoint closed segments in S such that the following conditions hold:

(1) $\sum_i |Q_i| < |S|/N$.

(2) *For each i, we have $d_X(\alpha(p_i), \alpha(q_i)) < |Q_i|$ where p_i and q_i denote the endpoints of Q_i.*

(3) *There is a collection of geodesics $\{\gamma_i : \overline{Q}_i \to X\}_i$ such that γ_i has endpoints $\alpha(p_i)$ and $\alpha(q_i)$ and replacing each $Q_i \to X$ in α with $\overline{Q}_i \to X$ results in a K-bi-Lipschitz embedding α' of a circle that is also 1-Lipschitz.*

Remark 6.0.2 If X is a graph and α is a combinatorial map then no Q_i can be contained in a single edge, as this would contradict $d_X(\alpha(p_i), \alpha(q_i)) < |Q_i|$. Then each Q_i contains a vertex in its interior and it follows that the family Q_i is necessarily finite. Moreover, since α' is bilipschitz, it is also a combinatorial map.

Corollary 6.0.3 *Let X be a Cayley graph. Suppose there exist $K > 1$, $0 < \lambda < 1$, M, and A such that every K-bi-Lipschitz loop α admits a filling of area at most A and mesh at most $\lambda|\alpha| + M$. Then there exists $0 < \lambda' < 1$ and M' such that every 1-Lipschitz loop α admits a filling of area at most $A + 3$ and mesh at most $\lambda'|\alpha| + M'$.*

Proof Take $N > 12$. Let $\theta > 1$ be as in Theorem 6.0.1 with respect to N and K. Let M' be the larger of M and the M of Theorem 6.0.1. Define

$$\lambda' := \max\Big\{\lambda,\ \frac{1+\theta}{2\theta},\ 2\Big(\frac{1}{3} + \frac{2}{N}\Big)\Big\} < 1.$$

Consider a 1-Lipschitz loop $\alpha\colon S \to X$. If there exist a pair of antipodal points p and $\bar{p} \in S$ such that

$$d_X(\alpha(p), \alpha(\bar{p})) < \frac{1}{\theta}\frac{|S|}{2}$$

then there is a diagram of area 2 filling α obtained by taking one 2-cell to have boundary the segment of α restricted to the interval of S from p to $\bar{p}$ followed by a geodesic δ from $\alpha(\bar{p})$ to $\alpha(p)$, and taking the other to have boundary $\bar{\delta}$ followed by the remaining subsegment of α. These loops have length

$$|S|/2 + |\delta| < \frac{1+\theta}{2\theta}|S| \leq \lambda'|S|.$$

If there is no such pair of antipodal points then Theorem 6.0.1 applies. We use notation as in the theorem. By hypothesis, there is a filling of α' of area at most A and mesh at most $\lambda|S'| + M \leq \lambda'|S| + M'$. We need to extend this to a filling of α. Pick points s_0, s_1, and s_2 in $S - \bigcup_i \mathring{Q}_i$ such that for all i, subscripts mod 3, we have $d_S(s_i, s_{i+1}) \leq \big(\frac{1}{3} + \frac{2}{N}\big)|S|$. Then we get 3 loops of the form: segment of α

restricted to the interval of S from s_i to s_{i+1} followed by reverse of the segment of α' from s_i to s_{i+1}. The length of such a loop is at most twice the α part, since α' is a tightening of α, so it is bounded above by $2\left(\frac{1}{3}+\frac{2}{N}\right)|S| \le \lambda'|S|$. Adding three 2-cells with these boundary loops to the diagram for α gives us the desired diagram for α. □

7 Bi-Lipschitz cycles in snowflake groups

Recall, Remark 3.2.12, that the group G_L is deemed to be fixed, so L, C, and α are constant. The other "constants" in this section implicitly depend on the choice of L.

In Section 7.1 we analyze bi-Lipschitz paths with endpoints in H. Sections 7.2 and 7.3 are preparatory for Section 8. In those sections we establish properties for central and noncentral regions of a putative filling of a bi-Lipschitz loop γ. Since "region" is a term applied to an already existing filling, we will phrase these properties in a way that is intrinsic to γ.

7.1 Bi-Lipschitz paths with endpoints in H

The main result of this subsection is:

Theorem 7.1.24 *For all $R > 1$ there exist $J > 1$ and $1 < K < 2$ such that if γ is a K-bi-Lipschitz path with endpoints in $H = \langle a, x, y\rangle$ then either $|\gamma| \le J$ or γ has either one or two long escapes that collectively account for at least $(R-1)/R$ fraction of its length. Furthermore, in the case of two long escapes they are one x-escape and one y-escape.*

If the endpoints of γ differ by a power of x or y, then there is only one long escape, and it is of the corresponding flavor, and if the endpoints of γ differ by a power of a then either there is one long a-escape or there is one long x-escape and one long y-escape.

We build up to this theorem through a string of auxiliary results. Most follow the same template: given a target R that quantifies the desired ratio of domination, it is possible to choose $K > 1$ sufficiently close to 1 so that any K-optimal path γ with endpoints in H must have one or two escapes that dominate the others. The complexity of the assumed structure of γ increases with each lemma.

There is also an issue that in the theorem the domination is phrased in terms of length, but for most of the auxiliary results it is more convenient to work with exponents, so we will have to make a conversion. Because the relationship between exponent and length is not monotonic, this conversion is nontrivial.

For the first step, we suppose that γ consists of two escapes of the same flavor:

Lemma 7.1.1 *There exists $J > 1$ such that for all $R > 1$ there exists $1 < K < 2$ such that for all $|m|, |n| \ge J$, if $g \in \{a, x, y\}$ and*

$$|g^m| + |g^n| \le K|g^{m+n}|$$

then $\max\{|n|/|m|, |m|/|n|\} > R$.

Proof First assume $g = a$. Let $J := 1 + (L+3)/2$. Let $m = \sum_{i=0}^{j} m_i L^i$ and $n = \sum_{i=0}^{k} n_i L^i$ be geodesic expressions for m and n satisfying condition (3.2.6). Since $|m|, |n| \geq J$, we have $j, k > 0$. Without loss of generality, assume $k \geq j$ and $n > 0$ so that $n_k \geq 1$.

We approximate as follows:

$$\begin{aligned}
\frac{|n|}{|m|} = \frac{|\sum_{i=0}^{k} n_i L^i|}{|\sum_{i=0}^{j} m_i L^i|} &= \frac{|n_k L^k + \sum_{i=0}^{k-1} n_i L^i|}{|m_j L^j + \sum_{i=0}^{j-1} m_i L^i|} \\
&\geq \frac{L^k + \sum_{i=0}^{k-1} -(L/2)L^i}{(L/2+2)L^j + \sum_{i=0}^{j-1}(L/2)L^i} \quad \text{(by (3.2.6))} \\
&= \frac{L^k - \frac{1}{2}\frac{L}{L-1}(L^k-1)}{\frac{1}{2}(L+4)L^j + \frac{1}{2}\frac{L}{L-1}(L^j-1)} \\
&= \frac{2(L-1)L^k - L(L^k-1)}{(L-1)(L+4)L^j + L(L^j-1)} \\
&= \frac{L^{k+1} - 2L^k + L}{L^{j+2} + 4L^{j+1} - 4L^j - L} = \frac{(L-2)L^{k-1} + 1}{(L+4)L^j - 4L^{j-1} - 1}.
\end{aligned}$$

Thus, $|n|/|m| > R$ if

$$(L-2)L^{k-1} + 1 - R\big((L+4)L^j - 4L^{j-1} - 1\big) > 0.$$

The left hand side is bigger than $(L-2)L^{k-1} - R(L+4)L^j$, so the claim holds if

$$k - j > 1 + \log_L\left(\frac{R(L+4)}{L-2}\right). \tag{7.1.2}$$

Otherwise, if $0 \leq k - j \leq 1 + \log_L(R(L+4)/(L-2))$, we choose K sufficiently close to 1 to make $|a^m| + |a^n| \leq K|a^{m+n}|$ impossible. To show that this can be done we use that $m+n = \sum_{i=0}^{k}(m_i + n_i)L^i$, where $m_i = 0$ for $i > j$, to provide an upper bound for $|a^{m+n}|$:

$$\begin{aligned}
|a^m| + |a^n| - K|a^{m+n}| &\geq |a^m| + |a^n| - K\left(\sum_{i=0}^{k} |m_i + n_i| 2^i + 4(2^k - 1)\right) \\
&= 4(2^j - 1) + 4(2^k - 1) - 4K(2^k - 1) + \sum_{i=0}^{j}(|m_i| + |n_i| - K|m_i + n_i|)2^i \qquad \text{(by (3.2.3))} \\
&\geq 4(2^j - 1) - (K-1)4(2^k - 1) + \sum_{i=0}^{k} -(K-1)2^i(|m_i| + |n_i|) \qquad \text{(triangle inequality)} \\
&\geq 4(2^j - 1) - (K-1)4(2^k - 1) - (K-1)\sum_{i=0}^{k} 2^i L - (K-1)(2^{k+1} + 2^{j+1}) \qquad \text{(by (3.2.6))} \\
&= 4(2^j - 1) - (K-1)\big(4(2^k - 1) + L(2^{k+1} - 1) + 2^{k+1} + 2^{j+1}\big) \\
&= 2^{j+2} - (K-1)\big(2^{k+2} + L(2^{k+1} - 1) + 2^{k+1} + 2^{j+1}\big) - 4(2 - K).
\end{aligned}$$

To make this expression positive, it suffices to have

$$1-\frac{2-K}{2^j} > 2^{k-j}(K-1)\left(1+\frac{L}{2}+2^{-1}+2^{j-1-k}\right). \tag{7.1.3}$$

The left hand side is at least $1/2$, and the right hand side is bounded above by

$$\left(\tfrac{1}{2}\right)\cdot 2^{k-j}(K-1)(L+4).$$

Thus, applying the bound on $k-j$ from the negation of (7.1.2), we see that (7.1.3) is satisfied for

$$K < 1+\frac{1}{2^{k-j}(L+4)} \le 1+\frac{1}{2(L+4)}\cdot\left(\frac{R(L+4)}{L-2}\right)^{-1/\alpha}.$$

This completes the proof when $g=a$.

If $g\in\{x,y\}$ then $|g^p| = |a^p|+2$ (by Lemma 3.3.1), so $|g^m|+|g^n| \le K|g^{m+n}|$ implies

$$|a^m|+|a^n| = |g^m|+|g^n|-4 \le K|g^{m+n}|-4 = K|a^{m+n}|+2K-4 \le K|a^{m+n}|.$$

Thus, the result follows from the result for a. □

Remark 7.1.4 In Lemma 7.1.1 we provide a condition to show that two quantities, let us call them A and B, are related by $A/B > R$ for some $R>0$. We will frequently pass without comment between equivalent formulations of this relation that we note here to avoid any confusion:

$$\frac{A}{B} > R \iff A > RB \iff A > \frac{R}{R+1}(A+B) \iff B < \frac{1}{R+1}(A+B).$$

Next, we consider multiple escapes, all of the same flavor and with all exponents positive:

Corollary 7.1.5 *For all $R>1$ there exist $J>1$ and $1<K<2$ such that if $g\in\{a,x,y\}$ and $m_1\ge m_2\ge\cdots\ge m_k>0$ satisfy $\sum_{i=1}^k m_i \ge J$ and*

$$\sum_{i=1}^k |g^{m_i}| \le K|g^{\sum_{i=1}^k m_i}|$$

then $m_1 > \frac{R}{R+1}\sum_{i=1}^k m_i$.

Proof Replace R by $\max\{R,5\}$. Observe that the result for the new R implies the result for the original R. Let $J' := 1+(L+3)/2$ as in Lemma 7.1.1, let $J := J'(R+1)$, and let K be a constant satisfying Lemma 7.1.1 with respect to R. By Lemma 7.1.1, we have one of the following possibilities:

- $m_1/(m_2+m_3+\cdots+m_k) > R$;
- $m_2+m_3+\cdots+m_k < J'$;
- $(m_2+m_3+\cdots+m_k)/m_1 > R$;
- $m_1 < J'$.

Since $\sum_{i=1}^k m_i \geq J$, either of the first two conditions imply $m_1 > \frac{R}{R+1}\sum_{i=1}^k m_i$. The disjunction of the last two conditions implies $m_2 + m_3 + \cdots + m_k > \frac{R}{R+1}\sum_{i=1}^k m_i$. Thus it suffices to rule out this latter possibility.

Suppose $m_2 + m_3 + \cdots + m_k > \frac{R}{R+1}\sum_{i=1}^k m_i$. Let ℓ be least such that $\sum_{i=1}^\ell m_i \geq \frac{1}{2}\sum_{i=1}^k m_i$. Then, for all i, $m_i \leq m_1 < \frac{1}{R}\sum_{i=1}^k m_i$, and so

$$\frac{1}{2}\sum_{i=1}^k m_i \leq \sum_{i=1}^\ell m_i < \frac{1}{2}\sum_{i=1}^k m_i + m_\ell < \frac{R+2}{2R}\sum_{i=1}^k m_i$$

holds. Then $\sum_{i=1}^\ell m_i \geq \frac{1}{2}\sum_{i=1}^k m_i \geq J/2 > J'$ and $\sum_{i=\ell+1}^k m_i > \frac{R-2}{2R}\sum_{i=1}^k m_i \geq \frac{R-2}{2R}J > J'$, since $R \geq 5$, so by Lemma 7.1.1, either

(I) $\sum_{i=1}^\ell m_i > R\sum_{i=\ell+1}^k m_i$, or

(II) $\sum_{i=\ell+1}^k m_i > R\sum_{i=1}^\ell m_i$.

We will show that either case results in a contradiction. In case (I) we have

$$\frac{R+2}{2R}\sum_{i=1}^k m_i > \sum_{i=1}^\ell m_i > R\sum_{\ell+1}^k m_i > \frac{R(R-2)}{2R}\sum_{i=1}^k m_i$$

which contradicts $R \geq 5$. In case (II) we have

$$\frac{1}{2}\sum_{i=1}^k m_i \geq \sum_{i=\ell+1}^k m_i > R\sum_1^\ell m_i \geq \frac{R}{2}\sum_{i=1}^k m_i$$

which again contradicts $R \geq 5$. □

Next we will generalize to the case that γ consists of escapes of the same flavor, this time allowing both positive and negative exponents. First, there is an auxiliary result, Corollary 7.1.6, that says either the positive exponents or the negative exponents dominate. Then, in Corollary 7.1.7, we show that a single escape dominates.

Corollary 7.1.6 *For all $R > 1$ there exists $J > 1$ such that for any sufficiently small $1 < K < 2$, independent of J, if $g \in \{a, x, y\}$, $\sum_{i=1}^k |m_i| \geq J$, and*

$$\sum_{i=1}^k |g^{m_i}| \leq K|g^{\sum_{i=1}^k m_i}|,$$

then for $I = \{i \mid m_i > 0\}$ and for one of $I^* \in \{I, I^c\}$ we have

$$\left|\sum_{i\in I^*} m_i\right| > \frac{R}{R+1}\sum_{i=1}^k |m_i|$$

and

$$\sum_{i\in I^*} |g^{m_i}| < \left(K + (K-1)CR^{(-1/\alpha)}\right)|g^{\sum_{i\in I^*} m_i}|.$$

Proof Let $J' := 1 + (L+3)/2$ and $J := J'(R+1)$ as in the previous corollary. By Lemma 7.1.1, if $K > 1$ is sufficiently small, then whenever $K|g^{m+n}| \geq |g^m| + |g^n|$ either

$$|n| > \frac{R}{R+1}(|m| + |n|), \quad |m| < J', \quad |m| > \frac{R}{R+1}(|m| + |n|), \quad \text{or} \quad |n| < J'.$$

When $|m| + |n| \geq J$, the second condition implies the first condition and the fourth condition implies the third. So, if

$$K|g^{\sum_{i=1}^k m_i}| \geq \sum_{i=1}^k |g^{m_i}|$$

and $\sum_{i=1}^k |m_i| \geq J$ then, since

$$\sum_{i=1}^k |g^{m_i}| = \sum_{i\in I} |g^{m_i}| + \sum_{i\notin I} |g^{m_i}| \geq |g^{\sum_{i\in I} m_i}| + |g^{\sum_{i\notin I} m_i}|,$$

we can set $n = \sum_{i\in I} m_i$ and $m = \sum_{i\notin I} m_i$ to obtain one of the two following conditions:

$$\left|\sum_{i\in I} m_i\right| > \frac{R}{R+1}\sum_{i=1}^k |m_i|, \quad \left|\sum_{i\notin I} m_i\right| > \frac{R}{R+1}\sum_{i=1}^k |m_i|.$$

By symmetry, the first condition holds without loss of generality so it remains to prove

$$\big(K + (K-1)CR^{(-1/\alpha)}\big)|g^{\sum_{i\in I} m_i}| > \sum_{i\in I} |g^{m_i}|.$$

We have

$$\begin{aligned}
|g^{\sum_{i\in I} m_i}| &\geq |g^{\sum_{i=1}^k m_i}| - |g^{\sum_{i\notin I} m_i}| \\
&\geq \frac{1}{K}\sum_{i\in I} |g^{m_i}| + \frac{1}{K}\sum_{i\notin I} |g^{m_i}| - |g^{\sum_{i\notin I} m_i}| \\
&\geq \frac{1}{K}\sum_{i\in I} |g^{m_i}| + \frac{1}{K}|g^{\sum_{i\notin I} m_i}| - |g^{\sum_{i\notin I} m_i}| = \frac{1}{K}\sum_{i\in I} |g^{m_i}| - \frac{K-1}{K}\cdot|g^{\sum_{i\notin I} m_i}|.
\end{aligned}$$

So,

$$\sum_{i\in I} |g^{m_i}| \leq K|g^{\sum_{i\in I} m_i}| + (K-1)|g^{\sum_{i\notin I} m_i}| \leq \left(K + (K-1)\cdot\frac{|g^{\sum_{i\notin I} m_i}|}{|g^{\sum_{i\in I} m_i}|}\right)|g^{\sum_{i\in I} m_i}|.$$

Finally, since we know from Theorem 3.2.9 that the distortion of $\langle g\rangle$ obeys a power law,

$$\left|\sum_{i\notin I} m_i\right| < \frac{1}{R}\left|\sum_{i\in I} m_i\right| \implies \frac{|g^{\sum_{i\notin I} m_i}|}{|g^{\sum_{i\in I} m_i}|} < CR^{(-1/\alpha)}.$$

□

Corollary 7.1.7 *For all $R > 1$ there exist $J, K > 1$ such that if $g \in \{a, x, y\}$, $\sum_{i=1}^k |m_i| \geq J$, and*

$$\sum_{i=1}^k |g^{m_i}| \leq K|g^{\sum_{i=1}^k m_i}|,$$

then $\max_i |m_i| > \frac{R}{R+1}\sum_{i=1}^k |m_i|$.

Proof We may assume $0 < m_1 = \max_i m_i$ and $0 > m_k = \min_i m_i$. Let $I := \{i \mid m_i > 0\}$. We will let $R' \geq 5$ be large enough that $\left(\frac{R'}{R'+1}\right)^2 > \frac{R}{R+1}$. Let $J_{7.1.6}, K_{7.1.6}$ and $J_{7.1.5}, K_{7.1.5}$ be the numbers provided by Corollaries 7.1.6 and 7.1.5, respectively, to R'. (Actually, $J_{7.1.6} = J_{7.1.5}$.) We claim it suffices to take any $J \geq \frac{R'+1}{R'} J_{7.1.5} > J_{7.1.6}$ and any

$$1 < K \leq \min\left\{ K_{7.1.6}, \frac{K_{7.1.5} + C(R')^{(-1/\alpha)}}{1 + C(R')^{(-1/\alpha)}} \right\}.$$

To see this, we first apply Corollary 7.1.6. Since $K \leq K_{7.1.6}$, K is sufficiently small for Corollary 7.1.6, so, since

$$\sum_{i=1}^{k} |m_i| \geq J > J_{7.1.6} \quad \text{and} \quad \sum_{i=1}^{k} |g^{m_i}| \leq K |g^{\sum_{i=1}^{k} m_i}|,$$

there is an $I^* \in \{I, I^c\}$ such that

$$\left| \sum_{i \in I^*} m_i \right| > \frac{R'}{R'+1} \sum_{i=1}^{k} |m_i|$$

and

$$\left(K + (K-1)C(R')^{(-1/\alpha)}\right) |g^{\sum_{i \in I^*} m_i}| > \sum_{i \in I^*} |g^{m_i}|.$$

Now,

$$K \leq \frac{K_{7.1.5} + C(R')^{(-1/\alpha)}}{1 + C(R')^{(-1/\alpha)}}$$

if and only if $(K + (K-1)C(R')^{(-1/\alpha)}) \leq K_{7.1.5}$, so $K_{7.1.5} |g^{\sum_{i \in I^*} m_i}| \geq \sum_{i \in I^*} |g^{m_i}|$. Furthermore,

$$\sum_{i \in I^*} |m_i| = \left| \sum_{i \in I^*} m_i \right| > \frac{R'}{R'+1} \sum_{i=1}^{k} |m_i| \geq \frac{R'}{R'+1} J \geq J_{7.1.5}.$$

Thus, the hypotheses of Corollary 7.1.5 for I^* with respect to R' are satisfied, and we get

$$\max_i |m_i| \geq \max_{i \in I^*} |m_i| > \frac{R'}{R'+1} \left| \sum_{i \in I^*} m_i \right| > \left(\frac{R'}{R'+1}\right)^2 \sum_{i=1}^{k} |m_i| > \frac{R}{R+1} \sum_{i=1}^{k} |m_i|. \qquad \square$$

In the previous results we picked out a "dominant" escape in terms of exponents. In the next result we show it is still dominant in terms of length.

Corollary 7.1.8 *For all $R > 1$ there exist $J, K > 1$ such that if $g \in \{a, x, y\}$ and $\gamma = \gamma_1 \gamma_2 \cdots \gamma_k$ is a concatenation of paths such that*

(1) *for all i, the endpoints of γ_i differ by g^{m_i},*

(2) $J \leq |\gamma| \leq K |g^{\sum_{i=1}^{k} m_i}|$,

then $\max_i |\gamma_i| > \frac{R}{R+1} \cdot |\gamma|$.

Proof Let $R_0 > 1$ and $J_0, K_0 > 1$ be such that Corollary 7.1.7 holds for $(R, J, K) = (R_0, J_0, K_0)$. Suppose $\gamma = \gamma_1\gamma_2\cdots\gamma_k$ satisfies conditions (1) and (2) for some $K \in (1, K_0)$, some $J > J_0^{1/\alpha}KC$ and some $g \in \{a, x, y\}$. Then

$$K|g^{\sum_{i=1}^k m_i}| \geq |\gamma| = \sum_{i=1}^k |\gamma_i| \geq \sum_{i=1}^k |g^{m_i}|$$

and

$$J_0^{1/\alpha} < \frac{J}{KC} \leq \frac{|\gamma|}{KC} \leq \frac{|g^{\sum_{i=1}^k m_i}|}{C} \leq \left|\sum_{i=1}^k m_i\right|^{1/\alpha} \leq \left(\sum_{i=1}^k |m_i|\right)^{1/\alpha}$$

so, by Corollary 7.1.7,

$$\max_i |m_i| > \frac{R_0}{R_0+1}\sum_{i=1}^k |m_i|.$$

If $i = \iota$ maximizes $|m_i|$ then this inequality is equivalent to $\sum_{i\neq\iota} |m_i| < (1/R_0)\cdot|m_\iota|$ and we have

$$\begin{aligned}\frac{1}{C}\cdot|g^{\sum_{i\neq\iota} m_i}| \leq \left|\sum_{i\neq\iota} m_i\right|^{1/\alpha} &\leq \left(\sum_{i\neq\iota}|m_i|\right)^{1/\alpha}\\ &< \frac{1}{R_0^{1/\alpha}}\cdot|m_\iota|^{\frac{1}{\alpha}} \leq \frac{1}{R_0^{1/\alpha}}\cdot|g^{m_\iota}| \leq \frac{1}{R_0^{1/\alpha}}\cdot\left(|g^{\sum_{i\neq\iota} m_i}| + |g^{\sum_{i=1}^k m_i}|\right),\end{aligned}$$

from which we can obtain

$$|g^{\sum_{i\neq\iota} m_i}| < \frac{C}{R_0^{1/\alpha} - C}\cdot|g^{\sum_{i=1}^k m_i}|$$

as long as $R_0 > C^\alpha$. Thus,

$$|g^{\sum_{i=1}^k m_i}| \leq |g^{m_1}| + |g^{\sum_{i\neq\iota} m_i}| < |\gamma_\iota| + \frac{C}{R_0^{1/\alpha} - C}\cdot|g^{\sum_{i=1}^k m_i}|$$

and so

$$\max_i |\gamma_i| \geq |\gamma_\iota| > \frac{R_0^{1/\alpha} - 2C}{R_0^{1/\alpha} - C}\cdot|g^{\sum_{i=1}^k m_i}| \geq \frac{R_0^{1/\alpha} - 2C}{K(R_0^{1/\alpha} - C)}\cdot|\gamma|$$

and the factor

$$\frac{R_0^{1/\alpha} - 2C}{K(R_0^{1/\alpha} - C)}$$

tends to 1 as K tends to 1 and R_0 tends to infinity.

To complete the proof of the corollary, we need to explain how we obtain the $J, K > 1$ of the corollary statement. To obtain $J, K > 1$ we pick $R_0 > C^\alpha$ large enough that

$$\frac{R_0^{1/\alpha} - 2C}{(R_0^{1/\alpha} - C)} > \frac{R}{R+1}.$$

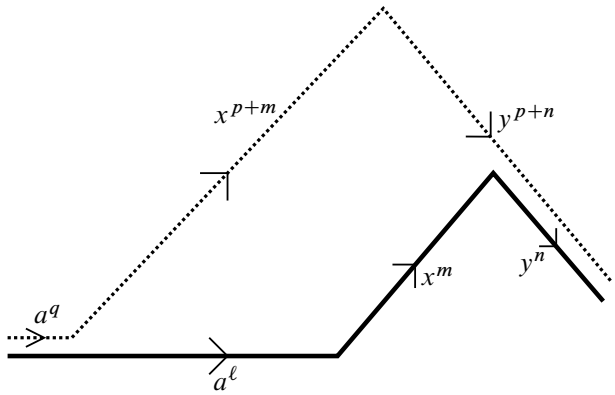


Figure 6: Illustration of the two decompositions of $a^\ell x^m y^n$.

Next, we obtain $K_0, J_0 > 1$ such that the conclusion of Corollary 7.1.7 holds for $(R, J, K) = (R_0, J_0, K_0)$. Finally, we choose $K \in (1, K_0)$ so that

$$\frac{R_0^{1/\alpha} - 2C}{K(R_0^{1/\alpha} - C)} > \frac{R}{R+1}$$

and then choose $J > J_0^{1/\alpha} KC$. □

We will now start allowing escapes of different flavors. The first result says if there is exactly one escape of each flavor and the a-escape is not tiny then it dominates.

Lemma 7.1.9 *For all $R > 1$ there exist $J, K > 1$ such that if $|\ell| + |m| + |n| \geq J$,*

$$|a^\ell| + |x^m| + |y^n| \leq K|a^\ell x^m y^n|,$$

and

$$|\ell| > \frac{|\ell| + |m| + |n|}{R},$$

then $|m|, |n| \leq (|\ell| + |m| + |n|)/R$.

Proof The result for large R yields J and K for which the same result is true for all smaller R, so we may assume in the proof that $R > 2$. We can multiply all three exponents by -1 without changing the hypothesized inequalities, so we may assume $\ell > 0$. Further, by the x/y-symmetry we may assume $|n| \leq |m|$.

There exist integers $p > 0$ and q with $|q| \leq L/2$ such that $\ell = pL + q$ and $|a^\ell| = |q| + |a^{pL}|$, by (3) of Lemma 3.3.1. The strategy for the proof is to compare the approximation of $|a^\ell x^m y^n|$ by $|a^\ell| + |x^m| + |y^n|$ (illustrated by the solid line in Figure 6), to the approximation given by $|a^q| + |x^{m+p}| + |y^{n+p}|$ (illustrated by the dotted line in Figure 6).

Notice that if $\ell \geq L + 1$ then we can verify the bounds

$$\frac{2}{L+2} \cdot \frac{1}{L} \leq \frac{p}{\ell} \leq \frac{L+2}{2} \cdot \frac{1}{L}. \tag{7.1.10}$$

Let $J' = (L+5)/2$ as in Lemma 7.1.1. Let $J := R(1+LJ')$. Let $R' := (1/2)L(L+2)(R-1)$, and let K' be as in Lemma 7.1.1 with respect to R'. We will choose $K > 1$ to be sufficiently small that the below inequalities labelled (♣) and (◇) are contradicted, and the below inequality labelled (♡) is satisfied.

Suppose that $|m| > (\ell + |m| + |n|)/R$. We will derive a contradiction. First observe that we have the bounds

$$\frac{1}{R-1} < \frac{\ell}{|m|} < R-1. \tag{7.1.11}$$

Thus, by combining (7.1.10) and (7.1.11) we deduce that

$$\frac{1}{R'} < \frac{p}{|m|} < \frac{(L+2)(R-1)}{2L} < \frac{L(L+2)(R-1)}{2} = R'. \tag{7.1.12}$$

As both $|m|$ and p are larger than J', we can combine Lemma 7.1.1 with (7.1.12) to see that

$$|x^m| + |x^p| > K'|x^{m+p}|. \tag{7.1.13}$$

We now deduce that

$$\begin{aligned} |q| + |x^p| + |x^m| + |y^p| + |y^n| &\le |a^l| + |x^m| + |y^n| && \text{(by Lemma 3.3.1(3))} \\ &\le K|a^\ell x^m y^n| && \text{(by assumption)} \\ &= K(|a^q x^{m+p} y^{n+p}|) \\ &\le K|q| + K(|x^{p+m}| + |y^{p+n}|) && \text{(triangle inequality).} \end{aligned}$$

Combining this with (7.1.13) yields

$$\Big(1 - \frac{K}{K'}\Big)(|x^p| + |x^m|) + |y^p| + |y^n| - K|y^{p+n}| < \frac{1}{2}(K-1)L. \tag{7.1.14}$$

First, suppose that either $|n| < J'$ or $|n| < p/R'$. Both of these conditions imply $|n| < p$, so by Theorem 3.2.9 we have $|y^n| < C|n|^{1/\alpha} < Cp^{1/\alpha} < C|y^p|$. Continuing from (7.1.14), we have

$$\begin{aligned} \frac{1}{2}(K-1)L &> \Big(1 - \frac{K}{K'}\Big)(|x^p| + |x^m|) + |y^p| + |y^n| - K|y^{p+n}| \\ &\ge \Big(1 - \frac{K}{K'}\Big)(|x^p| + |x^m|) - (K-1)(|y^p| + |y^n|) \\ &\ge \Big(1 - \frac{K}{K'}\Big)(|x^p| + |x^m|) - (K-1)|y^p|(1+C) \\ &= \Big(1 - \frac{K}{K'} - (K-1)(1+C)\Big)|x^p| + \Big(1 - \frac{K}{K'}\Big)|x^m|. \end{aligned}$$

To ensure that the coefficients in this final expression are nonnegative we will have assumed that K is chosen to satisfy the inequalities

$$1 < K < \min\left\{K',\ 1 + \frac{K'-1}{(C+1)K'+1}\right\}. \tag{♡}$$

By Theorem 3.2.9, the lower bounds on p and $|m|$ imply $|x^p|, |x^m| > (J')^{1/\alpha}$, so

$$\frac{1}{2}(K-1)L > \Big(1 - \frac{K}{K'} - (K-1)(1+C) + 1 - \frac{K}{K'}\Big)(J')^{1/\alpha}.$$

Solving for K gives us

$$K > \frac{LK' + 2(J')^{1/\alpha} \cdot (K'(1+C) + 2K')}{LK' + 2(J')^{1/\alpha} \cdot (K'(1+C) + 2)} > 1. \tag{♣}$$

Having assumed that $K > 1$ is chosen to be small enough that (♣) does not hold, we must in fact have $|n| \geq J'$ and $|n| \geq p/R'$. Then Lemma 7.1.1 implies $|y^p| + |y^n| > K'|y^{p+n}|$. In this case, continuing from (7.1.14) we have

$$\frac{1}{2}(K-1)L > \left(1 - \frac{K}{K'}\right)(|x^p| + |x^m| + |y^p| + |y^n|) > \left(1 - \frac{K}{K'}\right)(4(J')^{1/\alpha}).$$

Solving for K gives

$$K > \frac{8K'(J')^{1/\alpha} + LK'}{8(J')^{1/\alpha} + LK'} > 1. \tag{◇}$$

By having assumed that $K > 1$ is chosen small enough that (◇) does not hold, this cannot happen. Thus, for $J := R(1 + LJ')$ and K as specified, the hypotheses imply $|n| \leq |m| \leq (|\ell| + |m| + |n|)/R$. □

The next result, Lemma 7.1.15 is a stepping stone. It says that if we have a mixture of different flavors of escapes, then whenever one flavor makes up a nontrivial amount of the total, then there is an escape of that flavor that is dominant *among the other escapes of that flavor*. Moreover, if the a-escapes make up a nontrivial amount of the total then the x-escapes and y-escapes are negligible.

Lemma 7.1.15 *For all $R > 1$ there exist $J, K > 1$ such that if $T = \sum_i |\ell_i| + \sum_j |m_j| + \sum_k |n_k| \geq J$ and*

$$\sum_i |a^{\ell_i}| + \sum_j |x^{m_j}| + \sum_k |y^{n_k}| \leq K|a^{\sum_i \ell_i} x^{\sum_j m_j} y^{\sum_k n_k}| \tag{7.1.16}$$

then the following implications hold:

$$\sum_i |\ell_i| > \frac{1}{R} \cdot T \implies \max_i |\ell_i| > \frac{R-1}{R} \sum_i |\ell_i|,$$

$$\sum_j |m_j| > \frac{1}{R} \cdot T \implies \max_j |m_j| > \frac{R-1}{R} \sum_j |m_j|,$$

$$\sum_k |n_k| > \frac{1}{R} \cdot T \implies \max_k |n_k| > \frac{R-1}{R} \sum_k |n_k|.$$

Moreover, if $\sum_i |\ell_i| > \frac{1}{R} \cdot T$ then $\sum_j |m_j| \leq \frac{1}{R} \cdot T$ and $\sum_k |n_k| \leq \frac{1}{R} \cdot T$.

Proof We will prove the first implication. The remaining implications are proved similarly. Assume that (7.1.16) and the left-hand side of the first implication holds with

$$1 < K < 1 + \left(\frac{1}{2}\right)^{1+(\alpha-1)/\alpha} \cdot \frac{1}{CR^{1/\alpha}}. \tag{7.1.17}$$

Claim 7.1.18 $$|x^{\sum_j m_j} y^{\sum_k n_k}| < 2^{1-1/\alpha} C \left(\sum_j |m_j| + \sum_k |n_k|\right)^{1/\alpha}.$$

Proof of claim We observe that

$$\begin{aligned}\left(\sum_j |m_j| + \sum_k |n_k|\right)^{1/\alpha} &\geq \left(\left|\sum_j m_j\right| + \left|\sum_k n_k\right|\right)^{1/\alpha} \\ &\geq \left(\frac{1}{2}\right)^{(\alpha-1)/\alpha}\left(\left|\sum_j m_j\right|^{1/\alpha} + \left|\sum_k n_k\right|^{1/\alpha}\right) \quad \text{(by Fact 1.0.1)} \\ &> \frac{1}{C}\left(\frac{1}{2}\right)^{(\alpha-1)/\alpha}\left(|x^{\sum_j m_j}| + |y^{\sum_k n_k}|\right) \quad \text{(by Theorem 3.2.9)} \\ &\geq \frac{1}{C}\left(\frac{1}{2}\right)^{(\alpha-1)/\alpha}|x^{\sum_j m_j} y^{\sum_k n_k}|.\end{aligned}$$

□

Claim 7.1.19
$$\left|\sum_i \ell_i\right| > \frac{1}{(4C)^\alpha R}\cdot T.$$

Proof of claim For the sake of finding a contradiction, assume

$$\left|\sum_i \ell_i\right| \leq \frac{1}{(4C)^\alpha R}\cdot T.$$

Then, since $T < R\sum_i |\ell_i|$,

$$\left|\sum_i \ell_i\right| < \frac{1}{(4C)^\alpha}\sum_i |\ell_i|.$$

Then we have

$$\begin{aligned}|a^{\sum_i \ell_i}| &< C\left|\sum_i \ell_i\right|^{1/\alpha} && \text{(by Theorem 3.2.9)} \\ &< C\left(\frac{1}{(4C)^\alpha}\sum_i |\ell_i|\right)^{1/\alpha} && \text{(by the above)} \\ &\leq \frac{1}{4}\sum_i |\ell_i|^{1/\alpha} && \text{(by Fact 1.0.1)} \\ &< \frac{1}{4}\sum_i |a^{\ell_i}| && \text{(by Theorem 3.2.9).}\end{aligned}$$

Thus,

$$\begin{aligned}\sum_i |a^{\ell_i}| + |x^{\sum_j m_j} y^{\sum_k n_k}| &\leq \sum_i |a^{\ell_i}| + \sum_j |x^{m_j}| + \sum_k |y^{n_k}| \\ &\leq K|a^{\sum_i \ell_i} x^{\sum_j m_j} y^{\sum_k n_k}| \\ &\leq K|a^{\sum_i \ell_i}| + K|x^{\sum_j m_j} y^{\sum_k n_k}| < \frac{K}{4}\sum_i |a^{\ell_i}| + K|x^{\sum_j m_j} y^{\sum_k n_k}|,\end{aligned}$$

which implies

$$\frac{1}{2}\sum_i |a^{\ell_i}| < (K-1)|x^{\sum_j m_j} y^{\sum_k n_k}|$$

since (7.1.17) implies $K < 2$.

On the other hand, by Theorem 3.2.9, Claim 7.1.18 and Fact 1.0.1, we have

$$\begin{aligned}
\sum_i |a^{\ell_i}| &> \sum_i |\ell_i|^{1/\alpha} && \text{(by Theorem 3.2.9)}\\
&\geq \left(\sum_i |\ell_i|\right)^{1/\alpha} && \text{(by Fact 1.0.1)}\\
&> \frac{1}{R^{1/\alpha}} \cdot T^{1/\alpha} && \text{(by assumption)}\\
&\geq \frac{1}{R^{1/\alpha}}\left(\sum_j |m_j| + \sum_k |n_k|\right)^{1/\alpha}\\
&\geq \frac{1}{CR^{1/\alpha}}\left(\frac{1}{2}\right)^{\frac{\alpha-1}{\alpha}} |x^{\sum_j m_j} y^{\sum_k n_k}| && \text{(by Claim 7.1.18).}
\end{aligned}$$

Thus,

$$\frac{1}{2}\sum_i |a^{\ell_i}| < (K-1)|x^{\sum_j m_j} y^{\sum_k n_k}| < 2^{1-1/\alpha} C R^{1/\alpha}(K-1)\sum_i |a^{\ell_i}| < \frac{1}{2}\sum_i |a^{\ell_i}|$$

where the last inequality follows from (7.1.17). This is a contradiction. □

Then we have

$$\begin{aligned}
|a^{\sum_i \ell_i}| &> |\sum_i \ell_i|^{1/\alpha} && \text{(by Theorem 3.2.9)}\\
&> \frac{1}{4CR^{1/\alpha}} \cdot T^{1/\alpha} && \text{(by Claim 7.1.19)}\\
&\geq \frac{1}{4CR^{1/\alpha}} \cdot \left(\sum_j |m_j| + \sum_k |n_k|\right)^{1/\alpha} && \text{(by assumption)}\\
&\geq \frac{1}{4C^2R^{1/\alpha}} \cdot \left(\frac{1}{2}\right)^{\frac{\alpha-1}{\alpha}} \cdot |x^{\sum_j m_j} y^{\sum_k n_k}| && \text{(by Claim 7.1.18).}
\end{aligned}$$

This gives us

$$|x^{\sum_j m_j} y^{\sum_k n_k}| < 2^{3-1/\alpha} C^2 R^{1/\alpha} |a^{\sum_i \ell_i}|.$$

On the other hand,

$$\begin{aligned}
\sum_i |a^{\ell_i}| + |x^{\sum_j m_j} y^{\sum_k n_k}| &\leq \sum_i |a^{\ell_i}| + \sum_j |x^{m_j}| + \sum_k |y^{n_k}|\\
&\leq K|a^{\sum_i \ell_j} x^{\sum_j m_j} y^{\sum_k n_k}|\\
&\leq K|a^{\sum_i \ell_j}| + K|x^{\sum_j m_j} y^{\sum_k n_k}|.
\end{aligned}$$

So,

$$\begin{aligned}\sum_i |a^{\ell_i}| &\le K|a^{\sum_i \ell_j}| + (K-1)|x^{\sum_j m_j} y^{\sum_k n_k}| \\ &< K|a^{\sum_i \ell_j}| + 2^{3-1/\alpha} C^2 R^{1/\alpha}(K-1)|a^{\sum_i \ell_i}| \\ &= \big(K + 2^{3-1/\alpha} C^2 R^{1/\alpha}(K-1)\big)|a^{\sum_i \ell_j}| \\ &\xrightarrow{K\to 1} |a^{\sum_i \ell_j}|. \end{aligned} \tag{†}$$

Noting that if $T \ge J$ then $\sum_i |\ell_i| > (1/R)\cdot T \ge J/R$. Thus, by Corollary 7.1.7, if J is large enough and $K > 1$ is small enough that (†) satisfies the requisite assumption — all depending only on R and L — then $\max_i |\ell_i| > \frac{R-1}{R}\sum_i |\ell_i|$.

We now assume $\sum_i |\ell_i| > \frac{1}{R}\cdot T$ to prove the "moreover part". We will show that $\sum_j |m_j| \le \frac{1}{R}\cdot T$. A similar argument can be used to show that $\sum_k |n_k| \le \frac{1}{R}\cdot T$. For the sake of finding a contradiction, assume $\sum_j |m_j| > \frac{1}{R}\cdot T$. Without loss of generality, we have $\max_i |\ell_i| = |\ell_1|$ and $\max_j |m_j| = |m_1|$. Then

$$|\ell_1| > \frac{R-1}{R}\sum_i |\ell_i| > \frac{R-1}{R^2}\cdot T \quad \text{and} \quad |m_1| > \frac{R-1}{R}\sum_j |m_j| > \frac{R-1}{R^2}\cdot T. \tag{7.1.20}$$

Then, by Fact 1.0.1 and Theorem 3.2.9,

$$\begin{aligned} |a^{\ell_1}| &> |\ell_1|^{1/\alpha} && \text{(by Theorem 3.2.9)} \\ &> \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot T^{1/\alpha} && \text{(by (7.1.20))} \\ &\ge \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\left|\sum_{i\ge 2}\ell_i\right| + \left|\sum_{j\ge 2} m_j\right| + \left|\sum_k n_k\right|\right)^{1/\alpha} && \text{(by assumption)} \\ &\ge \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\frac{1}{3}\right)^{\frac{\alpha-1}{\alpha}}\cdot\left(\left|\sum_{i\ge 2}\ell_i\right|^{1/\alpha} + \left|\sum_{j\ge 2} m_j\right|^{1/\alpha} + \left|\sum_k n_k\right|^{1/\alpha}\right) && \text{(by Fact 1.0.1)} \\ &> \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\frac{1}{3}\right)^{\frac{\alpha-1}{\alpha}}\cdot\frac{1}{C}\cdot\big(|a^{\sum_{i\ge 2}\ell_i}| + |x^{\sum_{j\ge 2} m_j}| + |y^{\sum_k n_k}|\big) && \text{(by Theorem 3.2.9)} \\ &\ge \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\frac{1}{3}\right)^{\frac{\alpha-1}{\alpha}}\cdot\frac{1}{C}\cdot|a^{\sum_{i\ge 2}\ell_i} x^{\sum_{j\ge 2} m_j} y^{\sum_k n_k}| \end{aligned}$$

and, similarly,

$$|x^{m_1}| > \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\frac{1}{3}\right)^{(\alpha-1)/\alpha}\cdot\frac{1}{C}\cdot|a^{\sum_{i\ge 2}\ell_i} x^{\sum_{j\ge 2} m_j} y^{\sum_k n_k}|.$$

By Lemma 3.3.1(4), for sufficiently small $K' > 1$ we have

$$|a^{\ell_1} x^{m_1}| > \frac{1}{K'}(|a^{\ell_1}| + |x^{m_1}|) = \left(\frac{R-1}{R^2}\right)^{1/\alpha}\cdot\left(\frac{1}{3}\right)^{(\alpha-1)/\alpha}\cdot\frac{2}{CK'}\cdot|a^{\sum_{i\ge 2}\ell_i} x^{\sum_{j\ge 2} m_j} y^{\sum_k n_k}|. \tag{7.1.21}$$

But,

$$\begin{aligned}|a^{\ell_1}|+|x^{m_1}|+|a^{\sum_{i\geq 2}\ell_i}x^{\sum_{j\geq 2}m_j}y^{\sum_k n_k}| &\leq \sum_i |a^{\ell_i}|+\sum_j |x^{m_j}|+\sum_k |y^{n_k}|\\ &\leq K|a^{\sum_i \ell_i}x^{\sum_j m_j}y^{\sum_k n_k}|\\ &\leq K|a^{\ell_1}x^{m_1}|+K|a^{\sum_{i\geq 2}\ell_i}x^{\sum_{j\geq 2}m_j}y^{\sum_k n_k}|,\end{aligned}$$

so

$$\begin{aligned}|a^{\ell_1}|+|x^{m_1}| &\leq K|a^{\ell_1}x^{m_1}|+(K-1)|a^{\sum_{i\geq 2}\ell_i}x^{\sum_{j\geq 2}m_j}y^{\sum_k n_k}|\\ &< K|a^{\ell_1}x^{m_1}|+3^{\frac{\alpha-1}{\alpha}}\left(\frac{R^2}{R-1}\right)^{1/\alpha}\cdot\frac{CK'}{2}(K-1)|a^{\ell_1}x^{m_1}| \qquad \text{(by (7.1.21))}\\ &= \left(K+3^{\frac{\alpha-1}{\alpha}}\left(\frac{R^2}{R-1}\right)^{1/\alpha}\cdot\frac{CK'}{2}(K-1)\right)|a^{\ell_1}x^{m_1}| \xrightarrow{K\to 1} |a^{\ell_1}x^{m_1}|.\end{aligned}$$

Also, if $T\geq J$ then

$$|\ell_1|+|m_1| > \frac{2(R-1)}{R^2}\cdot T \geq \frac{2(R-1)}{R^2}\cdot TJ.$$

Thus, by Lemma 7.1.9, if J is large enough and $K>1$ is small enough — depending only on R, K', C and L — then $|m_1|<\frac{1}{R}|\ell_1|<\frac{1}{R}\cdot T$, which is a contradiction. Since K' and C depend only on L, we are done. □

Lemma 7.1.22 *If $R\geq 2$ and $|n_1|>\frac{R-1}{R}\sum_{i=1}^j |n_i|$ then*

$$\left|\frac{\sum_{i=1}^j n_i}{n_1}-1\right|<\frac{1}{R-1} \quad \text{and} \quad \frac{R-2}{R}<\frac{\left|\sum_{i=1}^j n_i\right|}{\sum_{i=1}^j |n_i|}\leq 1.$$

Proof The first and last inequalities are easy. For the remaining inequality, $|n_1|>(R-1)\sum_{i=2}^j |n_i|$, so

$$\frac{|\sum_{i=1}^j n_i|}{\sum_{i=1}^j |n_i|}\geq\frac{|n_1|-\sum_{i=2}^j |n_i|}{|n_1|+\sum_{i=2}^j |n_i|}>\frac{|n_1|-\frac{1}{R-1}|n_1|}{|n_1|+\frac{1}{R-1}|n_1|}=\frac{R-2}{R}.$$ □

Now we put it all together. Proposition 7.1.23 and Theorem 7.1.24 say basically the same thing: either there is a single dominant escape or there are two and they are an x-escape and a y-escape. The difference in the two results is that in Proposition 7.1.23 "dominance" is in terms of magnitude of exponent and in Theorem 7.1.24 it is in terms of length.

Proposition 7.1.23 *Suppose $R\geq 3L+4$ and J and K are as in Lemma 7.1.15, and*

$$\sum_i |a^{\ell_i}|+\sum_j |x^{m_j}|+\sum_k |y^{n_k}|\leq K|a^{\sum_i \ell_i}x^{\sum_j m_j}y^{\sum_k n_k}|.$$

Suppose $T:=\sum_i|\ell_i|+\sum_j|m_j|+\sum_k|n_k|\geq J$.

(1) *If $\sum_i|\ell_i|>T/R$ or exactly one of $\sum_j|m_j|$ and $\sum_k|n_k|$ is strictly greater than T/R then there is one maximal term whose absolute value is strictly greater than $((R-1)(R-2)/R^2)T$.*

(2) *If $\sum_j|m_j|>T/R$ and $\sum_k|n_k|>T/R$ then $\max_j|m_j|+\max_k|n_k|>\left(\frac{R-1}{R}\right)^2 T$.*

(3) *If* $x^z = a^{\sum_i \ell_i} x^{\sum_j m_j} y^{\sum_k n_k}$, *then* $\sum_j |m_j| > T/R$ *and* $\sum_i |\ell_i|, \sum_k |n_k| \le T/R$.

(4) *If* $a^z = a^{\sum_i \ell_i} x^{\sum_j m_j} y^{\sum_k n_k}$, *then either*

$$\sum_i |\ell_i| > \frac{T}{R} \quad \text{and} \quad \sum_j |m_j|, \sum_k |n_k| \le \frac{T}{R},$$

or

$$\sum_j |m_j|, \sum_k |n_k| > \frac{T}{R}, \quad \sum_i |\ell_i| \le \frac{T}{R}, \quad \max_j |m_j|, \max_k |n_k| > \frac{1}{2} \cdot \frac{(R-1)^2(R-2)}{R^3} \cdot T,$$
$$\text{and} \quad \left(\frac{R-2}{R+2}\right)^2 < \frac{\max_j |m_j|}{\max_k |n_k|} < \left(\frac{R+2}{R-2}\right)^2.$$

Proof The first case follows from Lemma 7.1.15, since if two of the sums are bounded above by T/R then the third is bounded below by $\frac{R-2}{R}T$.

The second case also follows immediately from Lemma 7.1.15, since

$$\max_j |m_j| + \max_k |n_k| > \frac{R-1}{R}\left(\sum_j |m_j| + \sum_k |n_k|\right) \ge \frac{R-1}{R}\left(T - \sum_i |\ell_i|\right) \ge \frac{R-1}{R}\left(1 - \frac{1}{R}\right)T.$$

In case (3), we have $\sum_i \ell_i = pL$, $z = p + \sum_j m_j$ and $0 = p + \sum_k n_k$. Suppose that $\sum_i |\ell_i| > T/R$. By Lemma 7.1.15,

$$\max_i |\ell_i| > \frac{R-1}{R} \sum_k |\ell_k| \quad \text{and} \quad \sum_j |m_j|, \sum_k |n_k| \le \frac{T}{R},$$

the latter implying $\sum_i |\ell_i| \ge \frac{R-2}{R}T$. Applying Lemma 7.1.22,

$$\frac{T}{R} \ge \sum_k |n_k| \ge \left|\sum_k n_k\right| = |p| = \frac{|\sum_i \ell_i|}{|L|} > \frac{1}{L} \cdot \frac{R-2}{R} \cdot \sum_i |\ell_i| \ge \frac{1}{L} \cdot \left(\frac{R-2}{R}\right)^2 T.$$

This gives a contradiction if $R \ge 3L + 4$.

Suppose that $\sum_k |n_k| > T/R$. By Lemma 7.1.15, $\sum_i |\ell_i| \le T/R$ and $\max_k |n_k| > \frac{R-1}{R} \sum_k |n_k|$ so, applying Lemma 7.1.22,

$$\frac{T}{R} \ge \sum_i |\ell_i| \ge |\sum_i \ell_i| = L|p| = L|\sum_k n_k| > L \cdot \frac{R-2}{R} \cdot \sum_k |n_k| > L \cdot \frac{R-2}{R} \cdot \frac{T}{R}$$

This implies $R < 2L/(L-1) \le 12/5$, which is a contradiction.

Since $R > 3$, we are left with the possibility that only $\sum_j |m_j|$ is strictly greater than T/R.

In case (4), if $\sum_i |\ell_i| > T/R$ then Lemma 7.1.15 lets us revert to case (1), so assume $\sum_i |\ell_i| \le T/R$. Without loss of generality, assume $\sum_j |m_j| \ge \sum_k |n_k|$, so that

$$\sum_j |m_j| \ge \frac{R-1}{2R} \cdot T > \frac{T}{R}.$$

By Lemma 7.1.15,

$$\max_j |m_j| > \frac{R-1}{R}\sum_j |m_j| \geq \frac{1}{2}\cdot\left(\frac{R-1}{R}\right)^2\cdot T.$$

The constraint $a^z = a^{\sum_i \ell_i} x^{\sum_j m_j} y^{\sum_k n_k}$ implies $\sum_j m_j = \sum_k n_k = \left(z - \sum_i \ell_i\right)/L$, so, applying Lemma 7.1.22,

$$\sum_k |n_k| \geq \left|\sum_k n_k\right| = \left|\sum_j m_j\right| \geq \frac{R-2}{R}\sum_j |m_j| \geq \frac{(R-1)(R-2)}{2R^2}\cdot T.$$

The given lower bound on R is enough to imply that this quantity is greater than T/R, so Lemma 7.1.15 applies to give

$$\max_k |n_k| > \frac{R-1}{R}\sum_k |n_k| \geq \frac{(R-1)^2(R-2)}{2R^3}\cdot T.$$

It remains to estimate $\max_j |m_j| / \max_k |n_k|$:

$$\frac{\max_j |m_j|}{\max_k |n_k|} \leq \frac{\sum_j |m_j|}{\frac{R-1}{R}\sum_k |n_k|} \leq \frac{T - \sum_k |n_k|}{\frac{R-1}{R}\sum_k |n_k|} \leq \frac{R}{R-1}\left(\frac{2R^2}{(R-1)(R-2)} - 1\right) < \left(\frac{R+2}{R-2}\right)^2.$$

The last inequality is true, in particular, when $R > 2$. A similar computation gives the lower bound. □

Theorem 7.1.24 *For any $R' > 1$ there exist $J' > 1$ and $2 > K > 1$ such that if γ is a K-bi-Lipschitz path with endpoints in $\langle a, x, y\rangle$ then either $|\gamma| \leq J'$ or γ has either one or two long escapes that collectively account for at least $(R'-1)/R'$ fraction of its length. Furthermore, in the case of two long escapes ϵ_0 and ϵ_1 they are one x-escape and one y-escape.*

If the endpoints of γ differ by a power of x or y, then there is only one long escape, and it is of the corresponding flavor, and if the endpoints of γ differ by a power of a then either there is one long a-escape or there is one long x-escape and one long y-escape.

Proof Since γ is K-bi-Lipschitz for $K < 2$, γ does not use any x or y edges, so toral subpaths are just powers of a. Thus, the a-escapes and/or maximal toral subpaths have endpoints that differ by $a^{\ell_1}, a^{\ell_2}, \dots$ and the x-escapes and y-escapes have endpoints that differ by $x^{m_1}, x^{m_2}, \dots$ and $y^{n_1}, y^{n_2}, \dots$, respectively.

Given R' we will determine in the course of the proof a bound such that the rest of the argument works provided R is larger than that bound and K and J are chosen as in Lemma 7.1.15 with respect to R. Since we are allowed to enlarge R we may assume $R \geq 3L + 4$ so that Proposition 7.1.23 applies. Let T be the usual sum of absolute values. Let $J' := JK$. If $T < J$ then $|\gamma| \leq KT < KJ$, so it suffices to take $|\gamma| \geq J'$ to guarantee $T \geq J$ so that we may apply Lemma 7.1.15.

By Lemma 7.1.15 there are either one or two long escapes. Proposition 7.1.23 bounds these in terms of the exponent of the difference between their endpoints. It remains only to recast these bounds in terms of path length.

Suppose there are two long escapes ϵ_1 and ϵ_2, the strategy in the other cases being similar. By Proposition 7.1.23, $\max_j |m_j| + \max_k |n_k| > \left(\frac{R-1}{R}\right)^2 T$. The complement of these two escapes in

γ consists of at most three subpaths, δ_1, δ_2, and δ_3. Let $I_{a,1}$ be the set of indices i such that the escape of γ corresponding to the term a^{ℓ_i} belongs to δ_1. Similarly, define $I_{a,2}$, $I_{a,3}$, $I_{x,1}$, etc. Then

$$\begin{aligned}
\frac{|\delta_1|+|\delta_2|+|\delta_3|}{|\gamma|} &\le \frac{K\sum_{h=1}^{3}|a^{\sum_{i\in I_{a,h}}\ell_i}x^{\sum_{j\in I_{x,h}}m_j}y^{\sum_{k\in I_{y,h}}n_k}|}{|\delta_1|+|\epsilon_1|+|\delta_2|+|\epsilon_2|+|\delta_3|}\\
&\le \frac{K\sum_{h=1}^{3}|a^{\sum_{i\in I_{a,h}}\ell_i}|+|x^{\sum_{j\in I_{x,h}}m_j}|+|y^{\sum_{k\in I_{y,h}}n_k}|}{|\epsilon_1|+|\epsilon_2|}\\
&\le \frac{KC\sum_{h=1}^{3}\big|\sum_{i\in I_{a,h}}\ell_i\big|^{1/\alpha}+\big|\sum_{j\in I_{x,h}}m_j\big|^{1/\alpha}+\big|\sum_{k\in I_{y,h}}n_k\big|^{1/\alpha}}{(\max_j|m_j|)^{1/\alpha}+(\max_k|n_k|)^{1/\alpha}}\\
&\le \frac{KC\cdot 9^{1-1/\alpha}\big(\sum_{h=1}^{3}\big|\sum_{i\in I_{a,h}}\ell_i\big|+\big|\sum_{j\in I_{x,h}}m_j\big|+\big|\sum_{k\in I_{y,h}}n_k\big|\big)^{1/\alpha}}{(\max_j|m_j|+\max_k|n_k|)^{1/\alpha}}\\
&\le \frac{KC\cdot 9^{1-1/\alpha}\big(\sum_{h=1}^{3}\sum_{i\in I_{a,h}}|\ell_i|+\sum_{j\in I_{x,h}}|m_j|+\sum_{k\in I_{y,h}}|n_k|\big)^{1/\alpha}}{(\max_j|m_j|+\max_k|n_k|)^{1/\alpha}}\\
&< \frac{KC\cdot 9^{1-1/\alpha}\big(\big(1-\big(\frac{R-1}{R}\big)^2\big)T\big)^{1/\alpha}}{\big(\big(\frac{R-1}{R}\big)^2T\big)^{1/\alpha}}\\
&= KC\cdot 9^{1-1/\alpha}\left(\frac{2R-1}{(R-1)^2}\right)^{1/\alpha}\\
&< 2C\cdot 9^{1-1/\alpha}\left(\frac{2R-1}{(R-1)^2}\right)^{1/\alpha}\xrightarrow{R\to\infty}0.
\end{aligned}$$

Thus, given $R'>1$, for sufficiently large R we have $(|\delta_1|+|\delta_2|+|\delta_3|)/|\gamma|<1/R'$, or, conversely, $(|\epsilon_1|+|\epsilon_2|)/|\gamma|>(R'-1)/R'$. □

7.2 The central "region" of a bi-Lipschitz cycle

The goal of this subsection is to show, as promised in Section 5.1, that the central region has at most four long sides.

Definition 7.2.1 Let $\gamma\colon S\to X$ be a cycle such that $\gamma\cap H\neq\varnothing$ and such that all of the escapes of γ from H have length at most $|\gamma|/2$. Let $\{\gamma|_{P_1},\gamma|_{P_2},\dots,\gamma|_{P_k}\}$ be a set of distinct escapes from H. The set is *manageable* if one of the following conditions referencing Figure 7 holds:

- $k=2$ and

(7A) $\gamma|_{P_1}$ and $\gamma|_{P_2}$ are of the same flavor with opposite sign.

- $k=3$ and

(7B) the $\gamma|_{P_i}$ are all of the same flavor but are not all of the same sign, or

(7C) there is exactly one escape of each flavor among the $\gamma|_{P_i}$ and the a-escape has sign opposite to that of the x-escape and the y-escape.

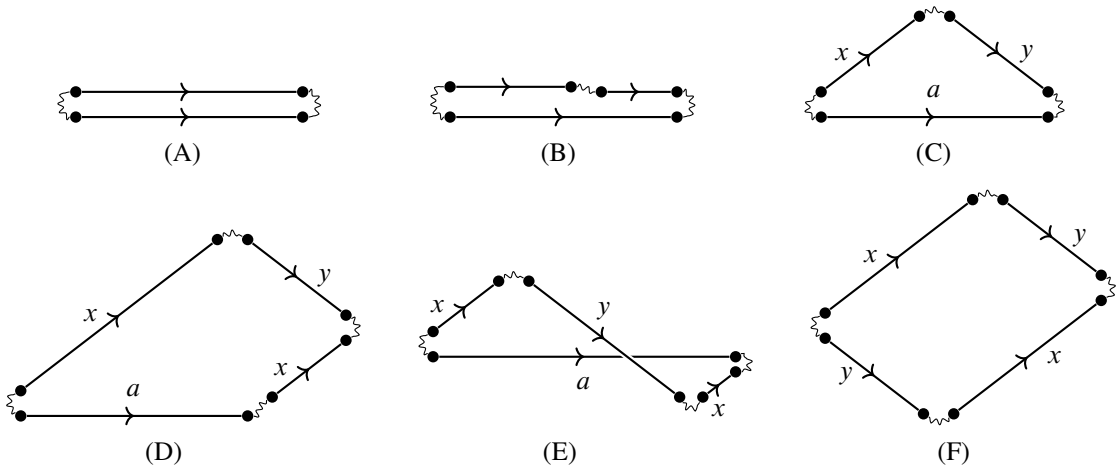


Figure 7: The possible configurations, up to symmetry, of the long escapes of a central region of the minimal area disk diagram of a bi-Lipschitz cycle. The directed segments represent the oriented corridors of the long escapes with labels indicating the flavor of the escape. (In the configurations without labels, all long escapes must have the same flavor but each of the three flavors is possible.) The squiggly paths represent the complement of the long escapes in the cycle.

- $k = 4$ and three consecutive escapes, $\gamma|_{P_{i_1}}$, $\gamma|_{P_{i_2}}$, and $\gamma|_{P_{i_3}}$, alternate between x and y flavors and the fourth escape $\gamma|_{P_{i_4}}$ has sign opposite to $\gamma|_{P_{i_2}}$ and either

(7D) $\gamma|_{P_{i_4}}$ is an a-escape and the escapes $\gamma|_{P_{i_1}}$ and $\gamma|_{P_{i_3}}$ have opposite sign,

(7E) $\gamma|_{P_{i_4}}$ is an a-escape and the escapes $\gamma|_{P_{i_1}}$ and $\gamma|_{P_{i_3}}$ have the same sign, or

(7F) $\gamma|_{P_{i_4}}$ is of the same flavor as $\gamma|_{P_{i_2}}$ and the escapes $\gamma|_{P_{i_1}}$ and $\gamma|_{P_{i_3}}$ have opposite sign.

Proposition 7.2.2 *For any $M \geq 24C + 1$ and $R > M - 1$ there are $J > 0$ and a $K \in (1, 2)$ such that the following holds.*

Let $\gamma\colon S \to X$ be a K-bi-Lipschitz cycle of length $|\gamma| \geq J$ such that $\gamma \cap H \neq \varnothing$ and every escape of γ from H has length at most $|\gamma|/2$. Let $\gamma|_{P_1}, \gamma|_{P_2}, \ldots, \gamma|_{P_k}$ be distinct escapes from H such that $k \leq 5$ and $\sum_{i=1}^{k} |\gamma|_{P_i}| \geq \frac{R-1}{R} \cdot |\gamma|$. Then

(I) *$k > 1$, and*

(II) *if $k \in \{2, 3, 4, 5\}$ then either the $\gamma|_{P_i}$ are manageable or $|\gamma|_{P_i}| \leq \frac{M-1}{R} \cdot |\gamma|$ for some i.*

Proof Given M and R as in the statement, let J and K be large enough with respect to R so that Theorem 7.1.24 is satisfied, and so that $J \geq R$.

We have $\sum_{i=1}^{k} |\gamma|_{P_i}| \geq \frac{R-1}{R} \cdot |\gamma| > \frac{1}{2} \cdot J > 0$, so $k > 0$. If $k = 1$ then by hypothesis

$$\frac{|\gamma|}{2} \geq |\gamma|_{P_1}| \geq \frac{R-1}{R} \cdot |\gamma| > \frac{1}{2} \cdot |\gamma|,$$

a contradiction. This establishes (I).

Let $\{Q_i\}_{i=1}^k$ be the set of closures of components of $S \setminus \bigcup_{i=1}^k P_i$ and let $\beta\colon Q \to X$ be the path obtained by concatenating translates of the $\gamma|_{Q_i}$. Note that if the endpoints of the concatenation of translates of the $\gamma|_{P_i}$ differ by $g \in G_L$ then the endpoints of β differ by g^{-1}.

For the cases $k \in \{2, 3, 4\}$, it will suffice to prove that if the $\gamma|_{P_i}$ are not manageable then $|\beta| \geq \frac{1}{2}|r|^{1/\alpha} - 5$ where g^r is the difference between the endpoints of some $\gamma|_{P_i}$ with $g \in \{a, x, y\}$. Indeed, then we have

$$\frac{|\gamma|}{R} \geq |\gamma| - \sum_{i=1}^{k} \left|\gamma|_{P_i}\right| = |\beta| \geq \frac{1}{2}|r|^{1/\alpha} - 5 \geq \frac{1}{2C} \cdot |g^r| - 5 \geq \frac{1}{2KC} \cdot \left|\gamma|_{P_i}\right| - 5$$

so, as long as $J \geq R$, we have

$$\left|\gamma|_{P_i}\right| \leq 2KC \cdot \left(\frac{|\gamma|}{R} + 5\right) < 4C \cdot \left(\frac{1}{R} + \frac{5}{J}\right) \cdot |\gamma| \leq \frac{24C}{R} \cdot |\gamma|,$$

and so we need only ensure that $M \geq 24C + 1$ to obtain $\left|\gamma|_{P_i}\right| \leq \frac{M-1}{R} \cdot |\gamma|$.

Consider the case $k = 2$. If the $\gamma|_{P_i}$ are unmanageable then either they differ in flavor or they have the same sign. Suppose that $\gamma|_{P_1}$ and $\gamma|_{P_2}$ have different flavors. Without loss of generality, $\gamma|_{P_1}$ is an x-escape and $\gamma|_{P_2}$ is either a y-escape or an a-escape. Let x^m be the difference between the endpoints of $\gamma|_{P_1}$ and let y^n and a^ℓ be the same for $\gamma|_{P_2}$ in each of the two cases. In the first case, Corollary 3.3.2 tells us that

$$|\beta| \geq |x^m y^n| \geq |m|^{1/\alpha} - 5,$$

so we are done. In the second case, choose p and q so that $\ell = pL + q$ with $0 \leq |q| < L$ and either $q = 0$ or q and ℓ have opposite sign. Corollary 3.3.2 tells us that

$$|\beta| \geq |a^\ell x^m| \geq |p|^{1/\alpha} - 5,$$

but our choice of p and q ensures that $|p| = |\ell - q|/L = (|\ell| + |q|)/L \geq |\ell|/L$ and so we have

$$|\beta| \geq \frac{|\ell|^{1/\alpha}}{L^{1/\alpha}} - 5 = \frac{1}{2}|\ell|^{1/\alpha} - 5.$$

It remains to consider the case where $\gamma|_{P_1}$ and $\gamma|_{P_2}$ have the same flavor $g \in \{x, y, a\}$ and the same sign. Suppose the endpoints of $\gamma|_{P_1}$ differ by g^{m_1} and those of $\gamma|_{P_2}$ differ by g^{m_2}. Then

$$|\beta| \geq |g^{m_1+m_2}| \geq |m_1 + m_2|^{1/\alpha} \geq |m_1|^{1/\alpha}.$$

This completes the proof for $k = 2$.

Consider the case $k = 3$. We split into subcases based on the number of flavors that appear among the $\gamma|_{P_i}$:

(i) exactly one flavor,

(ii) exactly two flavors, or

(iii) exactly three flavors.

If the $\gamma|_{P_i}$ are unmanageable then in case (i) they all have the same sign, in case (ii) up to symmetry the flavors that appear are (x,a,a), (x,x,a) or (x,x,y), and in case (iii) up to symmetry the x-escape and the a-escape have the same sign. Case (i) is treated similarly to the case $k=2$ with all escapes having the same flavor and sign. In case (iii), let the endpoints of the escapes differ by a^ℓ, x^m and y^n. If p and q are such that $\ell = pL+q$ with $0 \le |q| < L$ then ℓ and p have the same sign so that, by Corollary 3.3.2,

$$|\beta| \ge |a^\ell x^m y^n| \ge |m+p|^{1/\alpha} - 5 = (|m|+|p|)^{1/\alpha} - 5 \ge |m|^{1/\alpha} - 5,$$

so we are done. For case (ii), the (x,x,a) and (x,x,y) subcases are treated by again applying Corollary 3.3.2 to obtain

$$|\beta| \ge |m_1+m_2+p|^{1/\alpha} + |p|^{1/\alpha} - 5 \ge |p|^{1/\alpha} - 5 \ge \frac{|\ell|^{1/\alpha}}{L^{1/\alpha}} - 5 = \frac{|\ell|^{1/\alpha}}{2} - 5$$

(where the last inequality is obtained as in the (x,a) case of $k=2$ above) and

$$|\beta| \ge |m_1+m_2|^{1/\alpha} + |n|^{1/\alpha} - 5 \ge |n|^{1/\alpha} - 5,$$

respectively. It remains to consider the (x,a,a) subcase for which Corollary 3.3.2 gives us the inequality

$$|\beta| \ge |m+p_1+p_2|^{1/\alpha} + |p_1+p_2|^{1/\alpha} - 5.$$

If $|m| \le |p_1+p_2|$ we are done, so assume that $|p_1+p_2| \le |m|$. Then, by the triangle inequality we have $|m+p_1+p_2| \ge |m| - |p_1+p_2| \ge 0$ and so $|m+p_1+p_2|^{1/\alpha} \ge \big(|m|-|p_1+p_2|\big)^{1/\alpha} \ge |m|^{1/\alpha} - |p_1+p_2|^{1/\alpha}$ by Fact 1.0.1. Thus,

$$|\beta| \ge |m|^{1/\alpha} - |p_1+p_2|^{1/\alpha} + |p_1+p_2|^{1/\alpha} - 5 = |m|^{1/\alpha} - 5,$$

as required.

Consider the case $k=4$. Assume that the $\gamma|_{P_i}$ are unmanageable. Renumber the segments so that P_1, P_2, P_3 and P_4 appear consecutively along S. Then either $P_1 \cup P_2$ is contained in a segment of length at most $\frac{1}{2}|S|$ of S or $P_3 \cup P_4$ is contained in such a segment and similarly for $P_2 \cup P_3$ and $P_1 \cup P_4$. We again renumber the segments so that the first possibility holds in each case. Apply Theorem 7.1.24 to the minimal geodesic segments respectively containing $P_1 \cup P_2$ and $P_2 \cup P_3$ to conclude that either $\big|\gamma|_{P_i}\big| \le \frac{M-1}{R} \cdot |\gamma|$ for some i or that $\gamma|_{P_1}$, $\gamma|_{P_2}$ and $\gamma|_{P_3}$ alternate between x and y flavors. Then, up to symmetry, we can assume that $\gamma|_{P_1}$ and $\gamma|_{P_3}$ are x-escapes and $\gamma|_{P_2}$ is a y-escape. If $\gamma|_{P_4}$ were an x-escape then, applying Corollary 3.3.2 we would obtain

$$|\beta| \ge |m_1+m_3+m_4|^{1/\alpha} + |n_2|^{1/\alpha} - 5 \ge |n_2|^{1/\alpha} - 5,$$

as needed. It remains to consider the cases where $\gamma|_{P_4}$ is

(i) a y-escape or

(ii) an a-escape.

In case (i), since the $\gamma|_{P_i}$ are unmanageable, either the escapes $\gamma|_{P_1}$ and $\gamma|_{P_3}$ have the same sign or the escapes $\gamma|_{P_2}$ and $\gamma|_{P_4}$ have the same sign. Up to symmetry we can assume the former so that Corollary 3.3.2 gives us

$$|\beta| \geq |m_1 + m_3|^{1/\alpha} + |n_2 + n_4|^{1/\alpha} - 5 \geq |m_1|^{1/\alpha} - 5,$$

as needed. In case (ii), since the $\gamma|_{P_i}$ are unmanageable, the escapes $\gamma|_{P_2}$ and $\gamma|_{P_4}$ must have the same sign, but then Corollary 3.3.2 gives us

$$|\beta| \geq |m_1 + m_3 + p_4|^{1/\alpha} + |n_2 + p_4|^{1/\alpha} - 5 \geq |n_2|^{1/\alpha} - 5,$$

as needed.

Consider the case $k = 5$. The $\gamma|_{P_i}$ are unmanageable so we need to show that $|\gamma|_{P_i}| \leq \frac{M-1}{R} \cdot |\gamma|$ for some i. Suppose this does not hold. Apply Theorem 7.1.24 to conclude that no three consecutive escapes among the $\gamma|_{P_i}$ are contained in a subsegment of γ of length at most $\frac{1}{2}|\gamma|$ of γ. Then any two consecutive segments among the $\gamma|_{P_i}$ are contained in a subsegment of γ of length at most $\frac{1}{2}|\gamma|$. Then Theorem 7.1.24 implies that, for any two consecutive escapes, one of them is an x-escape and the other one is a y-escape. Since k is odd this is a contradiction. □

Corollary 7.2.3 *For all $M \geq 24C + 1$ and $R > M - 1$ there exist $J > 0$ and $1 < K < 2$ such that the following holds.*

Let $\gamma\colon S \to X$ be a K-bi-Lipschitz cycle of length $|\gamma| \geq J$ such that $\gamma \cap H \neq \varnothing$ and every escape of γ from H has length at most $|\gamma|/2$. Let $\gamma|_{P_1}, \gamma|_{P_2}, \ldots, \gamma|_{P_k}$ be distinct escapes from H such that $k \leq 5$ and $\sum_{i=1}^{k} |\gamma|_{P_i}| \geq \frac{R-1}{R} \cdot |\gamma|$. Then a subset $T \subseteq \{\gamma|_{P_i}\}_{i=1}^{k}$ of $|T| \in \{2, 3, 4\}$ longest escapes is manageable and satisfies

$$\sum_{\gamma|_P \in T} |\gamma|_P| \geq \frac{R - M^3}{R} \cdot |\gamma| \quad \text{and} \quad \min_{\gamma|_P \in T} |\gamma|_P| \geq (M-1) \max\left\{ \frac{|\gamma|}{R}, |\gamma| - \sum_{\gamma|_P \in T} |\gamma|_P| \right\}.$$

Proof Let $R_0 > M^4(M-1)$, and let $R_i := R_0 M^{-i}$ for $i \leq 4$, so that for all i, M and R_i satisfy the hypotheses of Proposition 7.2.2. Let $K \in (1, 2)$ be close enough to 1 and let $J > 0$ be large enough that Proposition 7.2.2 holds with respect to M and R_i, for all $0 \leq i \leq 4$.

Let $\gamma\colon S \to X$ be a K-bi-Lipschitz cycle of length $|\gamma| \geq J$ such that $\gamma \cap H \neq \varnothing$ and every escape of γ from H has length at most $|\gamma|/2$. Let $\gamma|_{P_1}, \gamma|_{P_2}, \ldots, \gamma|_{P_k}$ be distinct escapes from H such that $k \leq 5$ and

$$\sum_{i=1}^{k} |\gamma|_{P_i}| \geq \frac{R_0 - 1}{R_0} \cdot |\gamma|.$$

If every escape of $T_0 = \{\gamma|_{P_i}\}_{i=1}^{k}$ has length greater than $\frac{M-1}{R_0} \cdot |\gamma|$ then applying Proposition 7.2.2 with $R = R_0$ the escapes of T_0 must be manageable. Moreover, in this case we would have

$$\min_{\gamma|_P \in T_0} |\gamma|_P| > (M-1) \cdot \left(|\gamma| - \frac{R_0 - 1}{R_0} \cdot |\gamma| \right) \geq (M-1) \cdot \left(|\gamma| - \sum_{\gamma|_P \in T_0} |\gamma|_P| \right)$$

so we would have the required conclusion with $R = R_0$ and $T = T_0$. Hence we may assume that some escape of T_0 has length at most $\frac{M-1}{R_0} \cdot |\gamma|$.

Let T_1 be obtained from T_0 by removing a shortest escape. We have

$$\sum_{\gamma|_P \in T_1} |\gamma|_P| \geq \left(\frac{R_0 - 1}{R_0} - \frac{M-1}{R_0}\right) \cdot |\gamma| = \frac{R_0 - M}{R_0} \cdot |\gamma| = \frac{R_1 - 1}{R_1} \cdot |\gamma|$$

so, applying Proposition 7.2.2 with $R = R_1$, we can conclude that $2 \leq |T_1| = |T_0| - 1 \leq 4$. Moreover, if every escape of T_1 had length greater than $\frac{M-1}{R_1} \cdot |\gamma|$ then T_1 would be manageable and, as argued above, we would have $\min_{\gamma|_P \in T_1} |\gamma|_P| \geq (M-1) \cdot \big(|\gamma| - \sum_{\gamma|_P \in T_1} |\gamma|_P|\big)$. Thus we would have the required conclusion with $R = R_0$ and $T = T_1$. Hence we may assume that some escape of T_1 has length at most $\frac{M-1}{R_1} \cdot |\gamma|$.

Continuing in this way, we obtain T_2 from T_1 by removing a shortest escape and if $T = T_2$ is not as needed then we again remove a shortest escape to obtain T_3. By the same arguments, we have

$$\sum_{\gamma|_P \in T_3} |\gamma|_P| \geq \frac{R_3 - 1}{R_3} \cdot |\gamma| = \frac{R_0 - M^3}{R_0} \cdot |\gamma|$$

and $|T_3| = 2$. If the shorter escape of T_3 had length at most $\frac{M-1}{R_3} \cdot |\gamma|$ then the remaining escape would have length at least $\frac{R_4-1}{R_4} \cdot |\gamma|$ which, by Proposition 7.2.2 with $R = R_4$, is a contradiction. Thus the escapes of T_3 must all have length greater than $\frac{M-1}{R_3} \cdot |\gamma| \geq (M-1) \cdot \big(|\gamma| - \sum_{\gamma|_P \in T_3} |\gamma|_P|\big)$. Applying Proposition 7.2.2 with $R = R_3$ the escapes of T_3 must also be manageable, so we are done with $R = R_0$ and $T = T_3$. □

Corollary 7.2.4 *For all $R > 0$ there exist $J > 0$ and $1 < K < 2$ such that if $\gamma : S \to X$ is a K-bi-Lipschitz cycle of length $|\gamma| \geq J$ with $\gamma \cap H \neq \varnothing$ and all escapes of γ from H have length at most $|\gamma|/2$ then there is a manageable set T of longest escapes of γ from H that satisfies*

$$\sum_{\gamma|_P \in T} |\gamma|_P| \geq \frac{R-1}{R} \cdot |\gamma| \quad \text{and} \quad \min_{\gamma|_P \in T} |\gamma|_P| \geq R \cdot \max\left\{\frac{|\gamma|}{J}, |\gamma| - \sum_{\gamma|_P \in T} |\gamma|_P|\right\}.$$

Proof Let $M \geq R + 1$ be large enough that we can apply Corollary 7.2.3 and let R' be large enough that

$$\frac{R' - M^3}{R'} \geq \frac{R-1}{R}$$

and large enough that we can apply Corollary 7.2.3 with M and R'. Assume that $J \geq R'$ is large enough and $K > 1$ is close enough to 1 that the second paragraph of Corollary 7.2.3 holds.

Let $\gamma : S \to X$ be a K-bi-Lipschitz cycle of length $|\gamma| \geq J$. Let v be a vertex of S that maps to H. Let $\bar{v}$ be the antipode of v in S. Since the relators of G_L all have even length so then does S and so $\bar{v}$ is a vertex of S. Let Q'_1 and Q'_2 be the two geodesic segments of S between v and $\bar{v}$. If $\bar{v} \in P$ for some escape $\gamma|_P$ of A then, for each i, let $Q_i = Q'_i \setminus P^\circ$ where P° is the interior of P; otherwise, for each i, let $Q_i = Q'_i$. If J is large enough and K is close enough to 1, Theorem 7.1.24 implies that at most four

escapes occupy an $\frac{R'-1}{R'}$ proportion of $\gamma|_{Q_1}$ and $\gamma|_{Q_2}$. Thus a set T' of five longest escapes of A have length at least $\frac{R'-1}{R'}\cdot|\gamma|$. Applying Corollary 7.2.3 to T' we obtain a manageable set T of longest escapes of A such that

$$\sum_{\gamma|_P\in T}\big|\gamma|_P\big| \geq \frac{R'-M^3}{R'}\cdot|\gamma| \geq \frac{R-1}{R}\cdot|\gamma|,$$

$$\min_{\gamma|_P\in T}\big|\gamma|_P\big| \geq (M-1)\cdot\Big(|\gamma|-\sum_{\gamma|_P\in T}\big|\gamma|_P\big|\Big) \geq R\cdot\Big(|\gamma|-\sum_{\gamma|_P\in T}\big|\gamma|_P\big|\Big),$$

$$\min_{\gamma|_P\in T}\big|\gamma|_P\big| \geq (M-1)\cdot\frac{|\gamma|}{R'} \geq R\cdot\frac{|\gamma|}{J}.$$

□

7.3 Noncentral "regions"

Recall compound escapes from Definition 5.2.2. Fix R and let J and K be the corresponding constants from Theorem 7.1.24, and suppose that $\gamma\colon S\to X$ is a K-bi-Lipschitz cycle. Suppose that $\gamma\cap gH\neq\varnothing$ and suppose that γ has a designated compound escape $\gamma|_P$ from gH that has length greater than $|\gamma|/2$. Call $\gamma|_P$ the "central escape". We wish to describe the configuration of noncentral escapes of γ from gH. This is easier than in the previous subsection because $|\gamma|-\big|\gamma|_P\big|<|\gamma|/2$, so all of the noncentral escapes are contained in a common K-bi-Lipschitz path. This fact eliminates most of the cases considered in Figure 7; only the analogues of cases (A) and (C) can occur. Specifically, according to Theorem 7.1.24, there are only three possibilities:

(1) $|\gamma|-\big|\gamma|_P\big|<J$. In this case say *the noncentral part of γ at gH is very small.*

(2) The noncentral part is not very small, the central escape is a compound a-escape, and there are a noncentral x-escape and noncentral y-escape whose lengths sum to at least $\frac{R-1}{R}\big(|\gamma|-\big|\gamma|_P\big|\big)$. In this case say that the noncentral part of γ is *branching at gH*.

(3) The noncentral part is not very small, there is exactly one escape of length at least $\frac{R-1}{R}\big(|\gamma|-\big|\gamma|_P\big|\big)$, and it is of the same flavor as the central escape.

Instances of the third case can be chained together to form super-regions that we now describe:

Definition 7.3.1 Let $R>2$. Let γ be a g-escape for some $g\in\{a,x,y\}$. The *R-enfilade decomposition* of γ is the factoring of γ as

$$\gamma_0:=\gamma=\epsilon_0+\alpha_1+\epsilon_1+\cdots+\alpha_n+\epsilon_n+\gamma_n'+\epsilon_n^{-1}+\beta_n+\cdots+\epsilon_1^{-1}+\beta_1+\epsilon_0^{-1}$$

such that the following conditions hold (see Figure 8):

- Each α_i or β_i is a path whose endpoints differ by an element of H.
- For each i, we have $\epsilon_i\in\{s^{\pm1},t^{\pm1}\}$ and γ_i' is defined to be the path such that $\gamma_i=\epsilon_i+\gamma_i'+\epsilon_i^{-1}$.
- For each $i\in\{0,1,\dots,n-1\}$, we have $|\gamma_{i+1}|\geq\frac{R-1}{R}\cdot|\gamma_i'|$.

- For each $i \in \{0, 1, 2, \dots, n\}$,
$$\gamma_i = \epsilon_i + \alpha_{i+1} + \cdots + \epsilon_n + \gamma'_n + \epsilon_n^{-1} + \cdots + \beta_{i+1} + \epsilon_i^{-1}$$
is a g_i-escape, for some $g_i \in \{a, x, y\}$, whose endpoints differ by $g_i^{m_i}$ for some $m_i \in \mathbb{Z}$.
- n is maximal such that γ admits a decomposition satisfying these conditions.

Call the subsegment γ'_n the *end* of the enfilade.

Note that since $R > 2$ there is at most one candidate for γ_{i+1} for each i, so the decomposition is unique.

Lemma 7.3.2 *If $R > 2$ and $1 < K < 2$ and γ is a K-bi-Lipschitz g-escape for some $g \in \{a, x, y\}$ then in the R-enfilade decomposition of γ all of the exponents m_i have the same sign.*

Proof If m_{i-1} and m_i have different signs for some i then
$$|\beta_i| \geq |g_i^{m_{i-1}}| - |\alpha_i| + |g_i^{m_i}|,$$
but $\gamma'_{i-1} = \alpha_i + \gamma_i + \beta_i$, so
$$|\gamma'_{i-1}| - |\gamma_i| = |\alpha_i| + |\beta_i| \geq |g_i^{m_{i-1}}| + |g_i^{m_i}| \geq \frac{1}{K}(|\gamma'_{i-1}| + |\gamma_i|) \implies |\gamma_i| \leq \frac{K-1}{K+1}|\gamma'_{i-1}|.$$
But $|\gamma_i| > \frac{R-1}{R}|\gamma'_{i-1}|$, so $\frac{1}{2} < \frac{R-1}{R} < \frac{K-1}{K+1} < \frac{1}{3}$, and we have a contradiction. □

Lemma 7.3.3 *Assume that γ begins at 1 and that we have an enfilade decomposition of γ with notation as above. Let $h_i \in H$ be the difference between the endpoints of α_i. For any $k \in \mathbb{Z}$, we have that $g_0^k.\left(\epsilon_0 + \sum_{i=1}^n \alpha_i + \epsilon_i\right)$ is a path from g_0^k to $\left(\epsilon_0 \prod_{i=1}^n h_i \epsilon_i\right) g_{n+1}^k$.*

Proof This follows by induction on n using that facts that $g_i = \epsilon_{i-1}^{-1} g_{i-1} \epsilon_{i-1}$ and g_i and h_i commute. □

Remark 7.3.4 Let $R > 2$, let $2 > K > 1$, let γ be a K-bi-Lipschitz g-escape for some $g \in \{a, x, y\}$ and let
$$\gamma = \epsilon_0 + \alpha_1 + \cdots + \epsilon_n + \gamma'_n + \epsilon_n^{-1} + \cdots + \beta_1 + \epsilon_0^{-1}$$
be the R-enfilade decomposition of γ. By Theorem 7.1.24, there exists a $J > 1$ such that if K is close enough to 1 then either $|\gamma'_n| \leq J$ or the endpoints of γ'_n differ by a power of a and γ'_n has an x-escape and a y-escape that collectively account for an at least $\frac{R-1}{R}$ fraction of its length.

Proposition 7.3.5 (enfilades are short) *Let $R > 3$. There exists a $2 > K > 1$ and $J > 2$ such that if γ is a K-bi-Lipschitz g-escape for some $g \in \{a, x, y\}$, $|\gamma| \geq J$ and*
$$\gamma = \epsilon_0 + \alpha_1 + \cdots + \epsilon_n + \gamma'_n + \epsilon_n^{-1} + \cdots + \beta_1 + \epsilon_0^{-1}$$
is the R-enfilade decomposition of γ, then $|\gamma'_n| \geq \frac{R-3}{R} \cdot |\gamma|$.

Proof Since $R > 3$,
$$\frac{R-1}{R} \cdot \frac{R-2}{R} > \frac{R-3}{R}.$$

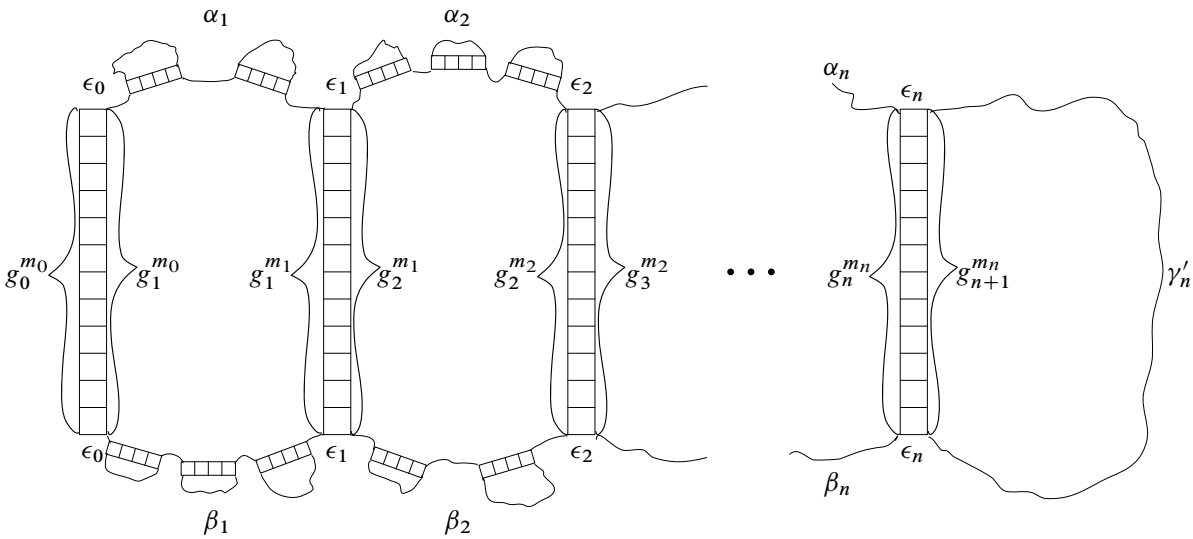


Figure 8: Enfilade of length n.

Let $R_0 > 1$ be large enough to satisfy $R_0/(R_0+1) > 2/R$ and

$$\frac{R-1}{R}\cdot\frac{R_0}{R_0+1} > \frac{R-2}{R}.$$

Let $J > 2$ and $2 > K > 1$ be such that the conclusion of Corollary 7.1.8 holds with $(R, J, K) = (R_0, J, K)$ and so that all of the following hold:

$$\frac{2}{R}+\frac{4}{J} < \frac{R_0}{R_0+1}, \qquad \frac{R-1}{R}\cdot\left(\frac{R_0}{R_0+1}-\frac{2}{J}\right) > \frac{R-2}{R}, \qquad \frac{R-1}{R}\cdot\frac{J-2}{J} \geq \frac{R-2}{R},$$
$$\frac{R-1}{R}\cdot\frac{R_0}{R_0+1}-\frac{4}{J} \geq \frac{R-2}{R}, \qquad \frac{R-1}{R}\cdot\frac{R-2}{R}\cdot\frac{J-2}{J} > \frac{R-3}{R}.$$

Let γ be a K-bi-Lipschitz g-escape for some $g \in \{a, x, y\}$, let $|\gamma| \geq J$ and consider the R-enfilade decomposition of γ with the same notation as in Definition 7.3.1.

First assume that $g_0 = a$, and notice that then $g_i = a$ whenever i is even. We have $\alpha_i + g_i^{m_i} + \beta_i = g_i^{m_{i-1}}$, for each $i \in \{1, 2, \dots, n\}$. By commutativity of H, the endpoints of $\alpha_i + \beta_i$ differ by $g_i^{m_{i-1}-m_i}$. Define $\delta_{2i} := \epsilon_{2i-2} + \alpha_{2i-1} + \epsilon_{2i-1} + \alpha_{2i} + \beta_{2i} + \epsilon_{2i-1}^{-1} + \beta_{2i-1} + \epsilon_{2i-2}^{-1}$. By the previous observation and an argument similar to Lemma 7.3.3, the endpoints of δ_{2i} differ by $a^{m_{2i-2}-m_{2i}}$, so the endpoints of $\delta_{2i} + \gamma_{2i}$ differ by $a^{m_{2i-2}-m_{2i}} = a^{m_{2i-2}}$, which is the same as the difference between the endpoints of γ_{2i-2}. Moreover, we have

$$\begin{aligned}|\delta_{2i} + \gamma_{2i}| &= |\epsilon_{2i-2} + \alpha_{2i-1} + \epsilon_{2i-1} + \alpha_{2i} + \beta_{2i} + \epsilon_{2i-1}^{-1} + \beta_{2i-1} + \epsilon_{2i-2}^{-1}| + |\gamma_{2i}| \\ &= |\epsilon_{2i-2} + \alpha_{2i-1} + \epsilon_{2i-1} + \alpha_{2i} + \gamma_{2i} + \beta_{2i} + \epsilon_{2i-1}^{-1} + \beta_{2i-1} + \epsilon_{2i-2}^{-1}| = |\gamma_{2i-2}|.\end{aligned}$$

By induction, the path $\delta_2 + \delta_4 + \cdots + \delta_{2i} + \gamma_{2i}$ has the same length and the same endpoints as γ. Let $k \in \mathbb{Z}$ be defined by $n - 1 \leq 2k \leq n$ and consider the path $\delta_2 + \delta_4 + \cdots + \delta_{2k} + \gamma_{2k}$. Since γ is K-bi-Lipschitz,

we have $J \leq |\gamma| = |\delta_2 + \delta_4 + \cdots + \delta_{2k} + \gamma_{2k}| = |\gamma| \leq K|a^{m_0}|$. Also the endpoints of δ_{2i} differ by $a^{m_{2i-2}-m_{2i}}$, for each i, and the endpoints of γ_{2k} differ by $a^{m_{2k}}$, so, by Corollary 7.1.8, either

$$|\delta_{2i}| > \frac{R_0}{R_0+1} \cdot |\gamma|,$$

for some $i \in \{1, 2, \ldots, k\}$, or

$$|\gamma_{2k}| > \frac{R_0}{R_0+1} \cdot |\gamma|.$$

But,

$$\begin{aligned}
|\delta_{2i}| &= |\epsilon_{2i-2} + \alpha_{2i-1} + \epsilon_{2i-1} + \alpha_{2i} + \beta_{2i} + \epsilon_{2i-1}^{-1} + \beta_{2i-1} + \epsilon_{2i-2}^{-1}| \\
&= |\epsilon_{2i-2} + \alpha_{2i-1} + \beta_{2i-1} + \epsilon_{2i-2}^{-1}| + |\epsilon_{2i-1} + \alpha_{2i} + \beta_{2i} + \epsilon_{2i-1}^{-1}| \\
&= \big(2 + |\gamma'_{2i-2}| - |\gamma_{2i-1}|\big) + \big(2 + |\gamma'_{2i-1}| - |\gamma_{2i}|\big) \\
&\leq |\gamma'_{2i-2}| - \frac{R-1}{R} \cdot |\gamma'_{2i-2}| + |\gamma'_{2i-1}| - \frac{R-1}{R} \cdot |\gamma'_{2i-1}| + 4 \\
&= \frac{1}{R} \cdot |\gamma'_{2i-2}| + \frac{1}{R} \cdot |\gamma'_{2i-1}| + 4 \\
&\leq \frac{2}{R} \cdot |\gamma| + 4 \\
&= \Big(\frac{2}{R} + \frac{4}{|\gamma|}\Big) \cdot |\gamma| \\
&\leq \Big(\frac{2}{R} + \frac{4}{J}\Big) \cdot |\gamma| \\
&< \frac{R_0}{R_0+1} \cdot |\gamma|,
\end{aligned}$$

where the first inequality follows from the definition of R-enfilade decomposition. Thus,

$$\frac{R_0}{R_0+1} \cdot |\gamma| < |\gamma_{2k}| = |\gamma'_{2k}| + 2 \leq \frac{R}{R-1} \cdot |\gamma_n| + 2.$$

We conclude that

$$\begin{aligned}
|\gamma'_n| = |\gamma_n| - 2 &> \frac{R-1}{R} \cdot \Big(\frac{R_0}{R_0+1} \cdot |\gamma| - 2\Big) - 2 \\
&= \frac{R-1}{R} \cdot \frac{R_0}{R_0+1} \cdot |\gamma| - 2 \cdot \frac{2R-1}{R} \\
&> \frac{R-1}{R} \cdot \frac{R_0}{R_0+1} \cdot |\gamma| - 4 \\
&= \Big(\frac{R-1}{R} \cdot \frac{R_0}{R_0+1} - \frac{4}{|\gamma|}\Big)|\gamma| \\
&\geq \Big(\frac{R-1}{R} \cdot \frac{R_0}{R_0+1} - \frac{4}{J}\Big)|\gamma| \\
&\geq \frac{R-2}{R} \cdot |\gamma|.
\end{aligned}$$

If $g_0 \neq a$ then $g_0 \in \{x, y\}$. If $n = 0$ we are done, so suppose not. Since $g_0 \neq a$, $g_1 = a$ and $g_i = a$ for all odd i. Apply the previous argument to γ_1 to get

$$|\gamma_n'| \geq \frac{R-2}{R} \cdot |\gamma_1| \geq \frac{R-2}{R} \cdot \frac{R-1}{R} \cdot |\gamma_0'| = \frac{R-2}{R} \cdot \frac{R-1}{R} \cdot (|\gamma| - 2)$$
$$= \frac{R-2}{R} \cdot \frac{R-1}{R} \cdot \frac{J-2}{J} \cdot |\gamma| \geq \frac{R-3}{R} \cdot |\gamma|. \qquad \square$$

8 Constructing fillings

We will show that for all $\lambda \in (0, 1)$ we have that for all sufficiently large R and Λ there exist $1 < K < 2$ and $A, J, M > 0$ such that every K-bi-Lipschitz loop γ can be filled by a diagram of mesh at most $\lambda|\gamma| + M$ of area at most A. We do so inductively. An overview of the inductive argument is given in Section 8.1. In Section 8.2 we establish basic filling results for the approximate polygons that will occur as vertex regions. In Section 8.3 we handle the base case: identifying and filling a central region. In Section 8.4 we cover three possible types of induction steps: small enough parts, enfilades, and branching regions. In many of these steps we will need to make assumptions on the relationships between the various constants. These are boxed for emphasis. In Theorem 9.0.1 we put all of the pieces of the proof together, including observing that all of the boxed relations can be satisfied simultaneously, so that we can actually, quantitatively, perform the induction described in Section 8.1.

8.1 Overview: constructing a filling inductively

In Lemma 8.3.1 we identify a coset gH from which γ has a manageable collection of at most 4 long escapes. Call these the 0^{th} *level escapes*. We resect those escapes and replace each with its trace in gH. We then subdivide each of the traces into segments of bounded exponent and replace each segment by a geodesic with the same endpoints. We find a filling diagram for the resulting loop γ' of controlled mesh and area. Call the filling diagram for γ' the 0^{th} *shell*.

It remains to take each of the 0^{th} level escapes, complete it to a loop using the same concatenation of geodesic subsegments as used to build γ', and construct a filling diagram of that loop.

For the induction, suppose we have an n^{th} level escape $\gamma|_P$ of γ and an n^{th} shell that includes a filling diagram with a boundary segment that is a subdivision of the trace of $\gamma|_P$. If $\gamma|_P$ is "small enough" we can fill it with a single 2-cell. This is quantified explicitly in (8.1.3) below. We will see in Section 8.4 that very small parts are small enough. If $\gamma|_P$ is not small enough then we look at its enfilade decomposition. Let $\gamma|_{P'}$ be the end of the enfilade, and let δ be the toral segment between the endpoints of $\gamma|_{P'}$. In Section 8.4 we describe how to fill the enfilade up to δ, which also yields a subdivision of δ. Then according to Section 7.3 there are two possibilities: either $\gamma|_{P'}$ is short enough to cap off with a single 2-cell, or it is branching. If it is branching then $\gamma|_{P'}$ has two long escapes and their traces make an approximate triangle with δ. We fill that triangle. The fillings of the enfilade part and the short/branching

part together make up a component of the $(n+1)^{\text{st}}$ shell. In the case of branching, the two long escapes are added to the list of $(n+1)^{\text{st}}$ level escapes.

We finish the argument by showing that after finitely many steps, independent of γ, every remaining branch is small enough that it can be capped off with a single 2-cell, so that no escapes get added to a deeper level of the induction. The diagram filling γ is the union of the shells, identified along common segments. The induction hypothesis is:

(8.1.1) For an n^{th} level escape that is not small enough to cap off, the n^{th} shell provides a subdivision of its trace into at most $2(\Lambda+1)$-many subsegments of exponent at most $E|\gamma|^\alpha \cdot (3/L)^n$, where E is the constant of Lemma 8.3.1, which is independent of γ.

Lemma 8.3.1 establishes (8.1.1) for $n = 0$. The inductive step is proved in Section 8.4. Let us suppose that (8.1.1) is true for all n and see how to finish the argument.

(8.1.2) Assume $R > 12C(\Lambda+1)/\lambda$.

Suppose γ is a K-bi-Lipschitz loop and gH is a coset such that:

(8.1.3) γ has a central escape, a compound escape from gH of length at least $|\gamma|/2$, whose trace is subdivided into at most $2(\Lambda+1)$-many subsegments of exponent bounded by $L(|\gamma|/R)^\alpha$.

Choose geodesic segments $\delta_1, \dots, \delta_m$, $m \le 2(\Lambda+1)$, with the same endpoints as the subdivision segments. Let $\gamma|_P$ be the complement of the central escape. It is K-bi-Lipschitz, with the same endpoints as $\delta_1 + \cdots + \delta_m$, so $|\gamma|_P| \le K \sum_{i=1}^m |\delta_i|$. Since the exponents of the δ_i are bounded, we apply Theorem 3.2.9 to get $|\delta_i| < C\left(L(|\gamma|/R)^\alpha\right)^{1/\alpha} = 2C|\gamma|/R$. Thus,

$$\begin{aligned}
|\gamma|_P| + \sum_{i=1}^m |\delta_i| &\le (K+1) \sum_{i=1}^m |\delta_i| \\
&< 3 \cdot 2(\Lambda+1) \cdot \frac{2C|\gamma|}{R} \\
&= 12C(\Lambda+1) \cdot |\gamma| \cdot \frac{1}{R} \\
&< 12C(\Lambda+1) \cdot |\gamma| \cdot \frac{\lambda}{12C(\Lambda+1)} = \lambda|\gamma|.
\end{aligned}$$

Thus, if (8.1.3) is satisfied then $\gamma|_P + \bar{\delta}_m + \cdots + \bar{\delta}_1$ is a loop of length less than $\lambda|\gamma|$, so it is small enough to cap off with a single 2-cell satisfying the desired mesh bound.

The induction hypothesis (8.1.1) says that an n^{th} level escape has trace subdivided into segments of exponent at most $E|\gamma|^\alpha \cdot (3/L)^n$, so it is guaranteed to be small enough to cap off with a single 2-cell once we have

(8.1.4) $$E|\gamma|^\alpha \cdot (3/L)^n \le L\left(\frac{|\gamma|}{R}\right)^\alpha \iff n \ge \log_{L/3}(ER^\alpha/L).$$

Since E and R are independent of γ, (8.1.4) gives a uniform upper bound on the number of shells we need to fill before we can cap off all remaining branches with a single 2-cell apiece.

Suppose that we have area bounds $\mathscr{C}$, $\mathscr{E}$, and $\mathscr{B}$ for our fillings of the central region, enfilades, and branching regions, respectively, by diagrams of controlled mesh. The worst-case scenario is that the central region has four long escapes, each belonging to an enfilade terminating in a branching region, and for each of these branching regions the two long escapes belong to enfilades terminating in a branching region, etc, carried out to the n^{th} shell, where n is the least integer satisfying (8.1.4), at which point all remaining regions are small enough and can be capped off with one additional 2-cell. The area of the resulting diagram is bounded above by

$$\mathscr{C} + 4(\mathscr{E} + \mathscr{B})(2^n - 1) + 2^{n+2}. \tag{8.1.5}$$

Thus, if we control the mesh and area of the filling diagrams used in the induction steps then we can assemble a uniformly bounded number of them to get a filling diagram for γ of controlled mesh and uniformly bounded area.

8.2 Approximate polygons and fillings

The construction of fillings for the different pieces of the argument can be unified in the framework of filling approximate polygons. An approximate polygon, for us, will mean a cycle, in H, of segments of a-lines, x-lines, and y-lines such that the terminal endpoint of one is close to the initial endpoint of the next. We construct fillings of certain approximate polygons. These will be the building blocks for fillings of bi-Lipschitz loops constructed in the next two sections.

Approximate polygons are close to true polygons We will quantify approximate polygons as *D-approximate* to mean that the distance between consecutive segments is as most D. A *true* polygon is a 0-approximate polygon.

Definition 8.2.1 Let a *D-approximate triangle* mean a collection of elements g_0, g_1, and g_2 of $\langle a, x, y\rangle$ and numbers $m_0, m_1, m_2 \in \mathbb{Z}$ such that the distances $d(g_0x^{m_0}, g_1)$, $d(g_1y^{m_1}, g_2)$, and $d(g_2a^{m_2}, g_0)$ are at most D.

Lemma 8.2.2 (approximate triangles) *For every such D-approximate triangle there exists a true triangle described by $g_0', g_1', g_2' \in \langle a, x, y\rangle$ and $m_0', m_1', m_2' \in \mathbb{Z}$ such that the distance between every corresponding pair of points is at most $2D + L$.*

Proof By Lemma 3.4.2, there exists $p \in \mathbb{Z}$ with $|p| \le L/2$ such that $g_0\langle x\rangle \cap g_1a^p\langle y\rangle \neq \varnothing$. Choose such a p, and let $g_1' := g_0\langle x\rangle \cap g_1a^p\langle y\rangle$. By Lemma 3.4.1, we can set $g_0' := g_0\langle x\rangle \cap g_2\langle a\rangle$ and $g_2' := g_2\langle a\rangle \cap g_1a^p\langle y\rangle = g_2\langle a\rangle \cap g_1'a^p\langle y\rangle$. Define m_i' so that $g_0'x^{m_0'} = g_1'$, $g_1'y^{m_1'} = g_2'$, and $g_2'a^{m_2'} = g_0'$. Note that $m_0' = m_1'$ and $m_2' = -Lm_0'$. See Figure 9.

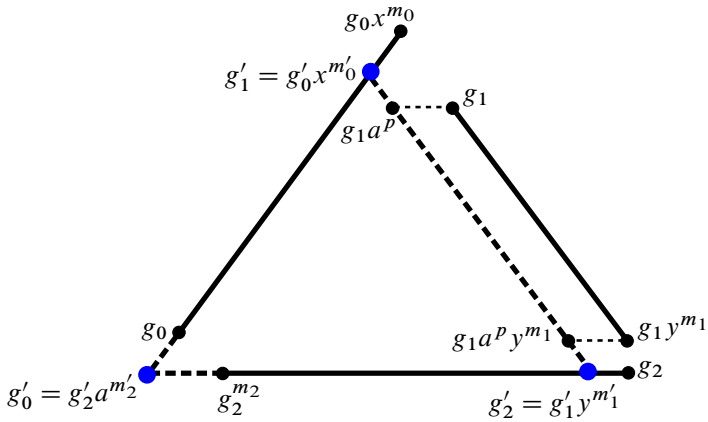


Figure 9: True triangle close to an approximate triangle.

It remains to bound the distance between corresponding points. We verify what we claim to be the worst case, and leave the remaining verifications to the reader:

$$\begin{aligned}
d(g_2, g_2') &\le d(g_2, g_1y^{m_1}) + d(g_1y^{m_1}, g_2') \\
&\le D + d(g_1y^{m_1}, g_2') \\
&= D + d(g_1y^{m_1}, g_1'y^{m_1'}) \\
&\le D + d(g_1y^{m_1}, g_1a^py^{m_1}) + d(g_1a^py^{m_1}, g_1'y^{m_1'}) \\
&= D + d(g_1y^{m_1}, g_1a^py^{m_1}) + d(g_1a^py^{m_1}, g_2\langle a\rangle) \\
&\le D + 2d(g_1y^{m_1}, g_1a^py^{m_1}) + d(g_1y^{m_1}, g_2\langle a\rangle) \\
&\le D + L + d(g_1y^{m_1}, g_2\langle a\rangle) \\
&\le D + L + d(g_1y^{m_1}, g_2) \le 2D + L.
\end{aligned}$$

□

Lemma 8.2.3 (approximate diamonds) *Let a **D-approximate diamond** mean a collection of elements g_1, g_2, h_1, $h_2 \in \langle a, x, y\rangle$ and numbers m_1, m_2, n_1, $n_2 \in \mathbb{Z}$ such that $d(g_1x^{m_1}, h_1)$, $d(h_1y^{n_1}, g_2)$, $d(g_2x^{m_2}, h_2)$, and $d(h_2y^{n_2}, g_1)$ are all at most D.*

For every such D-approximate diamond there exists a true diamond described by $g_1', g_2', h_1', h_2' \in \langle a, x, y\rangle$ and $m_1', m_2', n_1', n_2' \in \mathbb{Z}$ such that the distance between every corresponding pair of points is at most $D + 3L/2$.

Proof By Lemma 3.4.2, there exist numbers p_1, p_2, and p_3 with all $|p_i| \le L/2$ such that

$$g_1\langle x\rangle \cap h_1a^{p_1}\langle y\rangle \neq \varnothing, \quad g_1\langle x\rangle \cap h_2a^{p_2}\langle y\rangle \neq \varnothing, \quad g_2a^{p_3}\langle x\rangle \cap h_1a^{p_2}\langle y\rangle \neq \varnothing.$$

Choose such p_i. Define

- $h_1' := g_1\langle x\rangle \cap h_1a^{p_1}\langle y\rangle$,
- $g_2' := g_2a^{p_3}\langle x\rangle \cap h_1'\langle y\rangle$,

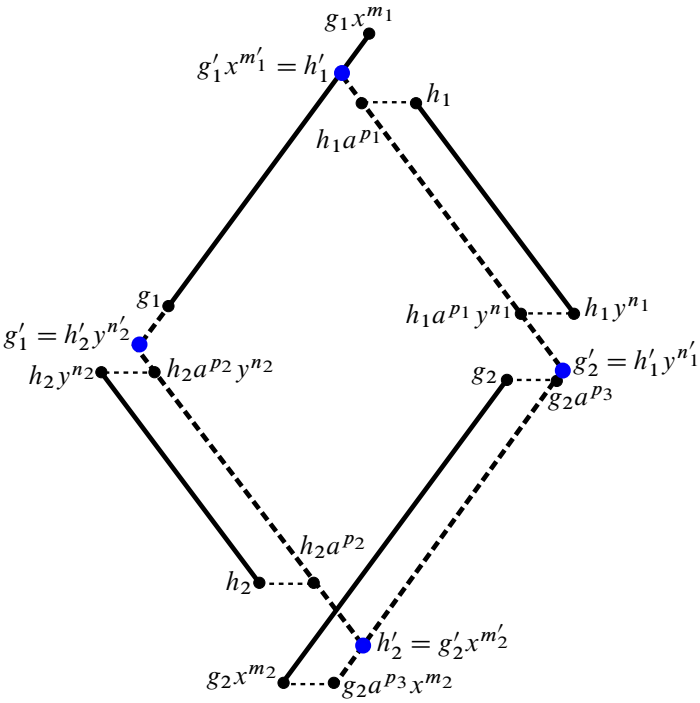


Figure 10: True diamond close to an approximate diamond.

- n_1' so that $h_1'y^{n_1'} = g_2'$,
- $h_2' := g_2'\langle x\rangle \cap h_2a^{p_2}\langle y\rangle$,
- m_2' so that $h_2' = g_2'x^{m_2'}$,
- $g_1' := h_2'\langle y\rangle \cap g_1\langle x\rangle = h_2a^{p_2}\langle y\rangle \cap g_1\langle x\rangle$,
- n_2' so that $g_1' = h_2'y^{n_2'}$,
- m_1' so that $g_1'x^{m_1'} = h_2'$.

Finally, note that $m_2' = -m_1'$ and $n_2' = -n_1'$. See Figure 10.

It remains to bound the distance between corresponding points. We will check what we claim is the worst case, and leave the other, similar, verifications to the reader:

$$\begin{aligned}
d(h_2, h_2') &\le d(h_2, h_2a^{p_2}) + d(h_2a^{p_2}, h_2')\\
&= d(h_2, h_2a^{p_2}) + d(h_2a^{p_2}, g_2'x^{m_2'})\\
&\le d(h_2, h_2a^{p_2}) + d(h_2a^{p_2}, g_2'\langle x\rangle)\\
&\le d(h_2, h_2a^{p_2}) + d(h_2a^{p_2}, g_2\langle x\rangle) + d_{\mathrm{Haus}}(g_2\langle x\rangle, g_2'\langle x\rangle)\\
&\le L + d(h_2a^{p_2}, g_2\langle x\rangle)\\
&\le L + d(h_2a^{p_2}, g_2x^{m_2})\\
&\le L + d(h_2a^{p_2}, h_2) + d(h_2, g_2x^{m_2}) \le L + L/2 + D.
\end{aligned}$$

□

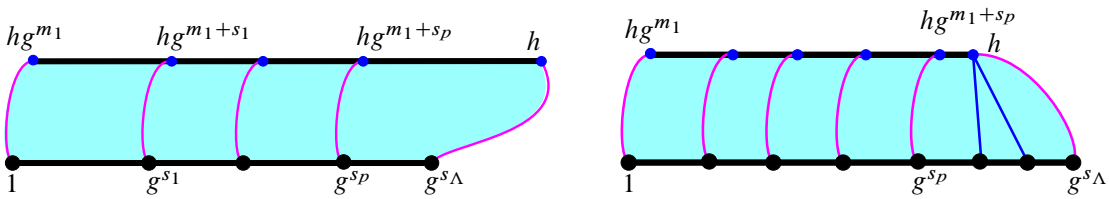


Figure 11: Two cases for filling a bigon.

Filling approximate polygons Let a *D-approximate bigon* mean a collection of elements $g_0, g_1 \in \langle a, x, y\rangle$ and $g \in \{a, x, y\}$, and numbers $m_0, m_1 \in \mathbb{Z}$ such that $d(g_i g^{m_i}, g_{i+1}) \leq D$ for all $0 \leq i \leq 1$ with indices mod 2.

For an approximate polygon, a collection of "corner paths" refers to a collection of paths connecting successive endpoints of sides.

Lemma 8.2.4 (filling bigons) *There exist constants M_0, M_1, and M_2 such that for all positive Λ, E, and D, given an approximate bigon, a collection of corner paths of length at most D, and a subdivision of one side of the bigon into Λ-many subsegments, each of exponent at most E, the following hold. There exists a subdivision of the opposite side of the bigon into Λ-many subsegments of exponent bounded by $E + LD^\alpha$. Furthermore, if we construct a loop γ out of the bigon with corner paths by replacing each of the subsegments of the sides by some geodesic, then there is a diagram for γ with mesh at most $M_0 + M_1 D + M_2 E^{1/\alpha}$ and area at most Λ. In fact, it suffices to take $M_0 := 0$, $M_1 := 2(2C + 1)$, and $M_2 := 2C$.*

Proof Assume that we have been given a subdivision of the g_0-side of the bigon. We refer to the g_0-side of the bigon as the "bottom" and the g_1-side as the "top". Up to isometry we may assume $g_0 = 1$ and $m_0 \geq 0$. Set $h := g_1$.

Let $s_0 := 0$ and let $s_i \geq 0$ be the sum of the first i exponents of the bottom, so that the bottom consists of g-segments $[g^{s_{i-1}}, g^{s_i}]$ for $0 < i \leq \Lambda$. Note that $s_\Lambda = m_0$. Let δ_0 be the given corner path from 1 to hg^{m_1}, and let δ_Λ be the corner path from g^{s_Λ} to h. Define p to be the lesser of $\Lambda - 1$ and the greatest index such that s_p is contained in the closed interval defined by 0 and m_1 (ie $[m_1, 0]$ or $[0, m_1]$ depending on the sign of m_1).

There are two cases according to whether or not $m_0 > -m_1$. See Figure 11.

First suppose, as in Figure 11, left, that $m_0 \leq -m_1$, so that $m_1 \leq 0$ and $p = \Lambda - 1$. For each $0 < i \leq p$, define $\delta_i = g^{s_i}.\delta_0$, so that δ_i is a path of length at most D between g^{s_i} and $g^{s_i} hg^{m_1} = hg^{m_1+s_i}$.

For each $i < p$ define a 2-cell with boundary consisting of a geodesic from $g^{s_{i-1}}$ to g^{s_i}, δ_i, a geodesic from $hg^{m_1+s_i}$ to $hg^{m+2+s_{i-1}}$, and $\bar{\delta}_{i-1}$. The δ sides have length at most D, while the other two sides have length bounded above by $CE^{1/\alpha}$, since $s_i - s_{i-1} \leq E$, so these 2-cells have perimeter of length at most $2D + 2CE^{1/\alpha} < M_1 D + M_2 E^{1/\alpha}$.

Now define a 2-cell whose boundary consists of a geodesic from g^{s_p} to g^{s_Λ}, δ_Λ, a geodesic from h to $hg^{m_1+s_p}$, and $\bar{\delta}_p$. The bottom has length less than $CE^{1/\alpha}$, since $p=\Lambda-1$, and the vertical sides each have length at most D by hypothesis. We have

$$|m_1+s_\Lambda|^{1/\alpha} < |g^{m_1+s_\Lambda}| = |(hg^{m_1+s_\Lambda})^{-1}h| \le d(hg^{m_1+s_\Lambda}, g^{s_\Lambda}) + d(g^{s_\Lambda}, h) \le |\delta_0| + |\delta_\Lambda| \le 2D.$$

So $|m_1+s_\Lambda| < (2D)^\alpha$, which implies $|m_1+s_p| < E+(2D)^\alpha = E+LD^\alpha$ and the length of a geodesic between $hg^{m_1+s_p}$ and h is less than $CE^{1/\alpha}+2CD$, so the perimeter has length less than $2(C+1)D+2CE^{1/\alpha} < M_1D+M_2E^{1/\alpha}$.

These 2-cells glue together along common δ_i to give the desired diagram.

Now suppose $m_0 > -m_1$, as in Figure 11, right. In this case it is possible that $m_1 \ge 0$ with $p=0$. For $i \le p$ define δ_i as in the previous case. For each $i<p$, as in the previous case, define a 2-cell with boundary consisting of a geodesic from $g^{s_{i-1}}$ to g^{s_i}, δ_i, a geodesic from $hg^{m_1+s_i}$ to $hg^{m+2+s_{i-1}}$, and $\bar{\delta}_{i-1}$.

For $p<i<\Lambda$ choose a geodesic δ_i from g^{s_i} to h. As in the previous case, $0 < s_\Lambda + m_1 \le (2D)^\alpha$, so for all $p<i\le\Lambda$ we also have $0 \le s_i+m_1 \le (2D)^\alpha$. This implies that for $i>p$ we have $|\delta_i| \le (2C+1)D$. It also implies that, when $m_1 \ge 0$, we have $|g^{m_1}| < C|m_1|^{1/\alpha} \le 2D$.

Define a 2-cell with boundary consisting of a geodesic from g^{s_p} to $g^{s_{p+1}}$, δ_{p+1}, a geodesic from h to $hg^{m_1+s_p}$, and $\bar{\delta}_p$. The bottom of this 2-cell has exponent bounded by E, so has length less than $CE^{1/\alpha}$. When $m_1<0$, the top of the 2-cell similarly has length bounded by $CE^{1/\alpha}$. Otherwise, the top has length $|g^{m_1}| < 2D$. The sides have length bounded by D and $(2C+1)D$, so the perimeter has length less than $(2C+4)D+2CE^{1/\alpha} < 2(2C+1)D+2CE^{1/\alpha} = M_1D+M_2E^{1/\alpha}$.

Finally, for $p<i<\Lambda$ define a 2-cell[3] with boundary consisting of a geodesic from g^{s_i} to g^{s_i+1}, δ_{i+1}, and $\bar{\delta}_i$. It has length less than $CE^{1/\alpha}+2(2C+1)D < M_1D+M_2E^{1/\alpha}$.

Again, glue the 2-cells together along common δ_i to get the desired diagram. □

Lemma 8.2.5 (filling triangles) *There exist constants M_0, M_1, and M_2 such that for all positive Λ, E, and D, given an approximate triangle, corner paths of length at most D, and a subdivision of the a-side of the triangle into Λ-many subsegments, each of exponent at most E, the following hold. There exist subdivisions of the x-side and y-side into Λ-many subsegments of exponent bounded by $1+E/L+D^\alpha$. Furthermore, if we construct a loop γ by replacing each of the subsegments by some geodesic then there is a diagram for γ with mesh at most $M_0+M_1D+M_2E^{1/\alpha}$ and area at most $(\Lambda^2+9\Lambda+6)/2$. In fact, it suffices to take $M_0:=4C$, $M_1:=6C+2$ and $M_2:=2C$.*

Proof By Lemma 8.2.2 there is a true triangle with corresponding points at distance at most $2D+L$. That is, for each of the three sides, the corresponding pairs of sides from the approximate and true triangle

[3]We are constrained by the hypothesis that the bottom of γ uses geodesics along the given subdivision of the bottom side, so if we just flip the construction from Figure 11, left, we cannot use the estimate for the length of a geodesic along the long side of the last 2-cell in the mesh calculation. We add the extra 2-cells and accept more area to get finer mesh.

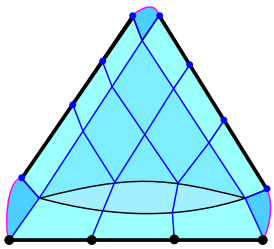


Figure 12: Filling an approximate triangle.

form a $(2D+L)$-approximate bigon. We construct a filling by filling the true triangle, using Lemma 8.2.4 to fill the three approximate bigons, and then adding one more 2-cell at each corner to close the corner path.

Apply Lemma 8.2.4 between the a-side approximate bigon. This gives a decomposition of the a-side of the true triangle into at most Λ-many subsegments of exponent at most $E+(2D)^\alpha$, and a diagram of area Λ and mesh at most $2(2C+1)D+2CE^{1/\alpha} < M_1 D + M_2 E^{1/\alpha}$.

It would be preferable, because we want to apply Lemma 3.4.4, if the subsegments of the a-side all had exponents that were multiplies of L. This can be achieved by moving the endpoints slightly along the side, each moving distance at most $L/2$. For the purposes of the diagram, we add an extra strip of 2-cells along the previous strip, the bottom of the new strip is identified with the top of the old strip and the top of the new strip is the a-side of the triangle with its new subdivision. A vertical division into 2-cells can be obtained by connecting the original subdivision points with the new ones by geodesics. The exponents of the new subdivision segments increase by at most L from the previous bound. The new strip has area Λ and mesh bounded by

$$\begin{aligned} C(E+(2D)^\alpha)^{1/\alpha} + C(L+E+(2D)^\alpha)^{1/\alpha} + \tfrac{1}{2}L + \tfrac{1}{2}L &\le 2CE^{1/\alpha} + 4CD + CL^{1/\alpha} + L \\ &= 2CE^{1/\alpha} + 4CD + 2C + L \\ &< M_0 + M_1 D + M_2 E^{1/\alpha}. \end{aligned}$$

Having made this adjustment, by Lemma 3.4.4, every subdivision point p of the a-side of the triangle has a unique closest point q on the x-side. Furthermore, from the proof of Lemma 3.4.4, every geodesic from p to q consists of a single geodesic y-escape, so they all have the same trace: it is the y-segment between p and q. A similar argument holds for p and the closest point on the y-side of the triangle.

Subdivide the triangle by taking the trace of a geodesic from each subdivision point of the a-side to its closest point on the x-side and the y-side. This divides the triangle into $(\Lambda^2-\Lambda)/2$-many small diamonds and Λ-many small triangles. See Figure 12.

The small triangles have as a-sides one of the subsegments of the a-side of the triangle, so we know they have exponent at most $L+E+(2D)^\alpha$. From Lemma 3.4.4, the resulting subdivisions of the x and y-sides have exponent that decreases by a factor of L, so all x and y-sides of all of the small triangles and small diamonds have exponents bounded by $(L+E+(2D)^\alpha)/L = 1+E/L+(2^\alpha/L)D^\alpha = 1+E/L+D^\alpha$.

From this subdivision of the triangle define a diagram with one 2-cell for each of the small diamonds and small triangles, where the 1-skeleton of the diagram is obtained by replacing each side of each of the small polygons by a geodesic with the same endpoints. The mesh of this diagram is at most

$$4\cdot C\left(\frac{L+E+(2D)^\alpha}{L}\right)^{1/\alpha} = 4C\cdot L^{-1/\alpha}(L^{1/\alpha}+E^{1/\alpha}+2D)$$
$$= 4C\cdot\frac{1}{2}(2+E^{1/\alpha}+2D) < M_0+M_1D+M_2E^{1/\alpha}.$$

The next step is to use Lemma 8.2.4 to fill the approximate bigons formed by the x-side of the approximate triangle and the x-side of the true triangle, and similarly for the y-sides. This results in subdivisions of the x-side and y-side of the approximate triangle with Λ-many pieces of exponent at most $(L+E+(2D)^\alpha)/L$ and two diagrams with area Λ and mesh at most

$$M_{1,8.2.4}D+M_{2,8.2.4}\left(\frac{L+E+(2D)^\alpha}{L}\right)^{1/\alpha} \le M_{1,8.2.4}D+M_{2,8.2.4}\left(1+\frac{E^{1/\alpha}}{2}+D\right)$$
$$= M_{2,8.2.4}+(M_{1,8.2.4}+M_{2,8.2.4})D+\left(\frac{M_{2,8.2.4}}{2}\right)E^{1/\alpha}$$
$$= 2C+(6C+2)D+CE^{1/\alpha} < M_0+M_1D+M_2E^{1/\alpha}.$$

To complete the diagram, at each corner add one 2-cell whose boundary is the given path between the sides of the approximate triangle, followed by a geodesic from one endpoint of one side of the approximate triangle to the corner of the true triangle, followed by a geodesic to the endpoint of the other side of the approximate triangle. The boundary of this 2-cell has length at most $D+2(2D+L) < M_0+M_1D$.

We have verified that the perimeter of every 2-cell of the diagram has length at most $M_0+M_1D+M_2E^{1/\alpha}$. The total area of our diagram is

$$\Lambda+\Lambda+\tfrac{1}{2}(\Lambda^2-\Lambda)+\Lambda+2\Lambda+3 = \tfrac{1}{2}(\Lambda^2+9\Lambda+6).$$ □

Lemma 8.2.6 (filling diamonds) *There exist constants M_0, M_1, and M_2 such that for all positive Λ, E, and D, given an approximate diamond, corner paths of length at most D, and subdivisions of one x-side and one y-side into at most Λ-many subsegments of exponent at most E, the following hold. There exist subdivisions of the other two sides into at most Λ-many subsegments of exponent at most $E+2LD^\alpha$. Furthermore, if we construct a loop γ by replacing each of the subsegments by some geodesic then there is a diagram for γ with mesh at most $M_0+M_1D+M_2E^{1/\alpha}$ and area at most $\Lambda^2+4\Lambda+4$. In fact, it suffices to take $M_0 := 3L$, $M_1 := 8C+2$, and $M_2 := 4C$.*

Proof The strategy is the same as in the triangle case. By Lemma 8.2.3 there is a true diamond whose corners are within distance $D+3L/2$ of the ends of the corresponding sides of the approximate diamond. Use Lemma 8.2.4 with the given subdivided sides of the approximate diamond to get subdivisions of two sides of the true diamond. Each of these consists of at most Λ-many segments of exponent at most $E+(2D)^\alpha$. In terms of diagrams we get two strips each with area at most Λ and mesh at most $2(2C+1)D+2CE^{1/\alpha} < M_1D+M_2E^{1/\alpha}$.

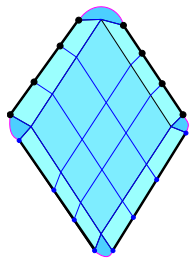


Figure 13: Filling an approximate diamond.

By Lemma 3.4.3 for each of the subdivision points p on the given x-side of the true diamond, there is a unique closest point q on the opposite x-side. Furthermore, by Proposition 3.1.1, every geodesic between them is a single geodesic y-escape, so all have the same trace; it is the y-segment between p and q. A similar statement holds for the y-sides.

Subdivide the diamond by taking the trace of a geodesic between each subdivision point and its closest point on the opposite side of the diamond. This divides the diamond into at most Λ^2-many small diamonds whose sides have the same exponent bound as we started with, $E+(2D)^\alpha$. See Figure 13.

From the subdivision of the diamond define a diagram with one 2-cell for each of the small diamonds, where the 1-skeleton is obtained by replacing each side of each small diamond by a geodesic. The mesh of this diagram is at most

$$4\cdot C\left(E+(2D)^\alpha\right)^{1/\alpha} \le 4CE^{1/\alpha}+8D < M_1D+M_2E^{1/\alpha}.$$

The next step is to use Lemma 8.2.4 to relate the remaining x and y-side of the true diamond with the corresponding sides of the approximate diamond. Each adds a subdiagram of area at most Λ and mesh at most

$$2(2C+1)D+2C(E+(2D)^\alpha)^{1/\alpha} \le 2CE^{1/\alpha}+(8C+2)D < M_1D+M_2E^{1/\alpha}.$$

The resulting subdivision of the sides of the approximate diamond have exponent at most

$$E+(2D)^\alpha+(2D)^\alpha = E+2LD^\alpha.$$

To complete the diagram, add a corner 2-cell at each of the four corners are in the triangle case. The boundaries of these 2-cells have length at most $D+2(D+3L/2)=3D+3L<M_0+M_1D$.

We have verified that the perimeter of every 2-cell of our diagram is bounded by $M_0+M_1D+M_2E^{1/\alpha}$. Its area is $2\Lambda+\Lambda^2+2\Lambda+4$. □

8.3 Base case: identifying and filling a central region

Lemma 8.3.1 *For all $\lambda\in(0,1)$ and all sufficiently large Λ depending on λ and all sufficiently large R depending on λ and Λ there exist $K>1$ depending on R, A depending on Λ, and $E:=(1/L\Lambda+2L/R^\alpha)$ and J depending on R and Λ such that for any K-bi-Lipschitz loop γ of length $|\gamma|>J$ we can construct a loop γ' from γ as follows*:

(1) *Identify a coset gH such that every escape of γ from gH has length at most $|\gamma|/2$.*

(2) *Select a manageable collection of long escapes of γ from gH.*

(3) *Subdivide the trace of each long escape into at most $2(\Lambda+1)$-many segments of exponent at most $E|\gamma|^\alpha$.*

(4) *Define γ' to be the loop obtained by replacing each long escape of γ by a concatenation of geodesics connecting the endpoint of these segments.*

The loop γ' can be filled by a diagram of area at most A and mesh at most $\lambda|\gamma|+4C$.

Proof Define

$$M_0 := \max\{M_{0,8.2.4}, M_{0,8.2.5}, M_{0,8.2.6}\} = 4C,$$
$$M_1 := \max\{M_{1,8.2.4}, M_{1,8.2.5}, M_{1,8.2.6}\},$$
$$M_2 := \tfrac{1}{2}\max\{M_{2,8.2.4}, M_{2,8.2.5}, M_{2,8.2.6}\}.$$

Assume that $\Lambda > (2M_2/\lambda)^\alpha$ and that $R > \max\{2M_1/\lambda, (L^2\Lambda/(L-1))^{1/\alpha}\}$. Let J and K be as in Corollary 7.2.4 with respect to R, and further increase J if necessary so that

$$J \geq \left(\frac{L-1}{L^2\Lambda} - \frac{1}{R^\alpha}\right)^{-1/\alpha}.$$

Assume $|\gamma| > J$. As described in step 2 of Section 5.2, any HNN diagram for γ has a central region. The boundary of the central region maps to a loop in some coset gH of the vertex group. Up to isometry, we may assume $g = 1$. By definition, the components of γ complementary to the central region have length at most $|\gamma|/2$. These components are compound escapes of γ from H, and they contain the simple escapes from H as subpaths, so all of the escapes of γ from H have length at most $|\gamma|/2$. Corollary 7.2.4 gives us a manageable collection of 2, 3, or 4 R-long escapes from H. Call the subsegments of γ between these long escapes the "junk". The total length of the junk is strictly less than $|\gamma|/R$.

Together the traces of the long escapes form an approximate 2-, 3-, or 4-gon, and the junk segments give corner paths of length less than $|\gamma|/R$. Furthermore, since each escape from the central region has length at most $|\gamma|/2$, its trace has exponent at most $(|\gamma|/2)^\alpha$. We will apply Lemmas 8.2.4, 8.2.5, and 8.2.6, in all cases with $D_{8.*} = |\gamma|/R$, $E_{8.*} = (|\gamma|/2)^\alpha/\Lambda$, and $\Lambda_{8.*} = \Lambda+1$. Then all three lemmas produce a diagram with

$$\textbf{(8.3.2)} \qquad \text{mesh} \leq M_0 + \frac{M_1|\gamma|}{R} + \frac{M_2|\gamma|}{\Lambda^{1/\alpha}}, \quad \text{area} \leq (\Lambda+1)^2 + 4(\Lambda+1) + 4.$$

To construct subdivisions and fillings there are six cases to consider, according to the configuration of traces of the escapes of the manageable collection as in Figure 7.

Case (7A) Approximate bigon: Choose a subdivision of one side into at most $(\Lambda+1)$-many segments of exponent at most $(|\gamma|/2)^\alpha/\Lambda$. Apply Lemma 8.2.4. This yields a subdivision of the opposite side into at most $(\Lambda+1)$-many segments of exponent at most $|\gamma|^\alpha(1/L\Lambda + L/R^\alpha)$ and a filling diagram with mesh and area bounded by (8.3.2).

Case (7C) Approximate triangle: Choose a subdivision of the a-side into at most $(\Lambda+1)$-many subsegments of exponent at most $(|\gamma|/2)^\alpha/\Lambda$. Apply Lemma 8.2.5. This yields subdivisions of the other two sides into at most $(\Lambda+1)$-many segments of exponent at most $1+|\gamma|^\alpha(1/(L^2\Lambda)+1/R^\alpha)$ and a filling diagram with mesh and area bounded by (8.3.2).

Case (7F) Approximate diamond: Choose subdivisions of one x-side and one y-side into at most $\Lambda+1$ many segments of exponent at most $(|\gamma|/2)^\alpha/\Lambda$. Apply Lemma 8.2.6. This yields subdivisions of the other two sides into $(\Lambda+1)$-many segments of exponent at most $|\gamma|^\alpha(1/(L\Lambda)+2L/R^\alpha)$ and a filling diagram with mesh and area bounded by (8.3.2).

Case (7B) The traces of the manageable escapes can be described as if they were an approximate triangle by elements g_0, g_1, and g_2 and numbers m_0, m_1, and m_2, with the exception that instead of the sides alternating x, y, and a, they are all of the same flavor $g \in \{a, x, y\}$. Let us assume the g_2-side is the long side. Let δ_i be the junk segment from $g_{i-1}g^{m_{i-1}}$ to g_i. Then $g^{m_0}\delta_0$ is a path from $g_2g^{m_2+m_0}$ to $g_0g^{m_0}$.

We treat this configuration as a pair of approximate bigons. One of these has side δ_0, top the g_0-side, side $g^{m_0}\bar{\delta}_0$, and bottom the part of the g_2-side from $g_2g^{m_2+m_0}$ to $g_2g^{m_2}$. The two sides have length $|\delta_0| < |\gamma|/R$. The second bigon has side $g^{m_0}\delta_0+\delta_1$, top the g_1-side, side $\bar{\delta}_2$, and bottom the part of the g_2-side from g_2 to $g_2g^{m_2+m_0}$. One side has length $|\delta_2| < |\gamma|/R$. The other has length $|\delta_0|+|\delta_1|$, which, since our junk bound is actually on the total length of the junk and not the individual segments, is still less than $|\gamma|/R$.

Choose a subdivision of the g_0 and g_1 sides into at most $(\Lambda+1)$-many subsegments of exponent at most $(|\gamma|/2)^\alpha/\Lambda$. Apply Lemma 8.2.4 to each of the two approximate bigons. The constants involved are exactly the same as in case (7A), except that the bounds for the area and number of induced subsegments of the g_2-side are doubled.

Case (7D) Use the same trick as in case (7B): use the group action to translate the y-side and its attached junk segments to the opposite end of the shorter x-side, and view this configuration as a combination of an approximate diamond and an approximate triangle, and fill them separately.

Case (7E) In this case think of the configuration as two approximate triangles meeting at the point where the a-side and y-side intersect. Fill them each separately and combine the fillings.

In all cases we produce diagrams with mesh bounded by $M_0+M_1|\gamma|/R+M_2|\gamma|/\Lambda^{1/\alpha}$, so since $M_0=4C$, $R>2M_1/\lambda$, and $\Lambda>(2M_2/\lambda)^\alpha$, we have mesh bounded by $\lambda|\gamma|+4C$.

In all cases we fill at most two approximate polygons, each with a diagram of area no more than $(\Lambda+1)^2+4(\Lambda+1)+4$, so it suffices to take A to be double this bound.

Scanning the six cases for exponent bounds, we have that the hypotheses

$$R>\left(\frac{L^2\Lambda}{L-1}\right)^{1/\alpha} \quad \text{and} \quad |\gamma|>J\geq\left(\frac{L-1}{L^2\Lambda}-\frac{1}{R^\alpha}\right)^{-1/\alpha}$$

imply that the largest of the bounds for the exponents of the subdivision segments is at most $|\gamma|^\alpha\big(\frac{1}{L\Lambda}+\frac{2L}{R^\alpha}\big)$, so it suffices to take $E := \big(\frac{1}{L\Lambda}+\frac{2L}{R^\alpha}\big)$. □

8.4 Inductive steps

Inductive step: capping off small enough parts Recall (8.1.3). There γ had a central escape, from some coset gH, whose trace in gH had been subdivided into at most $2(\Lambda+1)$-many subsegments of exponent at most $L(|\gamma|/R)^\alpha$, which were then replaced by geodesics $\delta_1,\dots,\delta_m$. Then if $\gamma|_P$ was the part of γ complementary to the central escape, we argued that the length of $\gamma|_P+\bar\delta_m+\cdots+\bar\delta_1$ was small enough to cap off with a single 2-cell satisfying the desired mesh bound.

Here we argue additionally that very small parts can be capped off. Recall from Section 7.3 that the noncentral part $\gamma|_P$ of γ at some coset of H is called "very small" when its length is less than J, where J is some prespecified constant, independent of γ. The setup is similar to the case described following (8.1.3): the central escape of γ from the coset has trace subdivided into at most $m=2(\Lambda+1)$-many subsegments, and there are given geodesics δ_i with the same endpoints as the subsegments.

(8.4.1) $$\text{Assume } M \ge (C(2(\Lambda+1))^{1-1/\alpha}+1)J.$$

Since $\gamma|_P$ has the same endpoints as the central escape and length less than J, by Theorem 3.2.9 the exponent of the central escape is less than J^α. Suppose that the exponent of δ_i is e_i. Then $|\sum_i e_i|<J^\alpha$. By Theorem 3.2.9, $|\delta_i|<C|e_i|^{1/\alpha}$. By reverse Hölder, $\sum_{i=1}^m |e_i|^{1/\alpha}\le m^{1-1/\alpha}\big|\sum_{i=1}^m e_i\big|^{1/\alpha}$. Therefore, the length of $\gamma|_P+\bar\delta_m+\cdots+\bar\delta_1$ is less than

(8.4.2) $$(Cm^{1-1/\alpha}+1)J \le \big(C(2(\Lambda+1))^{1-1/\alpha}+1\big)J \le M.$$

Thus, we can cap off a very small part $\gamma|_P$ of γ with a single 2-cell with boundary $\gamma|_P+\bar\delta_m+\cdots+\bar\delta_1$, whose length is at most M.

Inductive step: enfilades

(8.4.3) $$\text{Define } M_1 := 2C+1+C\cdot 2^{2+1/\alpha} \text{ and } M_2 := C.$$

(8.4.4) $$\text{Assume } R \ge \frac{2M_1}{\lambda} \text{ and } \Lambda \ge \left(\frac{2M_2}{\lambda}\right)^\alpha.$$

Suppose we have an escape $\gamma|_P$ in the n^{th} level, so that by (8.1.1) the central side has been subdivided into at most $2(\Lambda+1)$-many geodesic segments of exponent at most $|\gamma|^\alpha E(3/L)^n$, where E is the constant of Lemma 8.3.1, which depends on R and Λ, but not on γ. Suppose this exponent bound exceeds that of (8.1.3), so that the escape is not small enough to cap off. This means:

(8.4.5) $$|\gamma|^\alpha E\left(\frac{3}{L}\right)^n > L\left(\frac{|\gamma|}{R}\right)^\alpha \implies \frac{1}{R^\alpha} < \frac{3^n E}{L^{n+1}}.$$

(8.4.6) $$\text{Assume } J \text{ and } K \text{ satisfy Proposition 7.3.5 with } R_{7.3.5} = 3R.$$

Apply Proposition 7.3.5 with $R_{7.3.5} = 3R$. Then the distance between the endpoints of the escape and the corresponding endpoints of the end of the enfilade is at most $3|\gamma|_P|/R_{7.3.5} = |\gamma|_P|/R \leq |\gamma|/2R$. We claim that we can fill the enfilade with the same argument as Lemma 8.2.4, treating the enfilade as a $(|\gamma|/(2R))$-approximate bigon. Specifically, the point in that proof where we used that the two sides of the bigon were in the same coset of $\langle a, x, y\rangle$ was to say that we could translate δ_0 along the bottom side and obtain "parallel" copies, in the sense that $g^a.\delta_0$ and $g^b.\delta_0$ are both paths from the bottom side to the top side and their endpoint at the bottom and top both differ by g raised to the power $b-a$. By Lemma 7.3.3, exactly the same thing is true for enfilades, except that we possibly have different g's from the set $\{a, x, y\}$ on the two sides. As our length versus exponent estimates are uniform for $\{a, x, y\}$, this does not change the rest of the argument.

Apply that argument with $\Lambda_{8.2.4} := 2(\Lambda+1)$, $E_{8.2.4} := |\gamma|^\alpha E(3/L)^n$, and $D_{8.2.4} := |\gamma|/(2R)$, with $M_{i,8.2.4}$ denoting the constant M_i in the conclusion of Lemma 8.2.4. This yields a subdivision of the outgoing side into at most $2(\Lambda+1)$-many subsegments of exponent at most $E_{8.2.4} + LD^\alpha_{8.2.4}$ and a filling diagram of area $\Lambda_{8.2.4}$ and mesh at most

$$
\begin{aligned}
M_{0,8.2.4} + M_{1,8.2.4}D_{8.2.4} + M_{2,8.2.4}E^{1/\alpha}_{8.2.4} &= 0 + 2(2C+1)D_{8.2.4} + 2CE^{1/\alpha}_{8.2.4} \\
&= \frac{(2C+1)|\gamma|}{R} + 2C|\gamma|\left(E\left(\frac{3}{L}\right)^n\right)^{1/\alpha} \\
&= \frac{(2C+1)|\gamma|}{R} + 2CE^{1/\alpha}\left(\frac{3}{L}\right)^{n/\alpha}\cdot|\gamma| \\
&= \frac{(2C+1)|\gamma|}{R} + 2C\left(\frac{1}{L\Lambda} + \frac{2L}{R^\alpha}\right)^{1/\alpha}\cdot\left(\frac{3}{L}\right)^{n/\alpha}\cdot|\gamma| && \text{(by definition of } E\text{)} \\
&< \frac{(2C+1)|\gamma|}{R} + 2C\left(\frac{1}{L\Lambda} + \frac{2L}{R^\alpha}\right)^{1/\alpha}\cdot|\gamma| && \text{(because } L > 3\text{)} \\
&\leq \frac{(2C+1)|\gamma|}{R} + 2C\left(\frac{1}{2\Lambda^{1/\alpha}} + \frac{2^{1+1/\alpha}}{R}\right)\cdot|\gamma| && \text{(by reverse Hölder)} \\
&= \frac{(2C+1+C\cdot 2^{2+1/\alpha})|\gamma|}{R} + \frac{C|\gamma|}{\Lambda^{1/\alpha}} \\
&= \frac{M_1|\gamma|}{R} + \frac{M_2|\gamma|}{\Lambda^{1/\alpha}} && \text{(by definition of } M_i\text{).}
\end{aligned}
$$

Since $R \geq 2M_1/\lambda$ and $\Lambda \geq (2M_2/\lambda)^\alpha$, the mesh is at most $\lambda|\gamma|$.

Consider the ratio of outgoing to incoming exponent bounds,

$$
\begin{aligned}
\frac{E_{8.2.4} + LD^\alpha_{8.2.4}}{E_{8.2.4}} = 1 + \frac{L(|\gamma|/(2R))^\alpha}{|\gamma|^\alpha E(3/L)^n} = 1 + \frac{L^{n+1}}{2^\alpha 3^n ER^\alpha} &= 1 + \frac{L^n}{3^n ER^\alpha} \\
&< 1 + \frac{L^n}{3^n E}\cdot\frac{3^n E}{L^{n+1}} \quad \text{(by (8.4.5))} \\
&\leq 1 + \frac{1}{L} = \frac{L+1}{L}.
\end{aligned}
$$

Recall that for the induction we are aiming for an exponent ratio of $3/L$, and here we have only shown for enfilades that it is not too much larger than 1. The branching parts compensate.

Inductive step: branching parts

(8.4.7) Let $M_0 := 4C$, $M_1 := C \cdot 2^{3+1/\alpha} + 6C + 2$, and $M_2 := 2C$.

Assume (8.4.4) and (8.4.6) are satisfied, and furthermore:

(8.4.8) Assume $R \geq 2M_1/\lambda$ and $\Lambda \geq (2M_2/\lambda)^\alpha$ and $J \geq 2R$ and $M \geq M_0$.

Let E be the constant of Lemma 8.3.1, which depends on R and Λ.

Suppose now that we have reached a branching part of γ at the end of an enfilade as in the previous subsection. Then by the induction hypothesis (8.1.1) and the result of the previous subsection, the trace of the central escape is subdivided into at most $2(\Lambda+1)$-many subsegments of exponent at most $\frac{L+1}{L}|\gamma|^\alpha E(3/L)^n$. Again, we assume that the branching part is not small enough to cap off yet, so that (8.1.3) fails and we have

(8.4.9) $$\frac{L+1}{L}|\gamma|^\alpha E\left(\frac{3}{L}\right)^n > L\left(\frac{|\gamma|}{R}\right)^\alpha \implies \frac{1}{R^\alpha} < \frac{3^n E(L+1)}{L^{n+2}}.$$

The traces of the central and two long noncentral escapes form an approximate triangle, so we apply Lemma 8.2.5 with $\Lambda_{8.2.5} := 2(\Lambda+1)$, $E_{8.2.5} := \frac{L+1}{L}|\gamma|^\alpha E(3/L)^n$, and $D_{8.2.5} := |\gamma|/R$. This yields subdivisions of the other two sides into at most $2(\Lambda+1)$-many subsegments of exponent at most $1 + E_{8.2.5}/L + D_{8.2.5}^\alpha$ and a filling diagram of area quadratic in Λ and mesh at most

$$\begin{aligned}
M_{2,8.2.5}E_{8.2.5}^{1/\alpha} + M_{1,8.2.5}D_{8.2.5} + M_{0,8.2.5} &= 2CE_{8.2.5}^{1/\alpha} + (6C+2)D_{8.2.5} + 4C \\
&= 2C\left(\frac{L+1}{L}|\gamma|^\alpha E\left(\frac{3}{L}\right)^n\right)^{1/\alpha} + \frac{(6C+2)|\gamma|}{R} + 4C \\
&< 4C|\gamma|E^{1/\alpha} + \frac{(6C+2)|\gamma|}{R} + 4C && \text{(because } L \geq 6 > 3\text{)} \\
&= 4C|\gamma| \cdot \left(\frac{1}{L\Lambda} + \frac{2L}{R^\alpha}\right)^{1/\alpha} + \frac{(6C+2)|\gamma|}{R} + 4C && \text{(by definition of } E\text{)} \\
&\leq 4C|\gamma| \cdot \left(\frac{1}{2\Lambda^{1/\alpha}} + \frac{2^{1+1/\alpha}}{R}\right) + \frac{(6C+2)|\gamma|}{R} + 4C && \text{(by reverse Hölder)} \\
&= \frac{(C \cdot 2^{3+1/\alpha} + 6C + 2)|\gamma|}{R} + \frac{2C|\gamma|}{\Lambda^{1/\alpha}} + 4C \\
&= M_0 + \frac{M_1|\gamma|}{R} + \frac{M_2|\gamma|}{\Lambda^{1/\alpha}} && \text{(by definition of } M_i\text{)}.
\end{aligned}$$

Since $M \geq M_0$, $R \geq 2M_1/\lambda$, and $\Lambda \geq (2M_2/\lambda)^\alpha$, the mesh is less than $\lambda|\gamma| + M$.

Since $|\gamma| > J \geq 2R$, the ratio of outgoing to incoming exponents is at most

$$\begin{aligned}
\frac{1 + E_{8.2.5}/L + D^{\alpha}_{8.2.5}}{E_{8.2.5}} &= \frac{1}{L} + \frac{L}{L+1} \cdot \frac{L^n}{3^n E|\gamma|^{\alpha}} + \frac{L}{L+1} \cdot \frac{L^n}{3^n E R^{\alpha}} \\
&= \frac{1}{L} + \frac{L}{L+1} \cdot \frac{L^n}{3^n E} \left(\frac{R^{\alpha}}{|\gamma|^{\alpha}} + 1 \right) \frac{1}{R^{\alpha}} \\
&\leq \frac{1}{L} + \frac{L}{L+1} \cdot \frac{L^n}{3^n E} \cdot \frac{5}{4} \cdot \frac{1}{R^{\alpha}} && \text{(since } \alpha > 2 \text{ and } |\gamma| \geq 2R\text{)} \\
&\leq \frac{1}{L} + \frac{L}{L+1} \cdot \frac{L^n}{3^n E} \cdot \frac{5}{4} \cdot \frac{3^n E(L+1)}{L^{n+2}} && \text{(by (8.4.9))} \\
&= \frac{9}{4L}.
\end{aligned}$$

This means that if we take the ratio of the exponent bound when entering the enfilade to the bound when exiting through one of the long escapes of the branching part at the end of the enfilade, then we get $\frac{L+1}{L} \cdot \frac{9}{4L}$. Since we have $L > 3$, it follows that $\frac{L+1}{L} \cdot \frac{9}{4L} < \frac{3}{L}$, so this confirms (8.1.1).

9 Proof of the main theorem and closing remarks

Theorem 9.0.1 *For all even $L \geq 6$, every asymptotic cone of G_L is simply connected.*

Proof For $i \in \{1, 2\}$ let $M_i := \max\{M_{i,8.3.1}, M_{i,(8.4.3)}, M_{i,(8.4.7)}\}$. These depend only on C.

Pick any $\lambda \in (0, 1)$, any $\Lambda > (2M_2/\lambda)^{\alpha}$, and any $R > \max\left\{3, 2M_1\lambda, 12C(\Lambda+1)/\lambda, \left(L^2\Lambda/(L-1)\right)^{1/\alpha}\right\}$. Pick J large enough and $1 < K < 2$ close enough to 1 to satisfy Corollary 7.2.4 and Lemma 8.3.1 with respect to R and Proposition 7.3.5 with respect to $3R$. Assume further that $J \geq 2R$. Define $M = M_0 := \max\left\{4C, \left(C(2(\Lambda+1))^{1-1/\alpha} + 1\right)J\right\}$.

These choices simultaneously satisfy Lemma 8.3.1 and the assumptions (8.4.1), (8.4.3), (8.4.4), (8.4.6), (8.4.7), and (8.4.8).

Let $\mathscr{C}$ be the area bound given by Lemma 8.3.1, which is a quadratic function of Λ. Let

$$\mathscr{E} = \Lambda_{8.2.4} = 2(\Lambda + 1)$$

be the area bound for enfilades. Let $\mathscr{B}$ be the area bound for branching parts, which is quadratic in Λ. Let A be the area bound of (8.1.5) with respect to $\mathscr{C}$, $\mathscr{E}$, and $\mathscr{B}$.

If γ is a K-bi-Lipschitz loop of length at most J we fill it by a single 2-cell with mesh $J < M$.

If γ is a K-bi-Lipschitz loop of length greater than J then we apply the inductive argument of Section 8 to get a filling diagram of γ with area at most A and mesh at most the maximum of the meshes for the central part, very small parts, enfilades, and branching parts. We have chosen M_0, M_1, M_2, R, and Λ so that this is at most $\lambda|\gamma| + M$.

We conclude that for this choice of $K > 1$, every K-bi-Lipschitz loop γ can be filled by a diagram of area at most A and mesh at most $\lambda|\gamma| + M$, with A, λ, and M independent of γ. Now apply Corollary 6.0.3 and Theorem 2.3.2 to conclude that every asymptotic cone of G_L is simply connected. □

We close with a list of open questions:

(1) Is G_L a strongly shortcut group?

(2) If every asymptotic cone of a group is simply connected, is the group strongly shortcut?

(3) Is the strong shortcut property for Cayley graphs of groups invariant under change of generating set?

(4) Is the strong shortcut property for groups invariant under quasi-isometry?

(5) Can a group of polynomial growth have an asymptotic cone containing an isometrically embedded circle?

(6) Can a group with quadratic Dehn function have an asymptotic cone containing an isometrically embedded circle?

(7) Are all asymptotic cones of all of the Brady–Bridson snowflake groups $G_{p,q}$ of (3.0.2) simply connected, or is it important that we have taken $q = 1$?

Note that, by Theorem 9.0.1 and Corollary 4.0.2,

- if (1) has a negative answer then (2) has a negative answer and
- if (1) has a positive answer then (3) has a negative answer.

Thus (2) and (3) cannot both have positive answers.

References

[1] **G N Arzhantseva**, **C H Cashen**, **J Tao**, *Growth tight actions*, Pacific J. Math. 278 (2015) 1–49 MR Zbl

[2] **S P Avann**, *Metric ternary distributive semi-lattices*, Proc. Amer. Math. Soc. 12 (1961) 407–414 MR Zbl

[3] **H-J Bandelt**, *Hereditary modular graphs*, Combinatorica 8 (1988) 149–157 MR Zbl

[4] **J A Behrstock**, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. 10 (2006) 1523–1578 MR Zbl

[5] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, Geom. Topol. 21 (2017) 1731–1804 MR Zbl

[6] **J Behrstock**, **M Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, II: Combination theorems and the distance formula*, Pacific J. Math. 299 (2019) 257–338 MR Zbl

[7] **J Behrstock**, **B Kleiner**, **Y Minsky**, **L Mosher**, *Geometry and rigidity of mapping class groups*, Geom. Topol. 16 (2012) 781–888 MR Zbl

[8] **J A Behrstock**, **Y N Minsky**, *Dimension and rank for mapping class groups*, Ann. of Math. 167 (2008) 1055–1077 MR Zbl

[9] **N Brady**, **M R Bridson**, *There is only one gap in the isoperimetric spectrum*, Geom. Funct. Anal. 10 (2000) 1053–1070 MR Zbl

[10] **N Brady**, **T Riley**, **H Short**, *The geometry of the word problem for finitely generated groups*, Birkhäuser, Basel (2007) MR Zbl

[11] **M R Bridson**, *Asymptotic cones and polynomial isoperimetric inequalities*, Topology 38 (1999) 543–554 MR Zbl

[12] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR Zbl

[13] **C H Cashen**, *Quasi-isometries between tubular groups*, Groups Geom. Dyn. 4 (2010) 473–516 MR Zbl

[14] **J Chalopin**, **V Chepoi**, **A Genevois**, **H Hirai**, **D Osajda**, *Helly groups*, Geom. Topol. 29 (2025) 1–70 MR Zbl

[15] **V Chepoi**, *Graphs of some* CAT(0) *complexes*, Adv. in Appl. Math. 24 (2000) 125–179 MR Zbl

[16] **L van den Dries**, **A J Wilkie**, *Gromov's theorem on groups of polynomial growth and elementary logic*, J. Algebra 89 (1984) 349–374 MR Zbl

[17] **C Druțu**, *Quasi-isometry invariants and asymptotic cones*, Internat. J. Algebra Comput. 12 (2002) 99–135 MR Zbl

[18] **C Druțu**, **M Sapir**, *Tree-graded spaces and asymptotic cones of groups*, Topology 44 (2005) 959–1058 MR Zbl

[19] **M Gromov**, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. 53 (1981) 53–73 MR Zbl

[20] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR Zbl

[21] **M Gromov**, *Asymptotic invariants of infinite groups*, from "Geometric group theory, II" (G A Niblo, M A Roller, editors), Lond. Math. Soc. Lect. Note Ser. 182, Cambridge Univ. Press (1993) 1–295 MR Zbl

[22] **T Haettel**, **N Hoda**, **H Petyt**, *Coarse injectivity, hierarchical hyperbolicity and semihyperbolicity*, Geom. Topol. 27 (2023) 1587–1633 MR Zbl

[23] **F Haglund**, *Complexes simpliciaux hyperboliques de grande dimension*, preprint (2003) Available at https://www.imo.universite-paris-saclay.fr/~frederic.haglund/cpl_hyp_gde_dim.pdf

[24] **N Hoda**, *Quadric complexes*, Michigan Math. J. 69 (2020) 241–271 MR Zbl

[25] **N Hoda**, *Shortcut graphs and groups*, Trans. Amer. Math. Soc. 375 (2022) 2417–2458 MR Zbl

[26] **N Hoda**, *Strongly shortcut spaces*, Algebr. Geom. Topol. 24 (2024) 3291–3325 MR Zbl

[27] **T Januszkiewicz**, **J Świątkowski**, *Simplicial nonpositive curvature*, Publ. Math. Inst. Hautes Études Sci. 104 (2006) 1–85 MR Zbl

[28] **A Kar**, *Asymptotically* CAT(0) *groups*, Publ. Mat. 55 (2011) 67–91 MR Zbl

[29] **C Kent**, *Asymptotic cones of HNN extensions and amalgamated products*, Algebr. Geom. Topol. 14 (2014) 551–595 MR Zbl

[30] **E Le Donne**, **N Paddeu**, *Escape from compact sets of normal curves in Carnot groups*, preprint (2023) arXiv 2304.03205

[31] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR Zbl

[32] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR Zbl

[33] **L Nebeský**, *Median graphs*, Comment. Math. Univ. Carolinae 12 (1971) 317–325 MR Zbl

[34] **G A Niblo**, **L D Reeves**, *Coxeter groups act on* CAT(0) *cube complexes*, J. Group Theory 6 (2003) 399–413 MR Zbl

[35] **A Y Ol'shanskii**, **M V Sapir**, *Groups with non-simply connected asymptotic cones*, from "Topological and asymptotic aspects of group theory" (R Grigorchuk, M Mihalik, M Sapir, Z Šuniḱ, editors), Contemp. Math. 394, Amer. Math. Soc., Providence, RI (2006) 203–208 MR Zbl

[36] **D Osajda**, **T Prytuła**, *Classifying spaces for families of subgroups for systolic groups*, Groups Geom. Dyn. 12 (2018) 1005–1060 MR Zbl

[37] **P Pansu**, *Croissance des boules et des géodésiques fermées dans les nilvariétés*, Ergodic Theory Dynam. Systems 3 (1983) 415–445 MR Zbl

[38] **P Papasoglu**, *On the asymptotic cone of groups satisfying a quadratic isoperimetric inequality*, J. Differential Geom. 44 (1996) 789–806 MR Zbl

[39] **T R Riley**, *Higher connectedness of asymptotic cones*, Topology 42 (2003) 1289–1352 MR Zbl

[40] **M V Sapir**, **J-C Birget**, **E Rips**, *Isoperimetric and isodiametric functions of groups*, Ann. of Math. 156 (2002) 345–466 MR Zbl

[41] **V P Soltan**, **V D Chepoi**, *Conditions for invariance of set diameters under d-convexification in a graph*, Cybernetics 19 (1983) 750–756 MR Zbl

[42] **D T Wise**, *Sixtolic complexes and their fundamental groups*, unpublished manuscript (2003)

Faculty of Mathematics, University of Vienna
Vienna, Austria

Department of Mathematics, Cornell University
Ithaca, NY, United States

Mathematical Institute, University of Oxford
Oxford, United Kingdom

christopher.cashen@univie.ac.at, nima@nimahoda.net, daniel.woodhouse@mail.mcgill.ca

Received: 1 March 2023 Revised: 1 February 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3503–3521
DOI: 10.2140/agt.2025.25.3503
Published: 1 October 2025

Whitney tower concordance and knots in homology spheres

CHRISTOPHER W DAVIS

A Levine (Forum Math. Sigma 4 (2016) art. id. e34) proved the surprising result that there exist knots in homology spheres which are not smoothly concordant to any knot in S^3, where one allows for concordances in homology cobordisms. Since then subsequent works due to Hom, Levine and Lidman (Duke Math. J. 171 (2022) 3089–3131) and Zhou (J. Topol. 14 (2021) 1369–1395) have strengthened this result showing that there are many knots in homology spheres which are not smoothly concordant to knots in S^3. In this paper we present evidence that the opposite is true topologically. We study the Whitney tower filtration of concordance (Ann. of Math. 157 (2003) 433–519) and prove that modulo any term in this filtration every knot (or link) in a homology sphere is equivalent to a knot (or link) in S^3. As an application we recover a result of the author (J. Topol. 13 (2020) 343–355), namely that the solvable filtration similarly fails to distinguish links in homology spheres from links in S^3.

57K10, 57N70

1 Introduction and statement of results

Two links K and J in S^3 are called (*topologically*) *concordant* if $K\times\{1\}$ and $J\times\{0\}$ cobound a disjoint union of locally flat embedded annuli in $S^3\times[0,1]$. A link is (*topologically*) *slice* if it bounds a disjoint union of locally flat embedded disks, called slice disks, in the 4-ball, B^4. A link is slice if and only if it is concordant to the unlink. The set of concordance classes of links is denoted by $\mathcal{C}$. If these embedded annuli and disks are smooth, then we call the resulting links smoothly concordant and smoothly slice, respectively. We denote by $\mathcal{C}_{\mathrm{sm}}$ the set of links up to smooth concordance. The difference between $\mathcal{C}$ and $\mathcal{C}_{\mathrm{sm}}$ has been central to the interplay between topological and smooth 4-manifolds. For instance, the existence of links which are topologically slice but not smoothly slice can be used to produce smooth 4-manifolds which are homeomorphic but not diffeomorphic to $\mathbb{R}^4$ [9]. In this paper we present evidence for a surprising difference between smooth and topological concordance of links in homology spheres.

Conjecture 1.1 Let $K\subseteq Y$ be a link in a homology sphere. There is a link J in S^3 and a homology cobordism W from Y to S^3 in which K and J cobound a disjoint union of locally flat embedded annuli. Moreover, W can be chosen to be simply connected.

This conjecture is false in the smooth setting, as proved in [14] with subsequent improvements in [11; 19]. Recall that a homology cobordism between two 3-manifolds X and Y is informally a 4-manifold bounded by $X\sqcup -Y$ which has the same homology as $X\times[0,1]$. Homology cobordism of homology spheres was first studied in [10, Chapter 1], inspired by the study of h-cobordism of homotopy spheres of [12].

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

This conjecture fits into two extensions of concordance to the setting of links in homology spheres. The first is $\mathbb{Z}$-concordance. Two links K and J in homology spheres X and Y are called $\mathbb{Z}$-*concordant* if there is a homology cobordism from X to Y in which the components of K and J cobound disjoint locally flat embedded annuli. The $\mathbb{Z}$-concordance class of the unlink is the set of links whose components bound disjoint locally flat embedded disks in a homology ball. We denote by $\widehat{\mathcal{C}}_{\mathbb{Z}}$ the set of links in homology spheres up to $\mathbb{Z}$-concordance.

Suppose that W is the homology cobordism of the above paragraph. If $\pi_1(X)\to\pi_1(W)$ and $\pi_1(Y)\to\pi_1(W)$ are each surjective then W is called a *strong homology cobordism* and K and J are called *strongly concordant*. We denote by $\widehat{\mathcal{C}}$ the set of all links up to strong topological concordance. The strong concordance class of the unlink is the set of all links in homology spheres which bound disjoint unions of embedded disks in contractible 4-manifolds. There is an inclusion and a projection map

$$\mathcal{C} \xhookrightarrow{i} \widehat{\mathcal{C}} \xrightarrow{p} \widehat{\mathcal{C}_{\mathbb{Z}}}.$$

The injectivity of i follows from the 4-dimensional topological Poincaré conjecture [7]. Conjecture 1.1 can now be interpreted as saying that $\mathcal{C}\to\widehat{\mathcal{C}}$ is a surjection, and as a consequence is bijective.

In [5] the author presents evidence for Conjecture 1.1 in the language of the solvable filtration of Cochran, Orr and Teichner [3]. While this filtration has since received intense study and has been used to detect highly subtle structure in the study of link concordance, it is the weakest of three filtrations studied in that paper. The others are the symmetric Whitney tower filtration and the grope filtration. We focus on the Whitney tower filtration,

$$\mathcal{C} \supseteq \mathcal{W}_2 \supseteq \mathcal{W}_3 \supseteq \cdots.$$

A link in S^3 sits in $\mathcal{W}_h$ if it bounds a height $h-1$ symmetric Whitney tower in the 4-ball. While we give a brief discussion presently and a complete definition in Section 2, one can think of a Whitney tower as an immersed disk which comes close to admitting a regular homotopy to an embedded disk. Thus, if a link is in $\mathcal{W}_h$ then it is close to being slice. See, for example, [3, Definition 7.7; 4, Section 2.1]. A *Whitney tower*, T, is a 2-complex in a 4-manifold given by starting with an immersed disk, pairing up its double points with (possibly nonembedded) Whitney disks, as in Figure 1, left, pairing up the double points amongst these Whitney disks with Whitney disks as in Figure 1, middle, and so on. If any of these Whitney disks is embedded and has interior disjoint from all other surfaces in T, then the Whitney trick of Figure 1, right, can be used to remove these two intersection points. See, for example, [1, Section 1.2] for a discussion of the Whitney trick. If each round of Whitney disks added has interior disjoint from all of the previously added Whitney disks then T is called *symmetric* and the height of T records how many rounds of Whitney disks are added. For any $h\in\mathbb{N}$, $\mathcal{W}_h$ is the set of all links in homology spheres bounding immersed disks in contractible 4-manifolds that extend to height $h-1$ symmetric Whitney towers.

In [2, Definition 2.12], Cha extends $\mathcal{W}_h$ to an equivalence relation. We construct a similar equivalence relation for links in homology spheres as follows: Links in homology spheres are height h Whitney tower

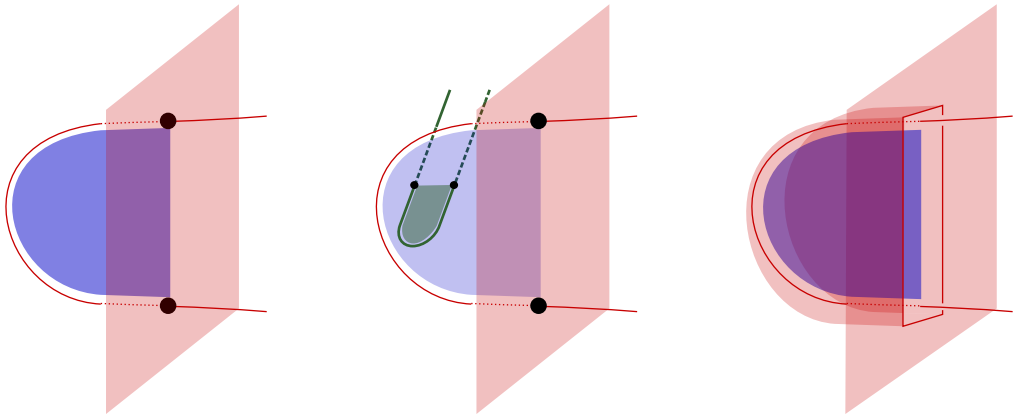


Figure 1: Left: A Whitney disk pairing up two points of intersection. Middle: A (portion of a) Whitney tower. Right: The Whitney moves canceling two points of intersection.

concordant if they cobound an immersed union of annuli in a strong homology cobordism that extends to a height $h-1$ Whitney tower. See Definition 2.5 for a more complete discussion. Our main result is that every link in a homology sphere is height h Whitney tower concordant to a link in S^3.

Theorem 1.2 *Let $h \in \mathbb{N}$ and K be a link in a homology sphere. There is some link in S^3 which is height h Whitney tower concordant to K.*

This theorem should be interpreted as saying that any link in a homology sphere cobounds with a link in S^3 an immersed union of annuli which is close to admitting a sequence of Whitney moves to a disjoint union of locally flat embedded annuli. Thus, any link in any homology sphere is close (from the perspective of Whitney towers) to being strongly concordant to a link in S^3.

As mentioned previously, there are two related filtrations of knot concordance, the solvable filtration $\mathcal{C} \supseteq \mathcal{F}_0 \supseteq \mathcal{F}_1 \supseteq \cdots$ and the grope filtration $\mathcal{C} \supseteq \mathcal{G}_2 \supseteq \mathcal{G}_3 \supseteq \cdots$ [3, Definitions 7.9 and 8.5]. The symmetric Whitney tower filtration gives stronger approximations to sliceness than the solvable filtration in that $\mathcal{W}_{h+2} \subseteq \mathcal{F}_h$ [3, Theorem 8.12]. In [5, Theorem 1.2] the author defines a sequence of equivalence relations called n-solvable concordance and proves that modulo n-solvable concordance every link in a homology sphere is equivalent to a link in S^3. In Section 5 we show that if links are height $h+2$ Whitney tower concordant then they are h-solvably concordant.

Proposition 5.4 *Let K and J be links in homology spheres. If K and J are height $h+2$ Whitney tower concordant, then K and J are h-solvably concordant.*

We recover the main result of [5] as a consequence.

Corollary 1.3 [5, Theorem 1.2] *Let $n \in \mathbb{N}$ and K be a link in a homology sphere. There is some link J in S^3 which is n-solvably concordant to K.*

The Grope filtration is at least as strong as the Whitney tower filtration, in that $\mathcal{G}_n \subseteq \mathcal{W}_n$ [17, Corollary 2], and is stronger than the solvable filtration in that $\mathcal{G}_n \subsetneq \mathcal{F}_{n+2}$ [15, Corollary 6.8]. We take a moment to explain what new idea allows our work to proceed that was not present in [5] and what prevents us from working in $\mathcal{G}_h$, the strongest of the three filtrations of [3]. The key tool in [5] used in the construction of the needed solvable cobordisms are n-solvable surgery curves [5, Definition 2.7]. These provide surgery instructions reducing a link in a homology sphere to a link in S^3 while preserving the n-solvable concordance class. Playing a similar role in this paper is a tool called the relative Whitney trick [6]. This tool attempts to use the double points in an immersed surface bounded by a link L to guide a homotopy of L to a new link bounding an embedded surface. The object used to guide this trick, called a relative Whitney tower, also allows for the construction of needed Whitney towers. Relative Whitney towers are detailed in Section 3.

The grope filtration is defined by setting $\mathcal{G}_h$ to be those links in S^3 which bound height h symmetric gropes in the 4-ball. This naturally extends to links in homology spheres by allowing gropes to sit in homology balls (or contractible 4-manifolds). It further extends to an equivalence relation by looking for annular gropes in a (strong) homology cobordism. See [2, Definition 2.14]. The obstacle that prevents us from working in this stronger setting is the lack of a suitable analogue to n-solvable surgery curves or relative Whitney disks. We close with a question.

Question 1.4 Does the grope filtration also fail to detect a difference between knots in homology spheres and knots in S^3? More precisely, for any $h \in \mathbb{N}$ and any link K in a homology sphere, is there is a link J in S^3 such that the components of K and J cobound a symmetric annular grope of height h in a (strong) homology cobordism?

Outline of paper

In Section 2 we give the formal definition of symmetric Whitney tower concordance and its relationship with the symmetric Whitney tower filtration of [3]. In Section 3 we translate the notion of a relative Whitney tower from [6] to the setting of symmetric relative Whitney towers and explain how use them to construct symmetric Whitney towers. In Section 4 we construct symmetric relative Whitney towers and complete the proof of Theorem 1.2. In Section 5 we recall the notion of solvable concordance and prove that Whitney tower concordance implies solvable concordance.

Acknowledgements

I would like to thank Taylor Martin and Carolyn Otto for a thorough reading of a preliminary draft of this manuscript and Mark Powell for suggesting the content of Section 5. I would also like to thank an anonymous referee for useful comments which improved the paper.

2 Symmetric Whitney towers

In order to make sense of Whitney disks and Whitney towers in a topological 4-manifold one must have notions like immersion, transversality, and tubular neighborhood. A careful exposition on these topics appears in [16, Section 3]. We give a very brief summary here. A continuous function $f : S \to W$ from a surface to a 4-manifold is called a *generic immersion* if it locally is a smooth generic immersion, meaning that there is an open cover $\{U_\alpha\}_{\alpha\in A}$ of W with homeomorphisms $\phi_\alpha : U_\alpha \to \mathbb{R}^4$ satisfying that for all α, the composition $\phi_\alpha \circ f|_{f^{-1}[U_\alpha]} : f^{-1}[U_\alpha] \to \mathbb{R}^4$ is a smooth generic immersion. Generic immersions admit linear normal bundles and so have tubular neighborhoods homeomorphic to a self plumbing of the total space of a vector bundle over S [8, Section 9.3]. From here on all immersions are assumed to be generic and all immersed surfaces are assumed to be transverse to each other and to the boundary of W.

Suppose that $p, q \in f[S]$ are double points of f with opposite sign. Then $f^{-1}\{p\} = \{p_1, p_2\}$ and $f^{-1}\{q\} = \{q_1, q_2\}$. For $i = 1, 2$, let α_i be a smoothly embedded arc in S running from p_i to q_i. Assume that other than its endpoints, $f(\alpha_i)$ is disjoint from all double points of f and that α_1 and α_2 are disjoint. The simple closed curve $\alpha := f(\alpha_1) * f(\overline{\alpha_2}) \subseteq W$ is called a *Whitney circle*. (Here $*$ indicates concatenation and $\overline{\alpha_2}$ is the reverse of α_2.) Assuming that α is nullhomotopic in W, there is an immersed disk $\Delta \subseteq W$ transverse to S bounded by α.

Being an immersed disk, Δ has a trivial normal bundle, N, admitting a unique trivialization $\psi : N \to D^2 \times \mathbb{R}^2$. In order for Δ to be called a Whitney disk, the restriction $\psi|_\alpha$ of ψ to $\alpha = \partial\Delta$ must agree with a trivialization coming from $f : S \to W$. See [2, Section 2]. In order to be explicit, we construct this trivialization. Let $U_1, U_2 \subseteq S$ be disjoint disk neighborhoods of α_1 and α_2 respectively. By taking U_1 and U_2 to be small enough, we arrange that $f|_{U_i} : U_i \to W$ is an embedding for $i = 1, 2$ and that $f[U_1] \cap f[U_2] = \{p_1, p_2\}$. At any point $x \in f(\alpha_i)$ let $\vec{u}_i$ be the tangent to $f(U_i)$ normal to α_i and $\vec{v}_i$ be normal to both Δ and $f(U_i)$. Since the points of intersection p and q have opposite signs, we may arrange that at p and q, $\vec{u}_1 = \vec{v}_2$ and $\vec{u}_2 = \vec{v}_1$. These vectors can be chosen to vary continuously and so give another trivialization τ of $N|_\alpha$. If $\tau = \psi|_\alpha$ then Δ is a *Whitney disk* for S pairing p and q.

Definition 2.1 [2, Definition 2.5; 3, Definition 7.7] Let S be a framed immersed surface in a 4-manifold. A *symmetric Whitney tower* T of height $h \in \mathbb{N}$ based on S is a sequence $T_0, T_1, \ldots, T_h$ of immersed surfaces such that

- $T_0 = S$,
- for $j \geq 1$, T_j is a collection of transverse framed immersed Whitney disks with disjoint boundaries pairing up all of the double points of T_{j-1}, and
- for $j > i$, T_j has interior disjoint from T_i.

We call T_i the i^{th} *stage of* T, and Whitney disks in T_i are called *height* i. If such a T exists then we say that S *extends to* T.

There is a more general notion simply called a Whitney tower. For a survey the reader is directed to [18]. All Whitney towers in this paper are symmetric. As a consequence any time we refer a Whitney tower T it should be understood that T is symmetric, even if we omit the word.

Remark 2.2 There is a discrepancy between the notion of height in [3] and [2]. In [3, Definition 7.7] the immersed surface S is called height 1, while in [2, Definition 2.5] it is height 0. The definition above follows the conventions of [2, Definition 2.5].

Remark 2.3 As is explored in [2, Remark 2.19], it is a consequence of results of Freedman and Quin on smoothing theory [8, Theorems 8.1A and 8.2A] that we could do equally well writing the definition above in the smooth setting. Indeed, suppose that $T \subseteq W$ is a Whitney tower in a 4-manifold. If W does not admit a smooth structure, then remove a point $p \notin T$ from W and appeal to [8, Theorems 8.2A] to see that $W \setminus \{p\}$ has a smooth structure. Briefly [8, Theorem 8.1 A] says that after changing T by a regular homotopy, which preserves the height of T, we may arrange that $T \subseteq W \setminus \{p\}$ is smooth. In the remainder of this paragraph, we fill in some of the details. We start with the lowest stages of T and work our way up. Indeed, suppose that we have already found a Whitney tower T of height h with the same boundary as T and for which $T_0 \cup \cdots \cup T_{i-1}$ is smoothly immersed. Let $S_i = T_i \setminus \nu(T_0 \cup \cdots \cup T_{i-1})$ be the (topologically) immersed surface in the smooth manifold $W \setminus (\{p\} \cup \nu(T_0 \cup \cdots \cup T_{i-1}))$ given by removing a neighborhood of the lower order stages. It has a tubular neighborhood $\nu(S_i)$ homeomorphic to a self plumbing of a 2-dimensional vector bundle over a surface (a disjoint union of disks if $i > 0$). Let X be such a self plumbing, $X_0 \subseteq X$ be its zero section and $h\colon (X, X_0) \to (\nu(S_i), S_i)$ be a homeomorphism. Being an open submanifold of a smooth manifold, $\nu(S_i) \subseteq W \setminus (\{p\} \cup \nu(T_0 \cup \cdots \cup T_{i-1}))$ is smooth. We now have that $h\colon X \to \nu(T_i)$ is a homeomorphism of smooth 4-manifolds and that $X_0 \subseteq X$ is a smooth 2-complex. By [8, Theorem 8.1A], we may change $X_0 \subset \nu(T)$ by a smooth regular homotopy and an ambient isotopy in X to get new 2-complex X_0' for which h is smooth on a small neighborhood of X_0'. Since regular homotopy consists of the Whitney move along embedded Whitney disks and its inverse, we may modify T by replacing S_i by $h[X_0']$ and arrive at a Whitney tower of the same height with $T_0, \ldots, T_i$ all smooth.

Definition 2.4 Let K be a link in a homology sphere Y and $\mathcal{B}$ be a contractible 4-manifold bounded by Y. We say that $K \in \mathcal{W}_h$ if there is a framed immersed union of disks in $\mathcal{B}$ bounded by K, with framing extending the 0-framing on K and which extends to a height $h-1$ Whitney tower.

This definition inspires an equivalence relation on the set of links in homology spheres. Instead of building a Whitney tower starting with a union of immersed disks, start with immersed annuli.

Definition 2.5 (compare with [2, Definition 2.12]) Let $h \in \mathbb{N}$, K and J be n-component links in homology spheres X and Y, and W be a strong homology cobordism from X to Y. For $i = 1, \ldots, n$ let A_i be an immersed annulus with $\partial A_i = K_i \cup -J_i$ and over which the 0-framing on $K_i \cup -J_i$ extends.

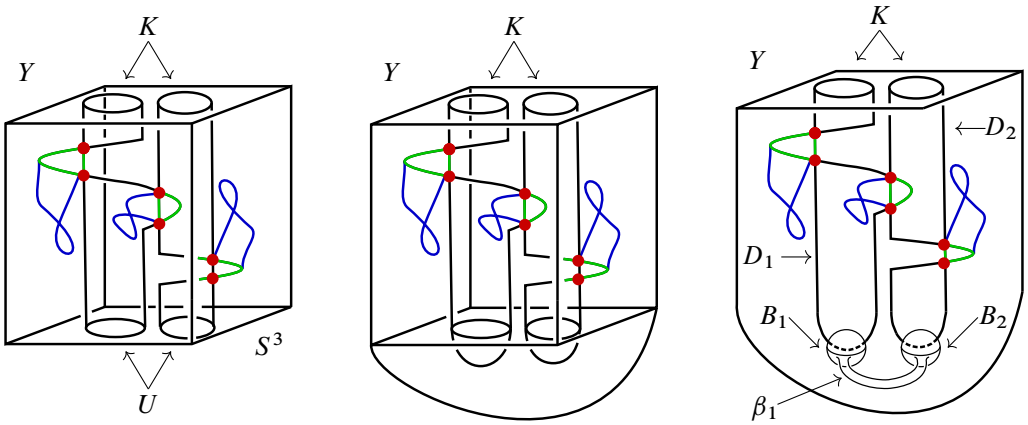


Figure 2: Left: An immersed union of annuli A bounded by $K \subseteq Y$ and $U \subseteq S^3$ extending to a height 1 Whitney tower. The curves in green represent Whitney circles. Middle: Capping A with disks bounded by U produces an immersed union of disks extending to a height 1 Whitney tower. Right: An immersed union of disks $D_1 \cup D_2$ bounded by a link K together with a pair of small 4-balls B_1 and B_2. Tube B_1 and B_2 together along the arc β_1 to get a single 4-ball B. Remove the interior of B to arrive at a Whitney tower concordance from K to the unlink.

If $A = A_1 \cup \cdots \cup A_n$ extends to a height $h-1$ symmetric Whitney tower T, then we say that K is *height h Whitney tower concordant* to J and write $K \sim_h J$. We call T a *height h Whitney tower concordance*.

In the above definition $-J$ indicates the reverse of J sitting in the orientation reverse of Y. We close this section by proving that the definitions of $\mathcal{W}_h$ and $\sim_h$ are compatible, in the sense that a link is in $\mathcal{W}_h$ if and only if it is $\sim_h$-equivalent to the unlink.

Proposition 2.6 *Let K be a link in a homology sphere and $h \in \mathbb{N}$. Then K is height h Whitney tower concordant to the unlink in S^3 if and only if $K \in \mathcal{W}_h$.*

Proof Suppose that $K \subseteq Y$ is a link in a homology sphere and $U \subseteq S^3$ is the unlink. Let W be a strong homology cobordism from Y to S^3 and $A \subseteq W$ be an immersed union of annuli bounded by K and U. Furthermore, assume that the 0-framing on $K \cup U$ extends over A and A extends to a height $h-1$ symmetric Whitney tower, T. A schematic appears in Figure 2, left. Let $\mathcal{B}$ be the contractible 4-manifold resulting from capping the S^3-boundary component of W with a 4-ball. It follows from the fact that W is a strong homology cobordism that $\mathcal{B}$ is a simply connected homology ball, and so is contractible. Let $D \subseteq B^4$ be a disjoint union of smoothly embedded disks bounded by U. Notice that A and D both induce the 0-framing on U so that $A \cup D \subseteq \mathcal{B}$ is an immersed union of disks in a contractible 4-manifold over which the 0-framing of K extends and which forms the base surface of a height $h-1$ symmetric Whitney tower, $T \cup D$. Thus, $K \in \mathcal{W}_h$. A schematic of $A \cup D \subseteq \mathcal{B}$ appears in Figure 2, middle.

Conversely, suppose that $\mathcal{B}$ is a contractible 4-manifold bounded by Y, that $D \subseteq \mathcal{B}$ is a framed immersed union of disks bounded by the 0-framing of K and that D extends to a height $h-1$ symmetric Whitney

tower, T in $\mathcal{B}$. For any i, let K_i be the i^{th} component of K and $D_i \subseteq D$ be the immersed disk bounded by K_i. Pick a point p_i interior to D_i, disjoint from all of the Whitney disks in T and a small closed 4-ball neighborhood $B_i \subseteq \mathcal{B}$ of p_i. By taking B_i small enough, we arrange that $\partial B_i \cap D_i$ is an unknot and that $B_i \cap D_i = B_i \cap T$ is an embedded disk bounded by this unknot. For $i = 1, \dots, n-1$, let β_i be an embedded arc in $\mathcal{B}$ running from B_i to B_{i+1}. A dimensionality argument allows us to arrange that β_i is disjoint from T. Let $B \subseteq \mathcal{B}$ be the 4-ball given by tubing together $B_1, \dots, B_n$ along $\beta_1, \dots, \beta_{n-1}$. See Figure 2, right, for a schematic.

We now see a strong homology cobordism $\mathcal{B} \setminus \operatorname{int}(B)$ from Y to S^3, in which $A = D \setminus \operatorname{int}(B)$ is an immersed union of annuli bounded by K and U. Here $\operatorname{int}(B)$ indicates the interior of B. That W is a strong homology cobordism follows from the fact that $\mathcal{B}$ is contractible. We restrict the framing on D to $D \cap B$. Since $D \cap B$ is a collection of slice disks for U, this framing restricts to the 0-framing on U. Thus, the 0-framing on $K \cup U$ extends over A. Finally A is the base surface for a height $h-1$ symmetric Whitney tower, $T \setminus \operatorname{int}(B)$. Thus, K is height h symmetric Whitney tower concordant to the unlink in S^3. □

3 Symmetric relative Whitney towers

Symmetric Whitney towers are a restriction of a larger class of immersed 2-complexes in 4-manifolds, simply called Whitney towers. In [6, Section 5, immediately before Remark 5.3] the author along with Orson and Park define a variation of a Whitney tower called a relative Whitney tower. Relative Whitney towers have two important properties. First, they exist much more readily than actual Whitney towers [6, Lemma 5.4]. Second, they can be used to guide a homotopy of an immersed surface which changes the boundary in a controlled way to a new immersed surface which extends to a Whitney tower [6, Lemma 5.5]. In order to prove Theorem 1.2 we extend these two results to the setting of symmetric Whitney towers.

Let $\psi : S \to W$ be an immersion of a surface (possibly with corners) into a 4-manifold. Assume that ψ is transverse to the boundary so that $\psi^{-1}[\partial W] \subseteq \partial S$ is a disjoint union of arcs and circles. Let p be a double point in $\psi(S)$ with preimage $\psi^{-1}\{p\} = \{p_1, p_2\}$. Let $\alpha_1, \alpha_2 : [0,1] \hookrightarrow S$ be disjoint embedded arcs in S running from p_1 and p_2 to points q_1 and q_2 interior to $\psi^{-1}(\partial W)$, α_1 and α_2 interior to S except for the endpoints at q_1 and q_2. Let α_3 be an arc in ∂W running from $\psi(q_1)$ to $\psi(q_2)$, which is otherwise disjoint from $\psi[S]$. The concatenation $\psi(\alpha_1) * \alpha_3 * \psi(\overline{\alpha_2})$ is a simple closed curve in W. If this curve bounds an immersed disk Δ transverse to S then Δ is called a *relative Whitney disk* associated to the double point p. The arc α_3 is called the *relative Whitney arc* of Δ. A schematic appears in Figure 3.

The reader will notice that while the definition of a Whitney disk requires some discussion of trivialization of normal bundles, the definition of a relative Whitney disk does not. This is because a trivialization of a vector bundle on an arc on the boundary of a of a disk automatically extends to a trivialization over the disk. Thus, if N is the normal bundle of Δ and $N|_{\psi(\overline{\alpha_2})*\psi(\alpha_1)}$ is its restriction to $\psi(\overline{\alpha_2}) * \psi(\alpha_1)$ then any trivialization of $N|_{\psi(\overline{\alpha_2})*\psi(\alpha_1)}$ automatically extends over N. We construct a trivialization of

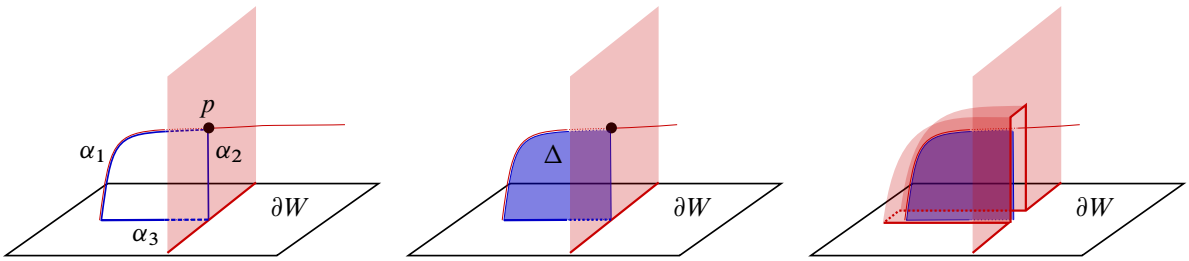


Figure 3: Left: A double point p in an immersed surface together with arcs α_1 and α_2 running from p to points in $\partial W \cap S$ and a third arc α_3 connecting the endpoints of α_1 and α_2. Middle: $\alpha_1 * \alpha_3 * \overline{\alpha_2}$ bounding an immersed disk, called a relative Whitney disk. Right: The relative Whitney trick along Δ removes the double point at p.

$N|_{\psi(\overline{\alpha_2})*\psi(\alpha_1)}$ as follows. At any point in $\psi(\alpha_i)$ let $\vec{u}_i$ be a tangent vector to $\psi(S)$ which is normal to $\psi(\alpha_i)$ and let $\vec{v}_i$ be normal to Δ and to $\psi(S)$. These may be chosen to vary continuously and so that at the double point p, $\vec{u}_1 = \vec{v}_2$ and $\vec{v}_2 = \vec{u}_1$. These two normal vectors together give a trivialization of $N|_{\psi(\overline{\alpha_2})*\psi(\alpha_1)}$ which we extend to give a framing of Δ. Notice that this imposes a framing on the relative Whitney arc, α_3. See [6, Section 2.2] for a more detailed discussion.

Definition 3.1 Let S be an immersed surface in a 4-manifold. A *symmetric relative Whitney tower* of height $h \in \mathbb{N}$ based on S is a sequence $T_0, T_1, \dots, T_h$ of immersed surfaces such that

- $T_0 = S$,
- for $j \geq 1$, T_j is a collection of transverse relative Whitney disks with disjoint boundaries consisting of one relative Whitney disk associated to each double point in T_{j-1}, and
- for $j > i$, T_j has interior disjoint from T_i.

We call T_i the i^{th} stage of T, and relative Whitney disks in T_i are called height i.

An important application of relative Whitney disks is the so-called relative Whitney trick, schematically described in Figure 3, right, and described in detail in [6, Subsection 2.2]. Another is the *partial relative Whitney trick* of Figure 4, right, [6, Proof of Lemma 5.5]. We give the properties of the partial relative Whitney trick. Let p be a double point in an immersed surface A. Let Δ be a relative Whitney disk associated with p and $\alpha \subseteq \partial W$ be its relative Whitney arc. The partial relative Whitney trick

- changes A by a homotopy which is constant away from a small neighborhood of α,
- introduces exactly one new double point q to A,
- replaces Δ with a Whitney disk pairing the intersection points p and q, and
- changes ∂A by a homotopy pushing an arc in ∂A along the framed relative Whitney arc over a meridian of ∂A, as in Figure 4, middle.

Let T be a symmetric relative Whitney tower. By performing the partial relative Whitney move on all of the relative Whitney disks of T one gets a symmetric Whitney tower.

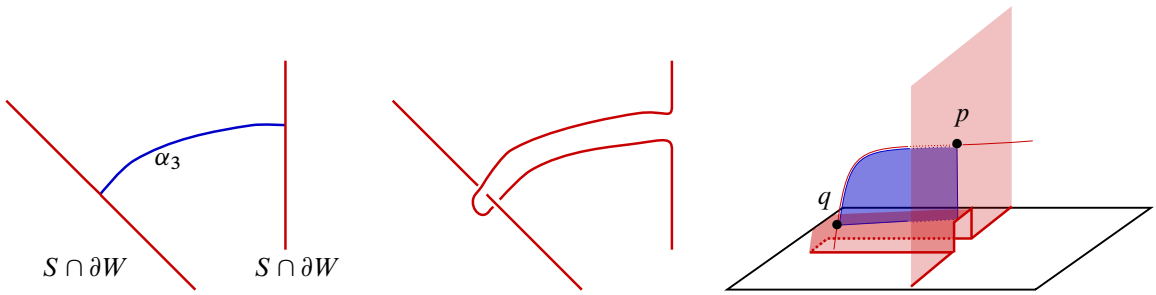


Figure 4: Left: Two arcs in $S \cap \partial W$ connected by the relative Whitney arc α_3 of a relative Whitney disk. Middle: The relative Whitney move changes $S \cap \partial W$ by sliding along α_3 over a meridian. Right: The partial relative Whitney trick introduces a new point of intersection and replaces a relative Whitney disk with a Whitney disk.

Lemma 3.2 (compare to [6, Lemma 5.5]) *Let W be a 4-manifold and T be a height h symmetric relative Whitney tower in W. If Y is a connected submanifold of ∂W and contains all of the relative Whitney arcs of T, then there exists a symmetric relative Whitney tower, T' of height h such that the height 0 surfaces of T and T' differ by a homotopy which is constant outside of a small neighborhood of Y.*

Proof As the proof is identical to that of [6, Lemma 5.5], we give only a summary. Start with a symmetric relative Whitney tower T of height h. Perform the partial relative Whitney move along each of the height h relative Whitney disks in T. These moves change T_{h-1} by a homotopy to T'_{h-1} (a new collection of relative Whitney disks if $h-1>0$). We now have $T_0 \cup T_1 \cup \cdots \cup T_{h-2} \cup T'_{h-1}$ is a height $h-1$ relative Whitney tower and T'_{h-1} extends to a height 1 symmetric Whitney tower.

Now do the same along all height $h-1$ relative Whitney disks, so that $T_0 \cup T_1 \cup \cdots \cup T_{h-3} \cup T'_{h-2}$ is a height $h-2$ relative Whitney tower and T'_{h-2} extends to a height 2 Whitney tower. Iterate until you have performed the relative Whitney trick along every relative Whitney disk. What results is a height 0 relative Whitney tower (ie an immersed surface) T'_0 which extends to a height h Whitney tower. Moreover T'_0 differs from T_0 by a homotopy supported in a small neighborhood of the relative Whitney arcs of T. As Y contains all of the relative Whitney arcs, this completes the proof. □

4 Constructing symmetric relative Whitney towers: proof of Theorem 1.2

Let S be an immersed surface and Δ be a Whitney disk pairing two double points of S. If Δ is embedded and has interior disjoint from S then the Whitney trick can be used to remove these points of intersection. The *finger move* is an inverse to the Whitney trick. Let $S \subseteq W$ be an immersed surface in a 4-manifold. Let β be framed arc running from a point interior to S to another point interior to S. A schematic for the result of changing S by a finger move along β is depicted in Figure 5. In the result of the finger move

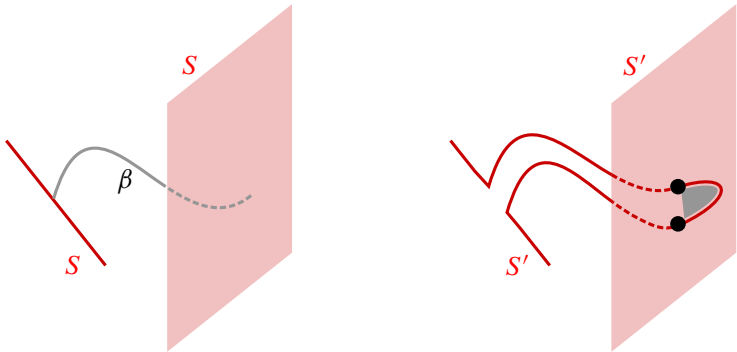


Figure 5: Left: An arc $\beta \subseteq W$ between two points in immersed surface $S \subseteq W$. Right: The result of the finger move β along with an embedded Whitney disk which undoes the finger move.

there are two new intersection points paired by an embedded Whitney disk. Performing the Whitney move along this disk results in a surface isotopic to S. See also [1, Section 11.2].

Lemma 3.2 gives a path to the proof of Theorem 1.2. Let K be a link in a homology sphere Y. By [8, Corollary 9.3] Y bounds a simply connected homology ball. By removing a small open 4-ball from its interior one gets a simply connected homology cobordism W from Y to S^3. Since W is simply connected, there is a free homotopy (a union of immersed annuli) in W from K to the unlink U in S^3. Call this immersed union of annuli A. If we can arrange that A extends to a height h symmetric relative Whitney tower, then we will be able to use Lemma 3.2 to replace it with a Whitney tower concordance from K to some link in S^3. The first step is arranging that $\pi_1(S^3 \setminus \nu(U)) \twoheadrightarrow \pi_1(W \setminus \nu(A))$ is surjective.

Lemma 4.1 *Let W be a 4-manifold with boundary. Let S be an immersed surface in W. Let $Y \subseteq \partial W$ be a connected 3-manifold. Assume that every component of S has nonempty intersection with Y and that the inclusion induced map $\pi_1(Y) \twoheadrightarrow \pi_1(W)$ is onto.*

There is an immersed surface S' in W satisfying that the inclusion induced map $\pi_1(Y \setminus \nu(S')) \twoheadrightarrow \pi_1(W \setminus \nu(S'))$ is onto and that S' differs from S by sequence of finger moves. In particular $\partial S = \partial S'$.

Proof Let W, S and Y be as in the lemma. Pick a base point y in $Y \setminus \nu(S)$. Let $\{g_1, \dots, g_k\}$ be a generating set for $\pi_1(W \setminus \nu(S), y)$. Since $\pi_1(Y, y) \twoheadrightarrow \pi_1(W, y)$ is surjective, for each i, there is some $g_i' \in \pi_1(Y, y)$ and an immersed disk E_i bounded by $g_i * \overline{g_i'}$. A dimensionality argument allows us to arrange that g_i' is disjoint from ∂S. If E_i is disjoint from S, then g_i is already in the image of $\pi_1(Y \setminus \nu(S)) \to \pi_1(W \setminus \nu(S))$. Otherwise, let p be a transverse point of intersection between E_i and S. Let α_1 and α_2 be arcs in S and E_i from p to points in Y. Since $\pi_1(Y, y) \twoheadrightarrow \pi_1(W, y)$, there is an arc α_3 in Y so that $\alpha_1 * \alpha_3 * \overline{\alpha_2}$ is nullhomotopic and so bounds a relative Whitney disk, Δ associated with p.

We would like to use the relative Whitney trick to eliminate the point of intersection at p, but Δ might be neither embedded nor disjoint from S. Let r be a transverse point of intersection in $\Delta \cap S$. Find an arc β

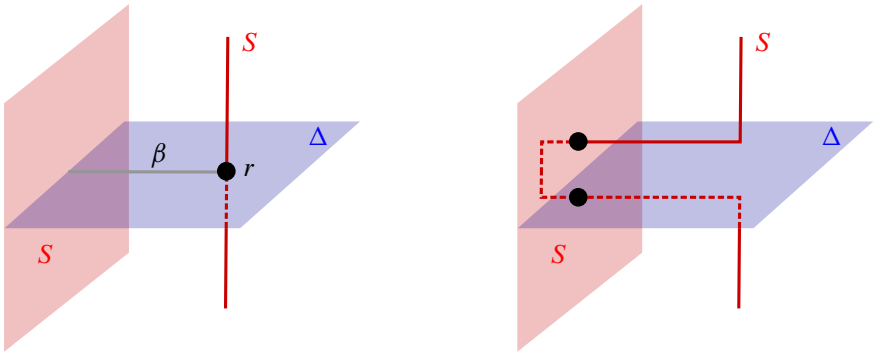


Figure 6: Left: A point q in the intersection of a surface S with a Whitney disk Δ associated to a point in $E_i \cap S$ together with an arc $\beta \subseteq \Delta$ from r to a point in $S \cap \partial\Delta$. Right: A finger move reduces the number of points in $S \cap \Delta$ at a cost of two new double points in S.

in Δ from r to a point interior to $\alpha_1 \subseteq S$. As in Figure 6 we modify S by a finger move along β in order to reduce the number of points in $\Delta \cap S$ by 1. This move adds two new double points to S. As is briefly explained in [13, Chapter XII Section 2], the finger move changes $\pi_1(W - \nu(S))$ by killing a commutator of two meridians of S. In particular, $g_1, \dots, g_n$ is still a generating set for $\pi_1(W - \nu(S))$. Iterating, we arrange that $\Delta \cap S = \varnothing$. Finally, we modify E_i by the relative Whitney trick along Δ. This eliminates the point of intersection $p \in S \cap E_i$. It may add double points to E_i and will change g_i'. Regardless, we now have that g_i is in the image of $\pi_1(Y \setminus \nu(S)) \to \pi_1(W \setminus \nu(S), y)$. By following the steps in this paragraph for every $i = 1, \dots, n$ we arrange that $\pi_1(Y \setminus \nu(S)) \twoheadrightarrow \pi_1(W \setminus \nu(S), y)$ is surjective. Notice that this process changes S by some finger moves and does not change ∂S. □

For any link K in any homology sphere Y we can find a simply connected homology cobordism W and an immersed union of annuli $A \subseteq W$ bounded by K and the unlink in S^3. Appealing to Lemma 4.1, we arrange that $\pi_1(S^3 \setminus \nu(U)) \twoheadrightarrow \pi_1(W \setminus \nu(A))$. An iterative application of the following lemma allows us to extend A to a symmetric relative Whitney tower of arbitrary height. We delay its proof until the end of the section.

Lemma 4.2 (compare to [6, Lemma 5.4]) *Let W be a 4-manifold with boundary. Let $T \subseteq W$ be a symmetric relative Whitney tower of height h. Let $Y \subseteq \partial W$ be a connected 3-manifold which contains all of the relative Whitney arcs of T. Assume that every component of T_0 has nonempty intersection with Y and that the inclusion induced map $\pi_1(Y \setminus \nu(T)) \twoheadrightarrow \pi_1(W \setminus \nu(T))$ is onto.*

Then T extends to a symmetric relative Whitney tower T' of height $h+1$ such that Y contains all of the relative Whitney arcs of T' and such that the inclusion induced map $\pi_1(Y \setminus \nu(T')) \twoheadrightarrow \pi_1(W \setminus \nu(T'))$ is onto.

Between Lemmas 4.1, 4.2, and 3.2 we can prove Theorem 1.2. We recall its statement.

Theorem 1.2 *Let $h \in \mathbb{N}$ and K be a link in a homology sphere. There is some link in S^3 which is height h Whitney tower concordant to K.*



Figure 7: Left: A double point in a framed surface S. Right: A pushoff S^+ intersects the original surface in two points with the same sign as that double point.

Proof Let K be a link in a homology sphere X. By [8, Corollary 9.3] X bounds a simply connected homology ball. Removing an open 4-ball from the interior, we see a simply connected homology cobordism W from X to S^3.

Let A be an immersed union of annuli bounded K and the unlink in S^3. This union of annuli exists since W is simply connected. By Lemma 4.1 we arrange that $\pi_1(S^3 \setminus \nu(A)) \twoheadrightarrow \pi_1(W \setminus \nu(A))$ is surjective. Since a height 0 symmetric relative Whitney tower is just an immersed surface, the assumptions of Lemma 4.2 are satisfied by $T = A$, $Y = S^3$, and $h = 0$. The tower produced by Lemma 4.2 satisfies all of the assumptions of Lemma 4.2 and so we can appeal to the lemma again. Iterating, we see that A extends to a height $h-1$ symmetric relative Whitney tower in W and that all of its relative Whitney arcs are contained in $S^3 \subseteq W$. Finally, by Lemma 3.2 there is a height $h-1$ symmetric Whitney tower whose base surface is an annulus bounded by K together with some link J in S^3.

It remains only to arrange that there is a trivialization of the normal bundle of A which restricts to the 0-framings on K and J. We do so one component at a time. First extend the 0-framing on K_i over A_i (the component of A bounded by $K_i \cup -J_i$). This can be done since an annulus deformation retracts to its either boundary component. This restricts to $-J_i$ to give the f_i-framing for some $f_i \in \mathbb{Z}$. Let $F_i \subseteq Y$ be a Seifert surface for K_i. Cap the Y-boundary component of W with a contractible 4-manifold. Call the result $\mathcal{B}$. Since $\mathcal{B}$ is a contractible 4-manifold bounded by S^3, $\mathcal{B}$ is homeomorphic to the 4-ball by the topological 4-dimensional Poincaré conjecture. Since the 0-framing of K_i extends over both A_i and F_i, $S_i = A_i \cap F_i$ is a framed surface in the 4-ball which restricts to the f_i-framing of $-J_i$. Then a framed pushoff S_i^+ is bounded by $(-J_i)^+$, the f_i-framed pushoff of $-J_i$.

Each double point of S_i results in two points of intersection with the same sign between S_i and S_i^+, as in Figure 7. We may as well assume that $h-1 \geq 1$, so that all of the self intersections of S_i are paired by Whitney disks, and in particular the signed count of the number of intersections between S_i and S_i^+ is zero. Finally since linking numbers can be computed as the signed count of the number of intersections between immersed surfaces in the 4-ball,

$$f_i = \mathrm{lk}(-J_i, (-J_i)^+) = S_i \cdot S_i^+ = 0.$$

Therefore, the 0-framings on K and J each extend over A, and A extends to a height $h-1$ Whitney tower. Thus, $K \sim_h J$ completing the proof □

Finally, we prove Lemma 4.2 by appealing to Lemma 4.1.

Proof of Lemma 4.2 Let W, T, and Y be as in the statement. Recall that T_h is the height h part of T. Let $p^1, p^2, \dots, p^k$ be the double points in T_h. Let $\psi\colon S \to W$ is the immersion of a surface parametrizing T_h. For any $j = 1, \dots, k$, let $\psi^{-1}\{p^j\} = \{p_1^j, p_2^j\}$. There are embedded arcs α_1^j and α_2^j in S running from p_1^j and p_2^j to points q_1^j and q_2^j interior to $\psi^{-1}(Y)$. We will further arrange that $\alpha_1^1, \alpha_2^1, \dots, \alpha_1^k, \alpha_2^k$ are all disjoint from each other. Such arcs exist when $h = 0$ because Y has nonempty intersection with each component of T_0 and when $h > 0$ because Y contains every relative Whitney arc. Since $\pi_1(Y \setminus \nu(T)) \twoheadrightarrow \pi_1(W \setminus \nu(T))$, for $j = 1, \dots, k$, there is some arc $\alpha_3^j \subseteq Y$ running between $\psi(q_1^j)$ and $\psi(q_2^j)$ so that $\psi(\alpha_1^j) * \alpha_3^j * \psi(\alpha_2^j)$ bounds an immersed disk Δ^j with interior disjoint from T. Notice that $T \cup (\Delta^1 \cup \cdots \cup \Delta^k)$ is now a height $h+1$ relative Whitney tower.

The assumptions of Lemma 4.1 are satisfied by $W' = W \setminus \nu(T)$, $Y' = Y \setminus \nu(T) \subseteq \partial W'$ and the immersed surface $S' = \Delta^1 \cup \cdots \cup \Delta^k \subseteq W'$. Thus, we can change $\Delta^1 \cup \cdots \cup \Delta^k$ by a sequence of finger moves in $W \setminus \nu(T)$ so that $\pi_1(Y \setminus \nu(T \cup \Delta^1 \cup \cdots \cup \Delta^k)) \twoheadrightarrow \pi_1(W \setminus \nu(T \cup \Delta^1 \cup \cdots \cup \Delta^k))$ is surjective, completing the proof. □

5 Whitney tower concordance and solvable concordance

For any link $K \subseteq S^3$, if $K \in \mathcal{W}_{h+2}$ then K is h-solvable [3, Theorem 8.12]. In this section we show that if two links in homology spheres are height $h+2$-Whitney tower concordant then they are h-solvably concordant. We begin by recalling the definition of h-solvable concordance.

Definition 5.1 [5, Definition 2.3] Let $K \subseteq X$ and $J \subseteq Y$ be links in homology spheres. We say that K is *n-solvably concordant* to J if there exists a cobordism W from X to Y in which the components of K and J cobound a disjoint collection of embedded locally flat annuli $A \subseteq W$ such that:

(1) $H_1(W) = 0$, so that W is an H_1-cobordism.

(2) There exist locally flat embedded closed oriented surfaces in W with trivial normal bundles $L_1, D_1, \dots, L_k, D_k$ all disjoint from C and from each other except that for each $i = 1, \dots, k$, the surfaces L_i and D_i intersect transversely in a single point.

(3) $\{[L_1], [D_1], \dots, [L_k], [D_k]\}$ generates $H_2(W)$.

(4) For all $i = 1, \dots, k$, the images of $\pi_1(L_i) \to \pi_1(E(A))$ and $\pi_1(D_i) \to \pi_1(E(A))$ are both contained in $\pi_1(E(A))^{(n)}$.

Here $E(A) = W \setminus \nu(A)$ is the exterior of A and $\pi_1(E(A))^{(n)}$ refers to the *derived series* of the group $\pi_1(E(A))$. It is defined recursively as

$$\pi_1(E(A))^{(0)} = \pi_1(E(A)) \quad \text{and} \quad \pi_1(E(A))^{(n+1)} = [\pi_1(E(A))^{(n)}, \pi_1(E(A))^{(n)}].$$



Figure 8: A pair of double points of opposite sign in an immersed surface may be canceled at the cost of increasing the genus of the surface by 1.

Remark 5.2 Similar to Remark 2.3, one gets an equivalent definition by requiring these surfaces to be smooth, either with regard to a smooth structure on $W \setminus \nu(A)$, if one exists, or a smooth structure on $W \setminus (\nu(A) \cup \{p\})$. Indeed by [8, Theorem 8.1A] $L_1 \cup D_1 \cup \cdots \cup L_k \cup D_k$ admits a regular homotopy inside of a small neighborhood to a smoothly immersed surface. Since a regular homotopy preserves intersection numbers, L_i still has algebraically zero self intersections, and so by increasing the genus as in Figure 8 we can arrange that the geometric intersections between $L_1 \cup D_1 \cup \cdots \cup L_k \cup D_k$ are preserved. As this occurs in a small neighborhood of $L_1 \cup D_1 \cup \cdots \cup L_k \cup D_k$, the image of $\pi_1(L_i) \to \pi_1(E(A))$ and $\pi_1(D_i) \to \pi_1(E(A))$ is still contained in $\pi_1(E(A))^{(n)}$.

Remark 5.3 The remaining result of this paper, Proposition 5.4, requires the construction of an h-solvable cobordism from a height $h+2$ symmetric Whitney tower concordance. We will need a strategy to demonstrate that certain elements of a fundamental group sit deep in the derived series of some fundamental group. We will use the following geometric translation of derived series of the fundamental group in the language of gropes. A height 1 symmetric grope is a compact oriented surface. A height $h+1$ symmetric grope consists of a surface together with a symplectic basis $\{a_1, b_1, \ldots, a_g, b_g\}$ and $2g$ many height h gropes bounded by $a_1, b_1, \ldots, a_g, b_g$. Generally in the literature, these are taken to be embedded in a 4-manifold and have some framing condition, but these conditions are not relevant to our work. For any 4-manifold X, a straightforward inductive argument reveals that $\gamma \in \pi_1(X)^{(h)}$ if and only if γ bounds an immersed height h grope in X.

Proposition 5.4 *Let K and J be links in homology spheres. If K and J are height $h+2$ Whitney tower concordant, then K and J are h-solvably concordant.*

Proof Let K and J be links in homology spheres X and Y, respectively. Let W be a homology cobordism from X to Y in which the components of K and J cobound an immersed union of annuli A which extends to a height $h+1$ symmetric Whitney tower T.

Let $\Delta_1, \ldots, \Delta_k$ be a complete list of all of the height $h+1$ Whitney disks in T_{h+1}. For every $j = 1, \ldots, k$, let γ_j be the framed curve given by pushing $\partial\Delta_j$ slightly into the interior of Δ_j. Let W' be the result of modifying W by surgery along $\gamma_1, \ldots, \gamma_k$.

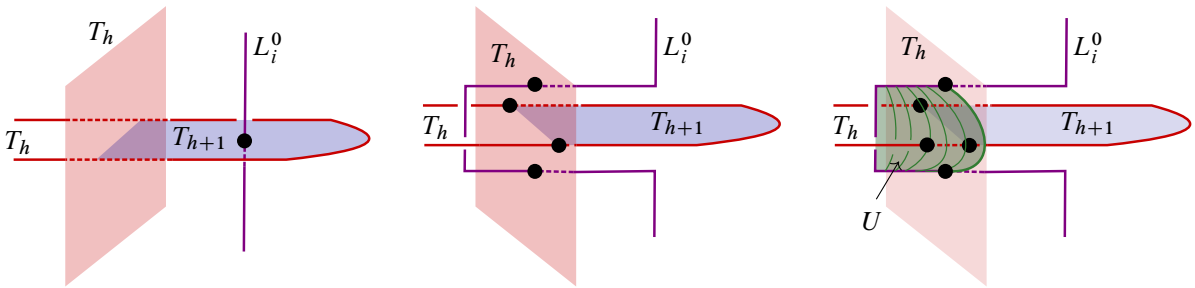


Figure 9: Left: L_i^0 intersects T_{h+1} transversely in a single point. Middle: After a finger move L_i^0 (purple) intersects T_h (red) in two points with opposite sign. Right: An embedded Whitney disk U pairing these two points of intersection and which intersects T_h transversely in one point.

First we exhibit a disjoint union of embedded annuli in W' cobounded by the components of K and the components of J. Let $T' \subseteq W'$ be the height $h+1$ Whitney tower given starting with T and replacing each $\Delta_j \subseteq T_{h+1}$ with the embedded disk glued in along γ_j. Notice that all of the height $h+1$ Whitney disks in T' are embedded and have interiors disjoint from all other surfaces in T'. Thus, we may use the Whitney trick to eliminate all of the intersections in T'_h. Next use these embedded height h Whitney disks to eliminate intersections in T'_{h-1}. Iterate until T'_0 is embedded. We now have that in W' the components of K and J cobound an embedded union of disjoint annuli, which we will call A. Observe that A is contained in a small neighborhood of T'. The remainder of the proof amounts to showing that the conditions of Definition 5.1 are satisfied.

As modifying a 4-manifold by surgery along nullhomologous simple closed curves does not change first homology, $H_1(W') = 0$. Therefore W' is an H_1-cobordism. Each time we perform surgery we increase the rank of second homology by 2. Let L_i^0 be a meridional 2-sphere for γ_i (an embedded sphere bounding a 3-ball in W which intersects γ_i transversely in a single point). Let $D_i^0 \subseteq W'$ be the immersed sphere given by capping the disk in Δ_i bounded by γ_i with the disk glued to the surgery curve γ. We now have that $\{[L_1^0], [D_1^0], \dots, [L_k^0], [D_k^0]\}$ forms a basis for $H_2(W')$. Being the boundary of an embedded 3-sphere in W, L_i^0 has trivial normal bundle. Since the framing on γ_i used to perform surgery was the framing induced by Δ_i, D_i^0 has trivial normal bundle.

Our next goal is to replace L_i^0 with an embedded surface homologous to L_i^0 which extends to a height $h+1$ grope in the exterior of T'. Currently, L_i^0 intersects T'_{h+1} transversely in a single point. By performing a finger move we isotope L_i^0 to be disjoint from T'_{h+1} and to instead intersect T'_h in two points of opposite sign. As in Figure 9, these two points of intersection are paired with an embedded Whitney disk U disjoint from T_{h+1} whose interior intersects T_h transversely in a single point. As in Figure 10 we tube L_i^0 to itself along the arc $\partial U \cap T_h$ to get an embedded genus-1 surface L_i disjoint from T. Also in Figure 10 we see a pair of disks U and V bounded by a symplectic basis for L_i. These disks each intersect T_h transversely in a single point and $L_i \cup U \cup V$ is a height 2 grope.

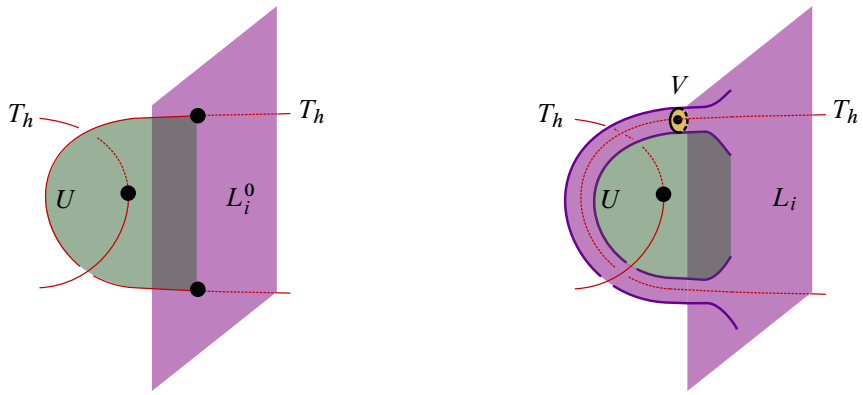


Figure 10: Left: Two points of intersection between L_i^0 and T_h which are paired by an embedded Whitney disk U intersecting T_h in one point. Right: Tubing produces a surface of genus 1 with a symplectic basis bounding embedded disks U and V each of which intersects T_h in a single point.

Next we apply exactly the same steps as the preceding paragraph to U and V this time instead of L^0. First, modify U and V by a finger move so that they are disjoint from T_h and each intersect T_{h-1} in two points paired with a Whitney disk. Then add a tube to each to replace U and V by genus-1 surfaces which admit symplectic bases bounding disks that each intersect T_{h-1} in a single point. The genus-1 surface L_i now forms the height 1 part of a height 3 grope G which is disjoint from T' except that the height 3 part of G consists of disks each intersecting the T'_{h-1} in a single point.

For any $k \in \{1,\dots,h+2\}$ we iterate this construction to produce a height k grope G_i with height 1 surface L_i, whose height k part consists of disjoint embedded disks each intersecting T'_{h+2-k} transversely in a single point, and which is otherwise disjoint from T'. Setting $k = h+2$ and dropping the disks that make up the height $h+2$ part of G, we see that L_i forms the height 1 part of a height $h+1$ grope in the exterior of T'. It follows that this grope is disjoint from A since A is contained in a neighborhood of T'. As the sphere L_i^0 has trivial normal bundle so does L_i. As L_i forms the height 1 part of a height $h+1$ grope disjoint from A, there is a generating set for $\pi_1(L_i)$ consisting of curves that bound height h gropes disjoint from A. We conclude using Remark 5.3 the image of $\pi_1(L_i) \to \pi_1(E(A))$ is contained in $\pi_1(E(A))^{(h)}$.

Next consider D_i^0, the immersed sphere formed from the union of Δ_i with the disk attached to γ_i. As the framing on γ_i used to perform surgery is the same as that coming from the normal bundle of Δ_i, D_i^0 has trivial normal bundle. These spheres are not embedded, nor are they disjoint from each other, but they are disjoint from T' and they are disjoint from $L_1 \cup \cdots \cup L_k$ except that D_i^0 intersects L_i transversely in a single point.

For each $i = 1,\dots,k$ and each point of self intersection of D_i^0, tube D_i^0 to a pushed off copy of L_i. This increases the genus of D_i^0 by 1, and reduces the number of self intersections by 1. The homology class $[D_i^0]$ is replaced by $[D_i^0] \pm [L_i]$ depending on the sign of the intersection point removed. In particular

$\{[L_1], [D_1^0], \dots, [L_k], [D_k^0]\}$ still generates $H_2(W')$. Since the result of tubing together framed surfaces is still framed, D_i^0 is still framed. By making this replacement at each point of self intersection, we arrange that D_i^0 is embedded. Similarly, we eliminate all points of intersection between any D_i^0 and D_j^0 by tubing D_i^0 into a pushed off copy of L_j. Let $D_1, \dots, D_k$ be the resulting surfaces. Each D_i consists of a union of pushoffs of $L_1, \dots, L_k$ together with a planar surface. As a consequence, the image of $\pi_1(D_i) \to \pi_1(E(A))$ is contained in the normal closure of the subgroup generated by the images of $\pi_1(L_1), \dots \pi_1(L_k)$. As we have already explained, this is contained in $\pi_1(E(A))^{(h)}$.

Thus, $L_1, D_1, \dots, L_k, D_k$ satisfy conditions (2), (3), and (4) of Definition 5.1. Therefore K and J are h-solvably concordant, completing the proof. □

References

[1] **S Behrens**, **B Kalmár**, **M H Kim**, **M Powell**, **A Ray** (editors), *The disc embedding theorem*, Oxford Univ. Press (2021) MR

[2] **J C Cha**, *Symmetric Whitney tower cobordism for bordered 3-manifolds and links*, Trans. Amer. Math. Soc. 366 (2014) 3241–3273 MR

[3] **T D Cochran**, **K E Orr**, **P Teichner**, *Knot concordance, Whitney towers and L^2-signatures*, Ann. of Math. 157 (2003) 433–519 MR

[4] **J Conant**, **R Schneiderman**, **P Teichner**, *Whitney tower concordance of classical links*, Geom. Topol. 16 (2012) 1419–1479 MR

[5] **C W Davis**, *Topological concordance of knots in homology spheres and the solvable filtration*, J. Topol. 13 (2020) 343–355 MR

[6] **C W Davis**, **P Orson**, **J Park**, *The relative Whitney trick and its applications*, Selecta Math. 28 (2022) art. id. 27 MR

[7] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR

[8] **M H Freedman**, **F Quinn**, *Topology of 4-manifolds*, Princeton Math. Ser. 39, Princeton Univ. Press (1990) MR

[9] **R E Gompf**, *An infinite set of exotic $\mathbb{R}^4$'s*, J. Differential Geom. 21 (1985) 283–300 MR

[10] **F J González Acuña**, *On homology spheres*, PhD thesis, Princeton University (1970) Available at `https://www.proquest.com/docview/302426826`

[11] **J Hom**, **A S Levine**, **T Lidman**, *Knot concordance in homology cobordisms*, Duke Math. J. 171 (2022) 3089–3131 MR

[12] **M A Kervaire**, **J W Milnor**, *Groups of homotopy spheres, I*, Ann. of Math. 77 (1963) 504–537 MR

[13] **R C Kirby**, *The topology of 4-manifolds*, Lecture Notes in Math. 1374, Springer (1989) MR

[14] **A S Levine**, *Nonsurjective satellite operators and piecewise-linear concordance*, Forum Math. Sigma 4 (2016) art. id. e34 MR

[15] **C Otto**, *The (n)-solvable filtration of link concordance and Milnor's invariants*, Algebr. Geom. Topol. 14 (2014) 2627–2654 MR

[16] **M Powell**, **A Ray**, **P Teichner**, *The* 4*-dimensional disc embedding theorem and dual spheres*, preprint (2020) arXiv 2006.05209

[17] **R Schneiderman**, *Whitney towers and gropes in* 4*-manifolds*, Trans. Amer. Math. Soc. 358 (2006) 4251–4278 MR

[18] **R Schneiderman**, *Introduction to Whitney towers*, Winter Braids Lect. Notes 7 (2020) art. id. 4 MR

[19] **H Zhou**, *Homology concordance and an infinite rank free subgroup*, J. Topol. 14 (2021) 1369–1395 MR

Department of Mathematics, University of Wisconsin–Eau Claire
Eau Claire, WI, United States

daviscw@uwec.edu

Received: 3 April 2023 Revised: 15 March 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3523–3583
DOI: 10.2140/agt.2025.25.3523
Published: 1 October 2025

The asymptotic behaviors of the colored Jones polynomials of the figure-eight knot, and an affine representation

HITOSHI MURAKAMI

We study the asymptotic behavior of the N-dimensional colored Jones polynomial of the figure-eight knot evaluated at $\exp(\kappa + 2p\pi\sqrt{-1}/N)$, where $\kappa := \operatorname{arccosh}(\frac{3}{2})$ and p is a positive integer. We can prove that it grows exponentially with growth rate determined by the Chern–Simons invariant of an affine representation from the fundamental group of the knot complement to the Lie group $\mathrm{SL}(2;\mathbb{C})$.

57K14; 57K10

1 Introduction

For a knot K in the three-dimensional sphere S^3 and a positive integer N, let $J_N(K;q)$ be the N-dimensional colored Jones polynomial associated with the N-dimensional irreducible representation of the Lie algebra $\mathfrak{sl}(2;\mathbb{C})$, where we normalize it as $J_N(U;q) = 1$ for the unknot U, and when $N = 2$ it satisfies the following skein relation:

$$q J_2\big(\times;q\big) - q^{-1} J_2\big(\times;q\big) = (q^{1/2} - q^{-1/2}) J_2\big(\;)\;(\;;q\big).$$

If we replace q with $e^{2\pi\sqrt{-1}/N}$, we obtain a complex number $J_N(K;e^{2\pi\sqrt{-1}/N})$, which is known as the Kashaev invariant; see Kashaev [14], and J Murakami and the author [25]. The volume conjecture (see Kashaev [15], and J Murakami and the author [25]) states that the series $\{J_N(K;e^{2\pi\sqrt{-1}/N})\}_{N=1,2,3,\dots}$ grows exponentially with growth rate proportional to the simplicial volume $\mathrm{Vol}(K)$ of $S^3 \setminus K$. Here the simplicial volume is also known as the Gromov norm; see Gromov [7]. It coincides with the hyperbolic volume divided by the volume v_3 of the ideal regular hyperbolic tetrahedron if the knot is hyperbolic, that is, its complement $S^3 \setminus K$ possesses a complete hyperbolic structure with finite volume. If the knot is not hyperbolic, then the simplicial volume is the sum of the hyperbolic volumes of the hyperbolic pieces of $S^3 \setminus K$ after the Jaco–Shalen–Johannson decomposition; see Jaco and Shalen [12] and Johannson [13].

Conjecture 1.1 (volume conjecture) *For any knot K in S^3, we have*

$$\lim_{N\to\infty} \frac{\log|J_N(K;e^{2\pi\sqrt{-1}/N})|}{N} = \frac{v_3\,\mathrm{Vol}(K)}{2\pi}.$$

The volume conjecture has been generalized in various ways. It can be complexified as follows; see J Murakami, Okamoto, Takata, Yokota, and the author [26]. Let $\mathscr{K} \subset S^3$ be a hyperbolic knot, and

$$\mathrm{cv}(\mathscr{K}) := \sqrt{-1}\,\mathrm{Vol}(\mathscr{K}) - 2\pi^2\,\mathrm{CS}^{\mathrm{SO}(3)}(\mathscr{K})$$

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

be the complex volume of $S^3 \setminus \mathscr{H}$, where $\mathrm{CS}^{\mathrm{SO}(3)}(\mathscr{H}) \pmod{\pi^2}$ is the Chern–Simons invariant of the Levi-Civita connection of $S^3 \setminus \mathscr{H}$ associated with the complete hyperbolic structure.

Conjecture 1.2 (complexification of the volume conjecture) *For any hyperbolic knot $\mathscr{H}$ in S^3, we have*

$$\lim_{N\to\infty} \frac{\log J_N(\mathscr{H}; e^{2\pi\sqrt{-1}/N})}{N} = \frac{\mathrm{cv}(\mathscr{H})}{2\pi\sqrt{-1}}.$$

The volume conjecture and its complexification can be refined as follows (see Gukov [8], and also Gukov and H Murakami [9], Dimofte, Gukov, Lenells, and Zagier [5], and Ohtsuki [30]):

Conjecture 1.3 (refined volume conjecture) *Let $\mathscr{H} \subset S^3$ be a hyperbolic knot. Then we have the asymptotic equivalence*

$$J_N(\mathscr{H}; e^{2\pi\sqrt{-1}/N}) \underset{N\to\infty}{\sim} \left(\frac{T(\mathscr{H})}{2\sqrt{-1}}\right)^{1/2} N^{3/2} \exp\left(\frac{\mathrm{cv}(\mathscr{H})}{2\pi\sqrt{-1}} N\right),$$

where $F(N) \underset{N\to\infty}{\sim} G(N)$ means $\lim_{N\to\infty} F(N)/G(N) = 1$, and $T(\mathscr{H})$ is the adjoint (cohomological) Reidemeister torsion twisted by the holonomy representation $\rho_0 \colon \pi_1(S^3 \setminus \mathscr{H}) \to \mathrm{SL}(2;\mathbb{C})$.

The refined volume conjecture has been proved for the figure-eight knot (see Andersen and Hansen [1]) and hyperbolic knots with at most seven crossings; see Ohtsuki [30; 31] and Ohtsuki and Yokota [32].

We can also generalize the refined volume conjecture by replacing $2\pi\sqrt{-1}$ in $e^{2\pi\sqrt{-1}/N}$ with a complex number.

Let $\rho_u \colon \pi_1(S^3 \setminus \mathscr{H}) \to \mathrm{SL}(2;\mathbb{C})$ be an irreducible representation, which is a small deformation of the holonomy representation ρ_0. Then it defines an incomplete hyperbolic structure of $S^3 \setminus \mathscr{H}$. Up to conjugation, we can assume that ρ_u sends the meridian and preferred longitude of $\mathscr{H}$ to

$$\begin{pmatrix} e^{u/2} & * \\ 0 & e^{-u/2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e^{v(u)/2} & * \\ 0 & e^{-v(u)/2} \end{pmatrix},$$

respectively. Then we can define the cohomological adjoint Reidemeister torsion $T_u(\mathscr{H})$ (see Porti [34]) and the Chern–Simons invariant $\mathrm{CS}_{u,v(u)}(\rho_u)$; see Kirk and Klassen [17].

The following conjecture was proposed by the author [23]; see also Gukov and Murakami [9] and Dimofte and Gukov [4].

Conjecture 1.4 (generalized volume conjecture) *For a hyperbolic knot $\mathscr{H}$, there exists a neighborhood $U \in \mathbb{C}$ of 0 such that if $u \in U \setminus \pi\sqrt{-1}\mathbb{Q}$, then we have the asymptotic equivalence*

$$J_N(\mathscr{H}; e^{(u+2\pi\sqrt{-1})/N}) \underset{N\to\infty}{\sim} \frac{\sqrt{-\pi}}{2\sinh(u/2)} T_u(\mathscr{H})^{1/2} \left(\frac{N}{u+2\pi\sqrt{-1}}\right)^{1/2} \exp\left(\frac{S_u(\mathscr{H})}{u+2\pi\sqrt{-1}} N\right),$$

where $S_u(\mathscr{H}) := \mathrm{CS}_{u,v(u)}(\rho_u) + u\pi\sqrt{-1} + \frac{1}{4}uv(u)$.

The generalized volume has been proved just for the figure-eight knot; see Yokota and the author [28]. The asymptotic equivalence in Conjecture 1.4 was also proved in the case where $0 < u < \kappa := \operatorname{arccosh}\left(\frac{3}{2}\right)$ by the author [23].

In the previous paper [24], the author proved the following theorem generalizing the result in [23]:

Theorem 1.5 *Let $\mathscr{E}$ be the figure-eight knot. For a real number u with $0 < u < \kappa$ and a positive integer p, we have*

$$J_N(\mathscr{E}; e^{(u+2p\pi\sqrt{-1})/N}) \underset{N\to\infty}{\sim} J_p(\mathscr{E}; e^{4N\pi^2/(u+2p\pi\sqrt{-1})}) \frac{\sqrt{-\pi}}{2\sinh\left(\frac{1}{2}u\right)} T_u(\mathscr{E})^{1/2} \left(\frac{N}{u+2p\pi\sqrt{-1}}\right)^{1/2} \exp\left(\frac{S_u(\mathscr{E})}{u+2p\pi\sqrt{-1}} N\right).$$

Note that in the case of the figure-eight knot, we have

$$T_u(\mathscr{E}) = \frac{2}{\sqrt{(2\cosh u+1)(2\cosh u-3)}},$$
$$S_u(\mathscr{E}) = \operatorname{Li}_2(e^{-u-\varphi(u)}) - \operatorname{Li}_2(e^{-u+\varphi(u)}) + u(\varphi(u) + 2\pi\sqrt{-1}),$$

where we put

$$\varphi(u) := \log\left(\cosh u - \tfrac{1}{2} - \tfrac{1}{2}\sqrt{(2\cosh u+1)(2\cosh u-3)}\right),$$

and $\operatorname{Li}_2(z) := -\int_0^z \log(1-x)/x\, dx$ is the dilogarithm function.

So it is impossible to extend Theorem 1.5 to the case where $u = \kappa$ because $T_u(\mathscr{E})$ is not defined. Topologically/geometrically speaking, the corresponding hyperbolic structure of the figure-eight knot complement collapses at $u = \kappa$.

On the other hand, for the figure-eight knot, we have the following theorems:

Theorem 1.6 (the author [21]) *If $\zeta \in \mathbb{C}$ satisfies the inequality $|\cosh\zeta - 1| < \frac{1}{2}$ and $|\operatorname{Im}\zeta| < \frac{1}{3}\pi$, then*

$$\lim_{N\to\infty} J_N(\mathscr{E}; e^{\zeta/N}) = \frac{1}{\Delta(\mathscr{E}; e^{\zeta})},$$

where $\Delta(K;t)$ is the Alexander polynomial of a knot K.

Theorem 1.7 (Hikami and the author [11]) *If $\zeta = \kappa$, then the colored Jones polynomial $J_N(\mathscr{E}; e^{\kappa/N})$ grows polynomially. More precisely,*

$$J_N(\mathscr{E}; e^{\kappa/N}) \underset{N\to\infty}{\sim} \frac{\Gamma\left(\frac{1}{3}\right)}{3^{2/3}} \left(\frac{N}{\kappa}\right)^{2/3},$$

where $\Gamma(x)$ is the gamma function.

Here we will extend Theorem 1.5 to the case $u = \kappa$.

Theorem 1.8 *Let $\mathscr{E}$ be the figure-eight knot, and $\xi := \kappa + 2p\pi\sqrt{-1}$ with $\kappa := \operatorname{arccosh}\left(\frac{3}{2}\right)$ and p a positive integer. Then we have the asymptotic equivalence*

$$J_N(\mathscr{E}; e^{\xi/N}) \underset{N\to\infty}{\sim} J_p(\mathscr{E}; e^{4\pi^2 N/\xi}) \frac{\Gamma\left(\frac{1}{3}\right) e^{\pi\sqrt{-1}/6}}{3^{1/6}} \left(\frac{N}{\xi}\right)^{2/3} \exp\left(\frac{S_\kappa(\mathscr{E})}{\xi} N\right),$$

where $S_\kappa(\mathscr{E}) := 2\kappa\pi\sqrt{-1}$, and we put $\xi^{1/3} := |\xi|^{1/3} e^{\arctan(2p\pi/\kappa)\sqrt{-1}/3}$.

As a corollary, we obtain a similar result for $J_N(\mathscr{E}; e^{\xi'/N})$ with $\xi' := -\kappa + 2p\pi\sqrt{-1}$.

Corollary 1.9 *We have*

$$J_N(\mathscr{E}; e^{\xi'/N}) \underset{N\to\infty}{\sim} J_p(\mathscr{E}; e^{4\pi^2 N/\xi'}) \frac{\Gamma\left(\frac{1}{3}\right) e^{\pi\sqrt{-1}/6}}{3^{1/6}} \left(\frac{N}{\xi'}\right)^{2/3} \exp\left(\frac{S_{-\kappa}(\mathscr{E})}{\xi'} N\right),$$

where we put $S_{-\kappa}(\mathscr{E}) := -2\kappa\pi\sqrt{-1}$.

See Section 6 for a topological interpretation of $S_u(\mathscr{E})$ for $|u| \le \kappa$. It is defined to be

$$\mathrm{CS}_{u,v(u)}(\rho_u) + u\pi\sqrt{-1} + \frac{1}{4} u v(u),$$

where

$$\mathrm{CS}_{u,v(u)}(\rho_u)$$

is the Chern–Simons invariant of a nonabelian representation $\rho_u : \pi_1(S^3 \setminus \mathscr{E}) \to \mathrm{SL}(2;\mathbb{C})$.

Remark 1.10 Since the highest-degree term of the Laurent polynomial $J_p(\mathscr{E}; q)$ is $q^{p(p-1)}$, we have $J_p(\mathscr{E}; e^{4\pi^2 N/\xi}) \underset{N\to\infty}{\sim} e^{4p(p-1)\pi^2 N/\xi}$. So we also have

$$J_N(\mathscr{E}; e^{\xi/N}) \underset{N\to\infty}{\sim} \frac{\Gamma\left(\frac{1}{3}\right) e^{\pi\sqrt{-1}/6}}{3^{1/6}} \left(\frac{N}{\xi}\right)^{2/3} \exp\left(\frac{4p^2\pi^2}{\xi} N\right),$$

because $2\kappa\pi\sqrt{-1}/\xi + 4p(p-1)\pi^2/\xi = 4p^2\pi^2/\xi + 2\pi\sqrt{-1}$. A similar result holds for ξ'.

There are two difficulties in proving Theorem 1.8.

The first one is that when we apply the saddle point method to the integral that approximates $J_N(\mathscr{E}; e^{\xi/N})$, the saddle point is of order two, that is, it looks like the saddle point of $\mathrm{Re}\, z^3$; see Figure 1.

To approximate the colored Jones polynomial by an integral as above, we use a quantum dilogarithm function, which converges to a function described by the dilogarithm function. However, the second difficulty is that our saddle point is on the boundary of the region of convergence. So we need to extend the domain of definition of the quantum dilogarithm slightly by using a functional identity.

The paper is organized as follows. In Section 2, we define the quantum dilogarithm and extend it as we require. In Section 3, we express the colored Jones polynomial as a sum of the terms described by the

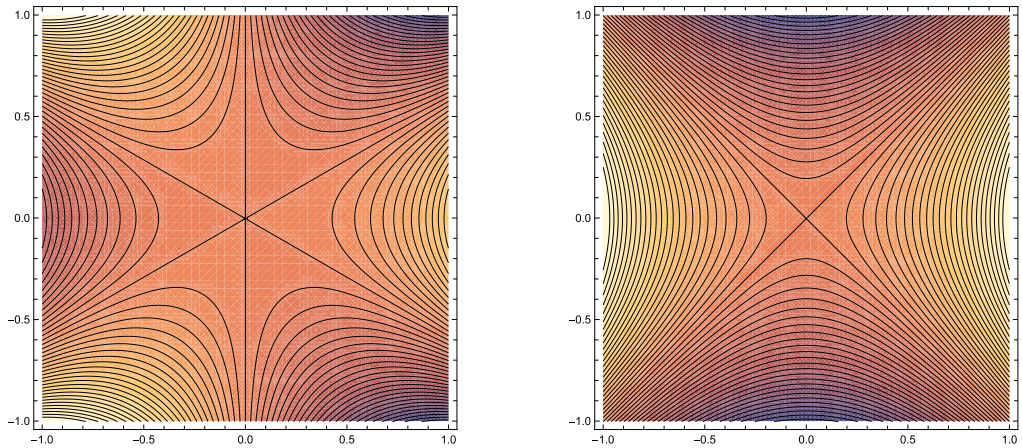


Figure 1: Contour plots of the functions $\operatorname{Re} z^3$ (left) and $\operatorname{Re} z^2$ (right) around their saddle points. The saddle point O of $\operatorname{Re} z^3$ is of order two, and that of $\operatorname{Re} z^2$ is of order one.

quantum dilogarithm. To approximate the sum by an integral, we use the Poisson summation formula in Section 4. Then in Section 5 we use the saddle point method to obtain the asymptotic formula, proving Theorem 1.8. Appendices A and B are devoted to proofs of the Poisson summation formula and the saddle point method, respectively. In Appendix C, we give some computer calculations about the asymptotic behavior of $J_N(\mathcal{S}; e^{(\pm\tilde{\kappa}+2\pi\sqrt{-1})/N})$ for the stevedore knot $\mathcal{S}$, where $\tilde{\kappa} := \log 2$. Since we know that $e^{\pm\kappa}$ ($e^{\pm\tilde{\kappa}}$, respectively) are zeros of the Alexander polynomial of the figure-eight knot (the stevedore knot, respectively), we try to generalize Theorem 1.8 to another knot in vain.

Acknowledgments The author is supported by JSPS KAKENHI grants JP20K03601, JP22H01117, and JP20K03931.

2 Quantum dilogarithm

In this section, we fix a complex number γ with $\operatorname{Re}\gamma > 0$ and $\operatorname{Im}\gamma < 0$. We will introduce a quantum dilogarithm [6]. See also [1; 15; 30].

We put

$$T_N(z) := \frac{1}{4}\int_{\widehat{\mathbb{R}}} \frac{e^{(2z-1)x}}{x\sinh(x)\sinh(\gamma x/N)}\,dx \tag{2-1}$$

for an integer $N > |\gamma|/\pi$, where $\widehat{\mathbb{R}} := (-\infty,-1]\cup\{w\in\mathbb{C} \mid |w|=1,\ \operatorname{Im} w \geq 0\}\cup[1,\infty)$ with orientation from $-\infty$ to ∞. Note that $\widehat{\mathbb{R}}$ avoids the poles of the integrand. We can prove that the integral above converges if $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 1+\operatorname{Re}\gamma/(2N)$.

Lemma 2.1 *The integral in the right side of (2-1) converges if* $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 1+\operatorname{Re}\gamma/(2N)$.

Proof First note that

$$\sinh(as) \underset{s\to\infty}{\sim} \tfrac{1}{2}e^{as} \quad \text{and} \quad \sinh(as) \underset{s\to-\infty}{\sim} -\tfrac{1}{2}e^{-as},$$

for a complex number a with $\operatorname{Re} a > 0$. So we have

$$\frac{e^{(2z-1)x}}{x\sinh(x)\sinh(\gamma x/N)} \underset{x\to\infty}{\sim} \frac{4}{x}\exp((2z-2-\gamma/N)x),$$
$$\frac{e^{(2z-1)x}}{x\sinh(x)\sinh(\gamma x/N)} \underset{x\to-\infty}{\sim} -\frac{4}{x}\exp((2z+\gamma/N)x),$$

since $\operatorname{Re}\gamma > 0$.

Therefore if $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 1+\operatorname{Re}\gamma/(2N)$, the integral converges. □

Thus $T_N(z)$ is a holomorphic function in the region $\{z\in\mathbb{C} \mid -\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 1+\operatorname{Re}\gamma/(2N)\}$.

We will study properties of $T_N(z)$, first introducing three related functions:

Definition 2.2 For a complex number z with $0 < \operatorname{Re} z < 1$, we put

$$\mathcal{L}_0(z) := \int_{\mathbb{R}} \frac{e^{(2z-1)x}}{\sinh(x)}\,dx, \quad \mathcal{L}_1(z) := -\frac{1}{2}\int_{\mathbb{R}} \frac{e^{(2z-1)x}}{x\sinh(x)}\,dx, \quad \mathcal{L}_2(z) := \frac{\pi\sqrt{-1}}{2}\int_{\mathbb{R}} \frac{e^{(2z-1)x}}{x^2\sinh(x)}\,dx.$$

In a similar way to the proof of Lemma 2.1, we can prove that the three integrals above converge if $0 < \operatorname{Re} z < 1$. The functions above can be expressed in terms of well-known functions.

Lemma 2.3 [27, Lemma 2.5] *We have the following formulas*:

$$\mathcal{L}_0(z) = \frac{-2\pi\sqrt{-1}}{1-e^{-2\pi\sqrt{-1}z}}, \tag{2-2}$$

$$\mathcal{L}_1(z) = \begin{cases} \log(1-e^{2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z \geq 0, \\ \pi\sqrt{-1}(2z-1)+\log(1-e^{-2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z < 0, \end{cases} \tag{2-3}$$

$$\mathcal{L}_2(z) = \begin{cases} \operatorname{Li}_2(e^{2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z \geq 0, \\ \pi^2\left(2z^2-2z+\frac{1}{3}\right)-\operatorname{Li}_2(e^{-2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z < 0. \end{cases} \tag{2-4}$$

Here the branch cuts of log *and* Li_2 *are* $(-\infty,0]$ *and* $[1,\infty)$, *respectively.*

The proof is similar to that of [27, Lemma 2.5], and so we omit it.

The function $\mathcal{L}_0(z)$ can be extended to the whole complex plane $\mathbb{C}$ except for integers. The functions $\mathcal{L}_1(z)$ and $\mathcal{L}_2(z)$ can be extended to holomorphic functions on $\mathbb{C}\setminus((-\infty,0]\cup[1,\infty))$ as follows.

Definition 2.4 For a complex number z in $\mathbb{C}\setminus((-\infty,0]\cup[1,\infty))$, we put

$$\mathcal{L}_1(z) = \begin{cases} \log(1-e^{2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z \geq 0, \\ \pi\sqrt{-1}(2z-1)+\log(1-e^{-2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z < 0, \end{cases} \tag{2-5}$$

$$\mathcal{L}_2(z) = \begin{cases} \operatorname{Li}_2(e^{2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z \geq 0, \\ \pi^2\left(2z^2-2z+\frac{1}{3}\right)-\operatorname{Li}_2(e^{-2\pi\sqrt{-1}z}) & \text{if } \operatorname{Im} z < 0. \end{cases} \tag{2-6}$$

Lemma 2.5 *When* $\operatorname{Im} z < 0$, *the functions* $\mathcal{L}_1(z)$ *and* $\mathcal{L}_2(z)$ *can also be written as*

$$\mathcal{L}_1(z) = \log(1-e^{2\pi\sqrt{-1}z}) + 2\lfloor \operatorname{Re} z\rfloor \pi\sqrt{-1}, \quad \mathcal{L}_2(z) = \operatorname{Li}_2(e^{2\pi\sqrt{-1}z}) - 2\pi^2\lfloor \operatorname{Re} z\rfloor(\lfloor \operatorname{Re} z\rfloor - 2z + 1),$$

where $\lfloor x\rfloor$ *is the greatest integer that does not exceed* x.

Proof For $\mathcal{L}_1(z)$, we have

$$\begin{aligned}\log(1-e^{-2\pi\sqrt{-1}z}) &= \log[(1-e^{2\pi\sqrt{-1}z})e^{-2\pi\sqrt{-1}z+\pi\sqrt{-1}}]\\ &= \log(1-e^{2\pi\sqrt{-1}z}) - 2\pi\sqrt{-1}z + \pi\sqrt{-1} + 2\lfloor \operatorname{Re} z\rfloor\pi\sqrt{-1}.\end{aligned}$$

The last equality follows because

- if $0 < \operatorname{Re} z - \lfloor \operatorname{Re} z\rfloor < \frac{1}{2}$, then $-\pi < \arg(1-e^{2\pi\sqrt{-1}z}) < 0$,
- if $\frac{1}{2} \le \operatorname{Re} z - \lfloor \operatorname{Re} z\rfloor < 1$, then $0 \le \arg(1-e^{2\pi\sqrt{-1}z}) < \pi$,

and so the imaginary part of the rightmost side is between $-\pi$ and π. Thus we obtain $\mathcal{L}_1(z) = \log(1-e^{2\pi\sqrt{-1}z}) + 2\lfloor \operatorname{Re} z\rfloor\pi\sqrt{-1}$ from (2-3).

For $\mathcal{L}_2(z)$, from the well known formula

$$\operatorname{Li}_2(w^{-1}) = -\operatorname{Li}_2(w) - \tfrac{1}{6}\pi^2 - \tfrac{1}{2}(\log(-w))^2, \tag{2-7}$$

we have

$$\begin{aligned}\operatorname{Li}_2(e^{-2\pi\sqrt{-1}z}) &= -\operatorname{Li}_2(e^{2\pi\sqrt{-1}z}) - \tfrac{1}{6}\pi^2 - \tfrac{1}{2}(2\pi\sqrt{-1}z - (2\pi\lfloor \operatorname{Re} z\rfloor + \pi)\sqrt{-1})^2\\ &= -\operatorname{Li}_2(e^{2\pi\sqrt{-1}z}) + \pi^2\big(2z^2 - 2z + \tfrac{1}{3}\big) + 2\pi^2\lfloor \operatorname{Re} z\rfloor^2 - 4\pi^2\lfloor \operatorname{Re} z\rfloor z + 2\pi^2\lfloor \operatorname{Re} z\rfloor,\end{aligned}$$

and the result follows. □

Corollary 2.6 *If* $\operatorname{Im} z < 0$, *then we have* $\mathcal{L}_1(z+1) - \mathcal{L}_1(z) = 2\pi\sqrt{-1}$ *and* $\mathcal{L}_2(z+1) - \mathcal{L}_2(z) = 4\pi^2 z$.

Lemma 2.7 *The derivatives of* $\mathcal{L}_i(z)$ *for* $i = 1, 2$ *are given as follows*:

$$\frac{d}{dz}\mathcal{L}_2(z) = -2\pi\sqrt{-1}\mathcal{L}_1(z), \tag{2-8}$$

$$\frac{d}{dz}\mathcal{L}_1(z) = -\mathcal{L}_0(z) = \frac{2\pi\sqrt{-1}}{1-e^{-2\pi\sqrt{-1}z}}. \tag{2-9}$$

Proof The first equality follows from the well-known equality $(d/dw)\operatorname{Li}_2(w) = -\log(1-w)/w$. The second one also follows easily. □

Now we will show three identities expressing the difference $T_N(z+a) - T_N(z)$ in terms of $\mathcal{L}_1$.

Lemma 2.8 *If* $|\operatorname{Re} z| < \operatorname{Re}\gamma/(2N)$, *then*

$$T_N(z) - T_N(z+1) = \mathcal{L}_1\Big(\frac{N}{\gamma}z + \frac{1}{2}\Big). \tag{2-10}$$

Remark 2.9 Since $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < \operatorname{Re}\gamma/(2N)$ and $1-\operatorname{Re}\gamma/(2N) < \operatorname{Re}(z+1) < 1+\operatorname{Re}\gamma/(2N)$, both z and $z+1$ are in the domain of T_N.

We will check that $Nz/\gamma+\frac{1}{2}$ is in $\mathbb{C}\setminus((-\infty,0]\cup[1,\infty))$, the domain of $\mathcal{L}_1$.

If not, then $(N/\gamma)z+\frac{1}{2}=s$ for $s\le 0$ or $s\ge 1$. Putting $s':=s-\frac{1}{2}$, we have $z=(\gamma/N)s'$ with $|s'|\ge\frac{1}{2}$, which implies $|\operatorname{Re} z|\ge \operatorname{Re}\gamma/(2N)$, a contradiction.

Proof By definition, we have

$$T_N(z)-T_N(z+1)=\frac{1}{4}\int_{\widehat{\mathbb{R}}}\frac{e^{(2z-1)x}-e^{(2z+1)x}}{x\sinh x\sinh(\gamma x/N)}\,dx=-\frac{1}{2}\int_{\widehat{\mathbb{R}}}\frac{e^{2zx}}{x\sinh(\gamma x/N)}\,dx.$$

Then setting $y:=\gamma x/N$, this equals

$$-\frac{1}{2}\int_{\widehat{\mathbb{R}}'}\frac{e^{2Nzy/\gamma}}{y\sinh y}\,dy=-\frac{1}{2}\int_{\widehat{\mathbb{R}}}\frac{e^{2Nzy/\gamma}}{y\sinh y}\,dy=\mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big),$$

where $\widehat{\mathbb{R}}'$ is obtained from $\widehat{\mathbb{R}}$ by multiplying by γ/N. The last equality follows since there are no poles of $1/(y\sinh y)$, that is, integer multiples of $\pi\sqrt{-1}$ between $\widehat{\mathbb{R}}$ and $\widehat{\mathbb{R}}'$. □

Lemma 2.10 *If* $0<\operatorname{Re} z<1$, *then*

$$T_N\Big(z-\frac{\gamma}{2N}\Big)-T_N\Big(z+\frac{\gamma}{2N}\Big)=\mathcal{L}_1(z). \tag{2-11}$$

Proof From the definition, we have

$$\begin{aligned}T_N\Big(z-\frac{\gamma}{2N}\Big)-T_N\Big(z+\frac{\gamma}{2N}\Big)&=\frac{1}{4}\int_{\widehat{\mathbb{R}}}\frac{e^{(2z-\gamma/N-1)x}-e^{(2z+\gamma/N-1)x}}{x\sinh x\sinh(\gamma x/N)}\,dx\\&=-\frac{1}{2}\int_{\widehat{\mathbb{R}}}\frac{e^{(2z-1)x}}{x\sinh x}\,dx=\mathcal{L}_1(z).\end{aligned}$$

□

The third one is a little tricky.

Lemma 2.11 *If* $|\operatorname{Re} z|<\operatorname{Re}\gamma/N<1$, *then*

$$T_N\Big(z+1-\frac{\gamma}{2N}\Big)-T_N\Big(z+\frac{\gamma}{2N}\Big)=\begin{cases}\mathcal{L}_1(z)-\mathcal{L}_1((N/\gamma)z) & \text{if } \operatorname{Re} z\ge 0 \text{ and } z\ne 0,\\ \mathcal{L}_1(z+1)-\mathcal{L}_1((N/\gamma)z+1) & \text{if } \operatorname{Re} z<0,\\ \log((\gamma/N)) & \text{if } z=0.\end{cases} \tag{2-12}$$

Remark 2.12 If $|\operatorname{Re} z|<\operatorname{Re}\gamma/N<1$, then $1-3\operatorname{Re}\gamma/(2N)<\operatorname{Re}(z+1-\gamma/(2N))<1+\operatorname{Re}\gamma/(2N)$ and $-\operatorname{Re}\gamma/(2N)<\operatorname{Re}(z+\gamma/(2N))<3\operatorname{Re}\gamma/(2N)$, and so both $z+1-\gamma/(2N)$ and $z+\gamma/(2N)$ are in the domain of T_N.

We will check that the arguments in the right-hand side are in $\mathbb{C}\setminus((-\infty,0]\cup[1,\infty))$, the domain of $\mathcal{L}_1$.

• Suppose $0\le\operatorname{Re} z$ and $z\ne 0$. Since $\operatorname{Re} z<\operatorname{Re}\gamma/N$, z is in the domain of $\mathcal{L}_1$ if N is sufficiently large. Suppose for a contradiction that $(N/\gamma)z$ is not in the domain of $\mathcal{L}_1$. Then $(N/\gamma)z\in(-\infty,0]\cup[1,\infty)$ and so $(N/\gamma)z=s$ for $s\ge 1$ or $s\le 0$. If $s\le 0$, then $\operatorname{Re} z=s\operatorname{Re}\gamma/N\le 0$ and so $z=s=0$, which is a contradiction. If $s\ge 1$, then $\operatorname{Re} z\ge s\operatorname{Re}\gamma/N\ge\operatorname{Re}\gamma/N$, which is also a contradiction.

• Suppose $\operatorname{Re} z < 0$. Then $z+1$ is in the domain of $\mathcal{L}_1$ because $1 - \operatorname{Re}\gamma/N < \operatorname{Re}(z+1) < 1$. Suppose for a contradiction that $Nz/\gamma + 1$ is not in the domain of $\mathcal{L}_1$. Then $(N/\gamma)z + 1 = s$ for $s \le 0$ or $s \ge 1$. Thus $\operatorname{Re} z = \operatorname{Re}((s-1)\gamma/N)$ and so we have $\operatorname{Re} z \le -\operatorname{Re}\gamma/N$, which is impossible.

Thus the arguments in the right-hand side are in the domain of $\mathcal{L}_1$.

Proof We first assume that $\operatorname{Re} z > 0$. Then from Lemmas 2.10 and 2.8 we have

$$T_N\left(z - \frac{\gamma}{2N}\right) - T_N\left(z + \frac{\gamma}{2N}\right) = \mathcal{L}_1(z), \quad T_N\left(z + 1 - \frac{\gamma}{2N}\right) - T_N\left(z - \frac{\gamma}{2N}\right) = -\mathcal{L}_1\left(\frac{N}{\gamma}z\right),$$

and the equality follows. Note that $-\operatorname{Re}\gamma/(2N) < \operatorname{Re}(z - \gamma/(2N)) < \operatorname{Re}\gamma/(2N)$ and so we can apply Lemma 2.8 to the second equality. Similarly, if $\operatorname{Re} z < 0$, we have

$$T_N\left(z+1-\frac{\gamma}{2N}\right)-T_N\left(z+1+\frac{\gamma}{2N}\right)=\mathcal{L}_1(z+1), \quad T_N\left(z+1+\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right)=-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right),$$

and the equality also holds.

When $\operatorname{Re} z = 0$, put $z := y\sqrt{-1}$ for $y \in \mathbb{R} \setminus \{0\}$ and consider the limit

$$\lim_{\varepsilon \to 0}\left(T_N\left(y\sqrt{-1} + 1 - \frac{\gamma}{2N} + \varepsilon\right) - T_N\left(y\sqrt{-1} + \frac{\gamma}{2N} + \varepsilon\right)\right).$$

Since T_N is a holomorphic function in $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 1 + \operatorname{Re}\gamma/(2N)$, the limit above coincides with the left-hand side of (2-10). From the result above, considering the limit from the right, we have

$$\begin{aligned} T_N\left(y\sqrt{-1} + 1 - \frac{\gamma}{2N}\right) - T_N\left(y\sqrt{-1} + \frac{\gamma}{2N}\right) &= \lim_{\varepsilon \searrow 0}\left(T_N\left(y\sqrt{-1} + 1 - \frac{\gamma}{2N} + \varepsilon\right) - T_N\left(y\sqrt{-1} + \frac{\gamma}{2N} + \varepsilon\right)\right) \\ &= \lim_{\varepsilon \searrow 0}\left(\mathcal{L}_1(y\sqrt{-1} + \varepsilon) - \mathcal{L}_1\left(\frac{N}{\gamma}(y\sqrt{-1} + \varepsilon)\right)\right) \\ &= \mathcal{L}_1(y\sqrt{-1}) - \mathcal{L}_1\left(\frac{N}{\gamma}(y\sqrt{-1})\right) \end{aligned}$$

if $y \neq 0$, because we extend $\mathcal{L}_1(z)$ to $\mathbb{C} \setminus ((-\infty, 0] \cup [1, \infty))$. Let us confirm that the limit from the left gives the same answer. We have

$$\lim_{\varepsilon \nearrow 0}\left(\mathcal{L}_1(y\sqrt{-1} + \varepsilon + 1) - \mathcal{L}_1\left(\frac{N}{\gamma}(y\sqrt{-1} + \varepsilon) + 1\right)\right) = \mathcal{L}_1(y\sqrt{-1} + 1) - \mathcal{L}_1\left(\frac{N}{\gamma}y\sqrt{-1} + 1\right),$$

which coincides with $\mathcal{L}_1(y\sqrt{-1}) - \mathcal{L}_1((N/\gamma)(y\sqrt{-1}))$ if $y \neq 0$ from Lemma 2.5, noting that

$$\operatorname{Im}((N/\gamma)y\sqrt{-1}) = Ny \operatorname{Re}\gamma/|\gamma|^2$$

has the same sign as y.

Now, we consider the case where $z = 0$. Since $\operatorname{Im}\gamma < 0$, we have $\operatorname{Im}(N/\gamma)\varepsilon > 0$ for $\varepsilon > 0$. Thus

$$\lim_{\varepsilon \searrow 0}\left(\mathcal{L}_1(\varepsilon) - \mathcal{L}_1\left(\frac{N}{\gamma}\varepsilon\right)\right) = \lim_{\varepsilon \searrow 0}(\log(1 - e^{2\pi\sqrt{-1}\varepsilon}) - \log(1 - e^{2N\varepsilon\pi\sqrt{-1}/\gamma})). \tag{2-13}$$

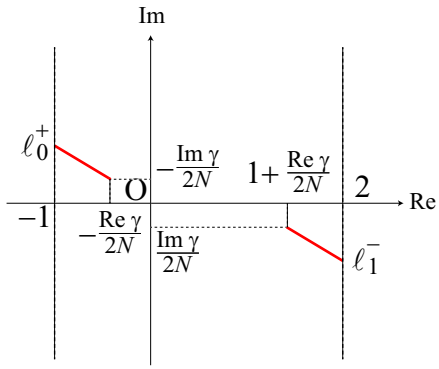


Figure 2: The region (2-14) is between the two thick dotted lines minus the two red lines ℓ_0^+ and ℓ_1^-.

Since we have that $\lim_{\varepsilon\searrow 0}\arg(1-e^{2\pi\sqrt{-1}\varepsilon})=-\frac{1}{2}\pi$, and $-\pi<\arg(1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma})<0$ because $\operatorname{Im}(1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma})<0$, (2-13) turns out to be

$$\lim_{\varepsilon\searrow 0}\log\frac{1-e^{2\pi\sqrt{-1}\varepsilon}}{1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma}}=\log\left(\frac{\gamma}{N}\right)$$

by l'Hôpital's rule,

Just for safety, we will check the other limit, $\lim_{\varepsilon\nearrow 0}(\mathcal{L}_1(\varepsilon+1)-\mathcal{L}_1((N/\gamma)\varepsilon+1))$. Since $\operatorname{Im}(N\varepsilon/\gamma+1)<0$ when $\varepsilon<0$, from Lemma 2.5

$$\begin{aligned}\lim_{\varepsilon\nearrow 0}&\left(\mathcal{L}_1(\varepsilon+1)-\mathcal{L}_1\left(\frac{N}{\gamma}\varepsilon+1\right)\right)\\&=\lim_{\varepsilon\nearrow 0}\left(\log(1-e^{2\pi\sqrt{-1}\varepsilon})-\log(1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma})-2\pi\sqrt{-1}\left\lfloor\operatorname{Re}\left(\frac{N}{\gamma}\varepsilon\right)+1\right\rfloor\right)\\&=\lim_{\varepsilon\nearrow 0}\log\frac{1-e^{2\pi\sqrt{-1}\varepsilon}}{1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma}}=\log\left(\frac{\gamma}{N}\right),\end{aligned}$$

where the second equality follows since $\lim_{\varepsilon\nearrow 0}\arg(1-e^{2\pi\sqrt{-1}\varepsilon})=\frac{1}{2}\pi$, $0<\arg(1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma})<\pi$ because $\operatorname{Im}(1-e^{2N\varepsilon\pi\sqrt{-1}/\gamma})>0$, and $\lim_{\varepsilon\nearrow 0}\lfloor\operatorname{Re}(N\varepsilon/\gamma)+1\rfloor=0$. □

We use Lemma 2.8 to extend the function T_N to the region

$$\{z\in\mathbb{C}\mid -1<\operatorname{Re}z<2\}\setminus(\ell_0^+\cup\ell_1^-), \tag{2-14}$$

where

$$\ell_0^+:=\left\{z\in\mathbb{C}\;\middle|\;z=s\gamma\text{ with }-\frac{1}{\operatorname{Re}\gamma}<s\le-\frac{1}{2N}\right\},\quad \ell_1^-:=\left\{z\in\mathbb{C}\;\middle|\;z=1+s\gamma\text{ with }\frac{1}{2N}\le s<\frac{1}{\operatorname{Re}\gamma}\right\}.$$

See Figure 2. Note that T_N is already defined for z with $-\gamma/(2N)<\operatorname{Re}z<1+\gamma/(2N)$.

If $-1<\operatorname{Re}z\le-\operatorname{Re}\gamma/(2N)$, then we use (2-10) to define

$$T_N(z):=T_N(z+1)+\mathcal{L}_1\left(\frac{N}{\gamma}z+\frac{1}{2}\right), \tag{2-15}$$

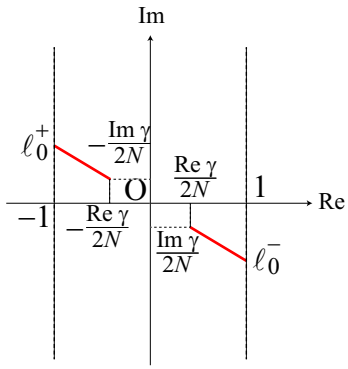


Figure 3: The region (2-17) is between the two thick dotted lines minus the two red lines ℓ_0^+ and ℓ_0^-.

noting that $z+1$ is in the domain of T_N. For the argument of $\mathcal{L}_1$, see Remark 2.13 below. Similarly, if $1+\operatorname{Re}\gamma/(2N) \le \operatorname{Re} z < 2$, we define

$$T_N(z) := T_N(z-1) - \mathcal{L}_1\Big(\frac{N}{\gamma}(z-1)+\frac{1}{2}\Big), \tag{2-16}$$

noting that $z-1$ is in the domain of T_N. For the argument of $\mathcal{L}_1$, see Remark 2.13 below.

Remark 2.13 Recall that $\mathcal{L}_1(z)$ is defined except for $z \in (-\infty, 0] \cup [1, \infty)$. Therefore $Nz/\gamma + \frac{1}{2}$ and $N(z-1)/\gamma + \frac{1}{2}$ are in the domain of $\mathcal{L}_1$ unless

- $-1 < \operatorname{Re} z \le -\operatorname{Re}\gamma/(2N)$ and $(N/\gamma)z + \frac{1}{2} = s$ for $s \in (-\infty, 0] \cup [1, \infty)$, or
- $1+\operatorname{Re}\gamma/(2N) \le \operatorname{Re} z < 2$ and $(N/\gamma)(z-1) + \frac{1}{2} = t$ for $t \in (-\infty, 0) \cup (1, \infty)$.

This is equivalent to

- $-1 < \operatorname{Re} z \le -\operatorname{Re}\gamma/(2N)$ and $z = s'\gamma$ with $|s'| \ge 1/(2N)$, or
- $1+\operatorname{Re}\gamma/(2N) \le \operatorname{Re} z < 2$ and $z = 1 + t'\gamma$ with $|t'| \ge 1/(2N)$.

Since $\operatorname{Re}\gamma > 0$, the condition above turns out to be $z \in \ell_0^+$ or $z \in \ell_1^-$.

We will also use $T_N(z)$ to denote the function extended by using (2-15) and (2-16). Then we have:

Lemma 2.14 *The function $T_N(z)$ extended as above also satisfies (2-10) for any z in the region*

$$\{z \in \mathbb{C} \mid -1 < \operatorname{Re} z < 1\} \setminus (\ell_0^+ \cup \ell_0^-) \tag{2-17}$$

with $\ell_0^- := \{z \in \mathbb{C} \mid z = s\gamma$ with $1/(2N) \le s < 1/\operatorname{Re}\gamma\}$; see Figure 3.

Remark 2.15 As in Remark 2.13, $Nz/\gamma + \frac{1}{2}$ is in the domain of $\mathcal{L}_1$ unless $z \in \ell_0^+ \cup \ell_0^-$.

Proof If $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < \operatorname{Re}\gamma/(2N)$, then (2-10) is proved in Lemma 2.8. If $-1 < \operatorname{Re} z \le -\operatorname{Re}\gamma/(2N)$ and $\operatorname{Re}\gamma/(2N) \le \operatorname{Re} z < 1$, then we define T_N by (2-15) and (2-16), respectively, so that (2-10) holds. □

Lemma 2.16 *The function $T_N(z)$ defined as above is holomorphic in the region (2-14).*

Proof From (2-1), $T_N(z)$ is holomorphic in $\{z\in\mathbb{C}\mid -\operatorname{Re}\gamma/(2N)<\operatorname{Re}z<1+\operatorname{Re}\gamma/(2N)\}$. Therefore from the definition using (2-15) and (2-16), $T_N(z)$ is holomorphic in the disjoint strips

$$\left\{z\in\mathbb{C}\mid -1<\operatorname{Re}z<-\frac{\operatorname{Re}\gamma}{2N}\right\}\sqcup\left\{z\in\mathbb{C}\mid 1+\frac{\operatorname{Re}\gamma}{2N}<\operatorname{Re}z<2\right\}.$$

So we need to confirm that $T_N(z)$ is holomorphic for z with $\operatorname{Re}z=-\operatorname{Re}\gamma/(2N)$ or $1+\operatorname{Re}\gamma/(2N)$.

Let B be an open disk centered at z (where $\operatorname{Re}z=-\operatorname{Re}\gamma/(2N)$) with radius less than $\operatorname{Re}\gamma/(2N)$. Then for $w\in B$ with $\operatorname{Re}w\le-\operatorname{Re}\gamma/(2N)$, we have

$$T_N(w)=T_N(w+1)+\mathcal{L}_1\Big(\frac{N}{\gamma}w+\frac{1}{2}\Big)$$

from (2-15). On the other hand, for $w\in B$ with $\operatorname{Re}w>-\operatorname{Re}\gamma/(2N)$, $T_N(w)$ is defined by using (2-1). However, from Lemma 2.8, this coincides with $T_N(w+1)+\mathcal{L}_1\big((N/\gamma)w+\frac{1}{2}\big)$. Therefore T_N is holomorphic in this case.

Similarly, we can prove the holomorphicity of T_N for the other case. □

Let Ω be the region defined as

$$\Omega:=\left\{z\in\mathbb{C}\;\middle|\;-1+\frac{\operatorname{Re}\gamma}{2N}<\operatorname{Re}z<2-\frac{\operatorname{Re}\gamma}{2N}\right\}\setminus(\Delta_0^+\cup\Delta_1^+),\tag{2-18}$$

where we put

$$\Delta_0^+:=\left\{z\in\mathbb{C}\;\middle|\;-1+\frac{\operatorname{Re}\gamma}{2N}<\operatorname{Re}z\le 0,\ \operatorname{Im}z\ge 0,\ \text{and }\operatorname{Im}\Big(\frac{z}{\gamma}\Big)\le 0\right\},$$

$$\Delta_1^-:=\left\{z\in\mathbb{C}\;\middle|\;1\le\operatorname{Re}z<2-\frac{\operatorname{Re}\gamma}{2N},\ \operatorname{Im}z\le 0,\ \text{and }\operatorname{Im}\Big(\frac{z-1}{\gamma}\Big)\ge 0\right\}.$$

See Figure 4. Note that Ω is contained in the region (2-14) because

$$\ell_0^+\cap\left\{z\in\mathbb{C}\;\middle|\;-1+\frac{\operatorname{Re}\gamma}{2N}<\operatorname{Re}z<2-\frac{\operatorname{Re}\gamma}{2N}\right\}$$

and

$$\ell_1^-\cap\left\{z\in\mathbb{C}\;\middle|\;-1+\frac{\operatorname{Re}\gamma}{2N}<\operatorname{Re}z<2-\frac{\operatorname{Re}\gamma}{2N}\right\},$$

are on the upper side of Δ_0^+ and the lower side of Δ_1^-, respectively.

Lemma 2.17 *The function $T_N(z)$ extended by using (2-10) satisfies (2-11) for $z\in\Omega$.*

Remark 2.18 The left-hand side of (2-11) is defined for z such that $z\pm\gamma/(2N)$ is in the region (2-14), that is, $z\pm\gamma/(2N)\notin\ell_0^+\cup\ell_1^-$. This is equivalent to saying that z is not on the two rays $\{s\gamma\in\mathbb{C}\mid s\le 0\}\cup\{1+s\gamma\in\mathbb{C}\mid s\ge 0\}$. Note that the ray $\{s\gamma\in\mathbb{C}\mid s\le 0\}$ includes the upper edge of Δ_0^+, and that the ray $\{1+s\gamma\in\mathbb{C}\mid s\ge 0\}$ includes the lower edge of Δ_1^-. The right-hand side of (2-11) is defined unless $z\in(-\infty,0]\cup[1,\infty)$.

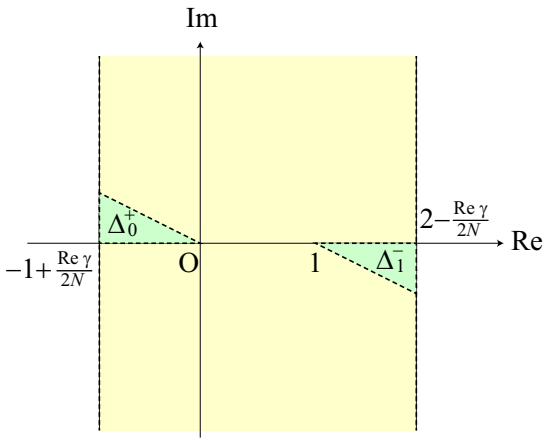


Figure 4: The yellow region is Ω. The green triangles are Δ_0^+ and Δ_1^-.

Proof We need to prove (2-11) for z with $-1+\operatorname{Re}\gamma/(2N) < \operatorname{Re} z \le 0$ or $1 \le \operatorname{Re} z < 2-\operatorname{Re}\gamma/(2N)$.

If $-\operatorname{Re}\gamma/(2N) < \operatorname{Re} z < 0$, from (2-12), we have

$$\begin{aligned} T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right) &= T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+1-\frac{\gamma}{2N}\right)+\mathcal{L}_1(z+1)-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right)\\ &= \mathcal{L}_1\left(\frac{N}{\gamma}\left(z-\frac{\gamma}{2N}\right)+\frac{1}{2}\right)+\mathcal{L}_1(z+1)-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right)\\ &= \mathcal{L}_1\left(\frac{N}{\gamma}z\right)+\mathcal{L}_1(z+1)-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right), \end{aligned}$$

where we use Lemma 2.14 for $z-\gamma/(2N)$ at the second equality. If $\operatorname{Im} z \ge 0$, then $\operatorname{Im}(z/\gamma) > 0$ from (2-18). So $\mathcal{L}_1(z+1)=\mathcal{L}_1(z)$ and $\mathcal{L}_1(Nz/\gamma+1)=\mathcal{L}_1(Nz/\gamma)$ from (2-5), which implies (2-11). If $\operatorname{Im} z < 0$, then we have $\operatorname{Im}(Nz/\gamma+1) = (N/|\gamma|^2)(\operatorname{Re}\gamma \operatorname{Im} z - \operatorname{Im}\gamma \operatorname{Re} z) < 0$. So $\mathcal{L}_1(z+1)=\mathcal{L}_1(z)+2\pi\sqrt{-1}$ and $\mathcal{L}_1(Nz/\gamma+1)=\mathcal{L}_1(Nz/\gamma)+2\pi\sqrt{-1}$ from Corollary 2.6, proving (2-11).

If $\operatorname{Re} z = 0$, then noting that 0 is not included in Ω, similarly we have

$$\begin{aligned} T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right) &= T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+1-\frac{\gamma}{2N}\right)+\mathcal{L}_1(z)-\mathcal{L}_1\left(\frac{N}{\gamma}z\right)\\ &= \mathcal{L}_1\left(\frac{N}{\gamma}\left(z-\frac{\gamma}{2N}\right)+\frac{1}{2}\right)+\mathcal{L}_1(z)-\mathcal{L}_1\left(\frac{N}{\gamma}z\right)=\mathcal{L}_1(z). \end{aligned}$$

If $\operatorname{Re} z = -\operatorname{Re}\gamma/(2N)$, then $\operatorname{Re}(z+\gamma/(2N))=0$ and $-1<\operatorname{Re}(z-\gamma/(2N))=-\operatorname{Re}\gamma/N<-\operatorname{Re}\gamma/(2N)$. Therefore from (2-15) we have

$$\begin{aligned} T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right) &= T_N\left(z-\frac{\gamma}{2N}+1\right)-T_N\left(z+\frac{\gamma}{2N}\right)+\mathcal{L}_1\left(\frac{N}{\gamma}\left(z-\frac{\gamma}{2N}\right)+\frac{1}{2}\right)\\ &= T_N\left(z-\frac{\gamma}{2N}+1\right)-T_N\left(z+\frac{\gamma}{2N}\right)+\mathcal{L}_1\left(\frac{N}{\gamma}z\right)\\ &= \mathcal{L}_1(z+1)-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right)+\mathcal{L}_1\left(\frac{N}{\gamma}z\right), \end{aligned} \tag{2-19}$$

where the last equality follows from Lemma 2.11 since $\operatorname{Re} z < 0$. If $\operatorname{Im} z \geq 0$, then $\operatorname{Im}(z/\gamma) > 0$, and so (2-19) turns out to be $\mathcal{L}_1(z)$. If $\operatorname{Im} z < 0$, then $\operatorname{Im}(z/\gamma) < 0$. Therefore (2-19) equals

$$\log(1 - e^{2\pi\sqrt{-1}z}) - 2\pi\sqrt{-1} = \mathcal{L}_1(z)$$

from Lemma 2.5 and Corollary 2.6.

We consider the case where $-1 + \operatorname{Re}\gamma/(2N) < \operatorname{Re} z < -\operatorname{Re}\gamma/(2N)$. Note that $-1 < \operatorname{Re}(z \pm \gamma/(2N)) < 0$. Therefore from (2-15) we have

$$\begin{aligned} &T_N\left(z - \frac{\gamma}{2N}\right) - T_N\left(z + \frac{\gamma}{2N}\right) \\ &\quad = T_N\left(z - \frac{\gamma}{2N} + 1\right) - T_N\left(z + \frac{\gamma}{2N} + 1\right) + \mathcal{L}_1\left(\frac{N}{\gamma}\left(z - \frac{\gamma}{2N}\right) + \frac{1}{2}\right) - \mathcal{L}_1\left(\frac{N}{\gamma}\left(z + \frac{\gamma}{2N}\right) + \frac{1}{2}\right) \\ &\quad = \mathcal{L}_1(z+1) + \mathcal{L}_1\left(\frac{N}{\gamma}z\right) - \mathcal{L}_1\left(\frac{N}{\gamma}z + 1\right), \end{aligned}$$

where we use (2-11) because $0 < \operatorname{Re}(z+1) < 1$. By the same reason as above, this equals $\mathcal{L}_1(z)$.

If $1 \leq \operatorname{Re} z < 1 + \operatorname{Re}\gamma/(2N)$, then from (2-16), we have

$$\begin{aligned} T_N\left(z - \frac{\gamma}{2N}\right) - T_N\left(z + \frac{\gamma}{2N}\right) &= T_N\left(z - \frac{\gamma}{2N}\right) - T_N\left(z + \frac{\gamma}{2N} - 1\right) + \mathcal{L}_1\left(\frac{N}{\gamma}\left(z + \frac{\gamma}{2N} - 1\right) + \frac{1}{2}\right) \\ &= \mathcal{L}_1(z-1) - \mathcal{L}_1\left(\frac{N}{\gamma}(z-1)\right) + \mathcal{L}_1\left(\frac{N}{\gamma}(z-1) + 1\right), \end{aligned}$$

where we use (2-12) for $z-1$ at the second identity, noting $1 \notin \Omega$. If $\operatorname{Im} z \geq 0$, then $\operatorname{Im}((z-1)/\gamma) = (1/|\gamma|^2)(\operatorname{Im}\gamma(1 - \operatorname{Re} z) + \operatorname{Re}\gamma \operatorname{Im} z) \geq 0$, so the last line equals $\mathcal{L}_1(z)$. If $\operatorname{Im} z < 0$, then $\operatorname{Im}((z-1)/\gamma) < 0$ from the definition of Δ_1^-, and so we have $\mathcal{L}_1(z-1) = \mathcal{L}_1(z) - 2\pi\sqrt{-1}$ and $\mathcal{L}_1(N(z-1)/\gamma + 1) = \mathcal{L}_1(N(z-1)/\gamma) + 2\pi\sqrt{-1}$ from Corollary 2.6, which implies (2-11).

Lastly, we consider the case where $1 + \operatorname{Re}\gamma/(2N) \leq \operatorname{Re} z < 2 - \operatorname{Re}\gamma/(2N)$. Since $1 \leq \operatorname{Re}(z \pm \gamma/(2N)) < 2$, from (2-16), we have

$$\begin{aligned} &T_N\left(z - \frac{\gamma}{2N}\right) - T_N\left(z + \frac{\gamma}{2N}\right) \\ &\quad = T_N\left(z - \frac{\gamma}{2N} - 1\right) - T_N\left(z + \frac{\gamma}{2N} - 1\right) - \mathcal{L}_1\left(\frac{N}{\gamma}\left(z - \frac{\gamma}{2N} - 1\right) + \frac{1}{2}\right) \\ &\qquad\qquad + \mathcal{L}_1\left(\frac{N}{\gamma}\left(z + \frac{\gamma}{2N} - 1\right) + \frac{1}{2}\right) \\ &\quad = \mathcal{L}_1(z-1) - \mathcal{L}_1\left(\frac{N}{\gamma}(z-1)\right) + \mathcal{L}_1\left(\frac{N}{\gamma}(z-1) + 1\right), \end{aligned} \tag{2-20}$$

using (2-11) at the last equality. If $\operatorname{Im} z \geq 0$, then $\operatorname{Im}((z-1)/\gamma) = (1/|\gamma|^2)(\operatorname{Im}\gamma(1 - \operatorname{Re} z) + \operatorname{Re}\gamma \operatorname{Im} z) > 0$ since $\operatorname{Re} z > 1$. So (2-20) equals $\log(1 - e^{2\pi\sqrt{-1}z}) = \mathcal{L}_1(z)$. If $\operatorname{Im} z < 0$, then $\operatorname{Im}((z-1)/\gamma) < 0$. Therefore (2-20) becomes

$$\log(1 - e^{2\pi\sqrt{-1}z}) + 2\pi\sqrt{-1} = \mathcal{L}_1(z)$$

from Lemma 2.5. □

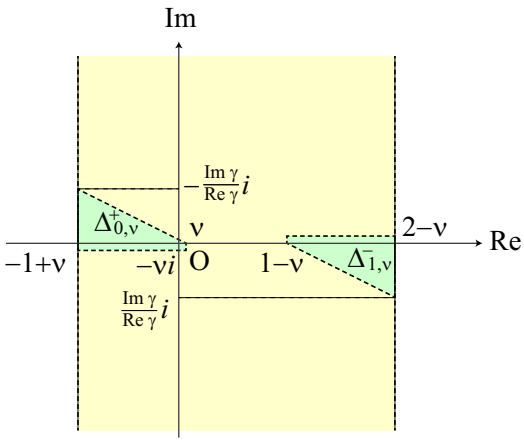


Figure 5: The yellow region is Ω^*_ν. The green trapezoids are $\Delta^+_{0,\nu}$ and $\Delta^-_{1,\nu}$.

Remark 2.19 Even if $z \in \operatorname{Int}\Delta^+_0 \cup \operatorname{Int}\Delta^-_1$, where Int means the interior, both sides of (2-11) are defined from Remark 2.18. However, if $z \in \operatorname{Int}\Delta^+_0$, then from the proof above,

$$T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right)=\mathcal{L}_1(z+1)+\mathcal{L}_1\left(\frac{N}{\gamma}z\right)-\mathcal{L}_1\left(\frac{N}{\gamma}z+1\right)=\mathcal{L}_1(z)-2\pi\sqrt{-1},$$

where the second equality follows from Corollary 2.6 since $\operatorname{Im} z > 0$ and $\operatorname{Im}(N/\gamma) < 0$. Similarly, if $z \in \operatorname{Int}\Delta^-_1$, we have

$$T_N\left(z-\frac{\gamma}{2N}\right)-T_N\left(z+\frac{\gamma}{2N}\right)=\mathcal{L}_1(z-1)-\mathcal{L}_1\left(\frac{N}{\gamma}(z-1)\right)+\mathcal{L}_1\left(\frac{N}{\gamma}(z-1)+1\right)=\mathcal{L}_1(z)-2\pi\sqrt{-1},$$

since $\operatorname{Im} z < 0$ and $\operatorname{Im}((N/\gamma)(z-1)) > 0$.

For a real number $0 < \nu < \frac{1}{2}$ and a positive real number M, we put

$$\Omega^*_\nu := \{z \in \mathbb{C} \mid -1+\nu \le \operatorname{Re} z \le 2-\nu, |\operatorname{Im} z| \le M\} \setminus (\Delta^+_{0,\nu} \cup \Delta^-_{1,\nu}), \tag{2-21}$$

where we put

$$\Delta^+_{0,\nu} := \left\{z \in \mathbb{C} \;\middle|\; -1+\nu \le \operatorname{Re} z < \nu,\ \operatorname{Im} z > -\nu,\ \text{and}\ \operatorname{Im}\left(\frac{z-\nu}{\gamma}\right) < 0\right\},$$
$$\Delta^-_{1,\nu} := \left\{z \in \mathbb{C} \;\middle|\; 1-\nu < \operatorname{Re} z \le 2-\nu,\ \operatorname{Im} z < \nu,\ \text{and}\ \operatorname{Im}\left(\frac{z-1+\nu}{\gamma}\right) > 0\right\}.$$

Note that $\Omega^*_\nu \subset \Omega$ if $N > \operatorname{Re}\gamma/(2\nu)$. Note also that $\Delta^+_{0,\nu} \cap \Delta^-_{1,\nu} = \varnothing$ since $\nu < \frac{1}{2}$; see Figure 5.

We can prove that $T_N(z)$ uniformly converges to $N/(2\pi\sqrt{-1}\gamma)\mathcal{L}_2(z)$ in Ω^*_ν. To do that, we prepare several lemmas.

Lemma 2.20 *Let ν and M be positive real numbers with $0 < \nu < \frac{1}{2}$. Then*

$$\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z-\frac{\gamma}{2N}\right)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z+\frac{\gamma}{2N}\right)=\mathcal{L}_1(z)+O(N^{-2})$$

as $N\to\infty$ for z in the region

$$\{z\in\mathbb{C}\mid -1+\nu\le \operatorname{Re} z\le 2-\nu,\ |\operatorname{Im} z|\le M\}\setminus(\square_\nu^-\cup\square_\nu^+), \tag{2-22}$$

where

$$\square_\nu^-:=\{z\in\mathbb{C}\mid -1+\nu\le\operatorname{Re} z\le\nu,\ |\operatorname{Im} z|\le\nu\},\quad \square_\nu^+:=\{z\in\mathbb{C}\mid 1-\nu\le\operatorname{Re} z\le 2-\nu,\ |\operatorname{Im} z|\le\nu\}.$$

This means that there exists a constant $c>0$ that does not depend on z such that

$$\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z-\frac{\gamma}{2N}\right)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z+\frac{\gamma}{2N}\right)-\mathcal{L}_1(z)\right|<\frac{c}{N^2}$$

for sufficiently large N.

Proof Note that if z is in the region (2-22), then $z\pm\gamma/(2N)$ is also in the same region, assuming that N is large enough. Note also that $\mathcal{L}_1$ and $\mathcal{L}_2$ are holomorphic there.

Since

$$\mathcal{L}_2'(z)=-2\pi\sqrt{-1}\mathcal{L}_1(z),\quad \mathcal{L}_2''(z)=\frac{4\pi^2}{1-e^{-2\pi\sqrt{-1}z}}\quad\text{and}\quad \mathcal{L}_2^{(3)}(z)=2\pi^3\sqrt{-1}\csc^2(\pi z)$$

($\csc x=1/\sin x$ is the cosecant of x, as you may know) from Lemma 2.7, we have

$$\mathcal{L}_2\left(z\pm\frac{\gamma}{2N}\right)=\mathcal{L}_2(z)\mp 2\pi\sqrt{-1}\mathcal{L}_1(z)\frac{\gamma}{2N}+\frac{2\pi^2}{1-e^{-2\pi\sqrt{-1}z}}\frac{\gamma^2}{4N^2}$$
$$\pm\frac{\pi^3\sqrt{-1}}{3\sin^2(\pi z)}\frac{\gamma^3}{8N^3}+\sum_{j=4}^{\infty}\frac{2\pi^3\sqrt{-1}}{j!}\frac{d^{j-3}\csc^2(\pi z)}{dz^{j-3}}\left(\pm\frac{\gamma}{2N}\right)^j$$

if N is large enough that $z\pm\gamma/(2N)$ is contained in the region (2-22). So

$$\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z-\frac{\gamma}{2N}\right)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\left(z+\frac{\gamma}{2N}\right)$$
$$=\mathcal{L}_1(z)-\sum_{k=1}^{\infty}\frac{\pi^2}{(2k+1)!}\frac{d^{2k-2}\csc^2(\pi z)}{dz^{2k-2}}\left(\frac{\gamma}{2N}\right)^{2k}. \tag{2-23}$$

From Lemma 2.21 below, we have

$$\sin^{2k}(\pi z)\frac{d^{2k-2}\csc^2(\pi z)}{dz^{2k-2}}=2\pi^{2k-2}\sum_{j=0}^{k-1}a_{2k-2,2j}\cos(2j\pi z)$$

with $a_{2k-2,2j}>0$ for $j=0,1,\dots,k-1$ and $\sum_{j=0}^{k-1}a_{2k-2,2j}=\frac{1}{2}(2k-1)!$. Letting L be the maximum of $|\cos(z)|$ in the closure of (2-22), we have

$$\left|\sin^{2k}(\pi z)\frac{\pi^2}{(2k+1)!}\frac{d^{2k-2}\csc^2(\pi z)}{dz^{2k-2}}\left(\frac{\gamma}{2N}\right)^{2k}\right|\le\frac{\pi^2}{(2k+1)!}2\pi^{2k-2}Lk\frac{(2k-1)!}{2}\left(\frac{|\gamma|}{2N}\right)^{2k}$$
$$=\frac{L}{2(2k+1)}\left(\frac{\pi|\gamma|}{2N}\right)^{2k}.$$

Let l be the minimum of $|\sin(\pi z)|$ in the closure of the region (2-22). Since the closure is compact and does not contain the zeros of $\sin(\pi z)$, we conclude that $l > 0$. So

$$N^2\Big|\sum_{k=1}^{\infty}\frac{\pi^2}{(2k+1)!}\frac{d^{2k-2}\csc^2(\pi z)}{dz^{2k-2}}\Big(\frac{\gamma}{2N}\Big)^{2k}\Big| < \frac{L\pi^2|\gamma|^2}{8l^2}\sum_{k=1}^{\infty}\frac{1}{2k+1}\Big(\frac{\pi|\gamma|}{2lN}\Big)^{2k-2},$$

which converges if $N > \pi|\gamma|/(2l)$.

Therefore the right-hand side of (2-23) turns out to be $\mathcal{L}_1(z) + O(N^{-2})$, completing the proof. □

Lemma 2.21 *Let m be a positive integer. The m^{th} derivative of $\csc^2(\pi z)$ can be expressed as*

$$\frac{d^m\csc^2(\pi z)}{dz^m} = 2(-\pi)^m\csc^{m+2}(\pi z)P_m(z),$$

where $P_m(z)$ is of the form

$$P_m(z) = \sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} a_{m,j}\cos(j\pi z)$$

with

(i) $a_{m,j} > 0$ *for* $0 \le j \le m$ *and* $j \equiv m \pmod 2$,

(ii) $\sum_{0\le j\le m,\, j\equiv m \pmod 2} a_{m,j} = \frac{1}{2}(m+1)!$, *and*

(iii) $a_{m,m} = 1$.

Proof First of all, recall that $\csc'(x) = -\cos(x)\csc^2(x)$.

We proceed by induction on m.

For $m = 1$, since $(d/dz)\csc^2(\pi z) = 2\csc(\pi z)(-\pi\cos(\pi z)\csc^2(\pi z)) = -2\pi\csc^3(\pi z)\cos(\pi z)$, we have $P_1(z) = \cos(\pi z)$, which agrees with (i)–(iii).

Suppose that the lemma is true for m. We calculate the $(m+1)^{\text{st}}$ derivative by using the inductive hypothesis for $P_m(z)$. We have

$$\begin{aligned}
&\frac{d^{m+1}\csc^2(\pi z)}{dz^{m+1}}\\
&\quad= 2(-\pi)^m\,\frac{d}{dz}(\csc^{m+2}(\pi z)P_m(z))\\
&\quad= 2(-\pi)^m\big((m+2)\csc^{m+1}(\pi z)(-\pi\cos(\pi z)\csc^2(\pi z))P_m(z) + \csc^{m+2}(\pi z)P_m'(z)\big)\\
&\quad= 2(-\pi)^m\csc^{m+3}(\pi z)[-(m+2)\pi\cos(\pi z)P_m(z) + \sin(\pi z)P_m'(z)]\\
&\quad= 2(-\pi)^m\csc^{m+3}(\pi z)\\
&\qquad\cdot\Big[-(m+2)\pi\cos(\pi z)\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} a_{m,j}\cos(j\pi z) - \sin(\pi z)\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} j\pi a_{m,j}\sin(j\pi z)\Big]\\
&\quad= 2(-\pi)^{m+1}\csc^{m+3}(\pi z)\\
&\qquad\cdot\Big[(m+2)\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} a_{m,j}\cos(\pi z)\cos(j\pi z) + \sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} j a_{m,j}\sin(\pi z)\sin(j\pi z)\Big].
\end{aligned}$$

Now we will calculate the terms inside the square brackets. We write $x := \pi z$. From the product–sum identities, we have

$$\sin(x)\sin(jx) = \tfrac{1}{2}\cos((j-1)x) - \tfrac{1}{2}\cos((j+1)x),$$
$$\cos(x)\cos(jx) = \tfrac{1}{2}\cos((j-1)x) + \tfrac{1}{2}\cos((j+1)x).$$

So we have

$$\begin{aligned}
(2\text{-}24)\quad (m+2) &\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} a_{m,j}\cos(x)\cos(jx) + \sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} j a_{m,j}\sin(x)\sin(jx)\\
&= \frac{1}{2}\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} (m+2)a_{m,j}\big(\cos((j-1)x)+\cos((j+1)x)\big)\\
&\qquad + \frac{1}{2}\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} j a_{m,j}\big(\cos((j-1)x)-\cos((j+1)x)\big)\\
&= \frac{1}{2}\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} \big((m+j+2)a_{m,j}\cos((j-1)x) + (m-j+2)a_{m,j}\cos((j+1)x)\big)\\
&= \frac{1}{2}\sum_{\substack{-1\le k\le m-1\\ k\equiv m+1 \pmod 2}} (m+k+3)a_{m,k+1}\cos(kx) + \frac{1}{2}\sum_{\substack{1\le k\le m+1\\ k\equiv m+1 \pmod 2}} (m-k+3)a_{m,k-1}\cos(kx)\\
&= \begin{cases}
\frac{1}{2}\displaystyle\sum_{\substack{0\le k\le m-1\\ k\equiv m+1 \pmod 2}} ((m+k+3)a_{m,k+1}+(m-k+3)a_{m,k-1})\cos(kx)\\
\qquad\qquad +a_{m,m}\cos((m+1)x) & \text{if } m \text{ is odd},\\
\frac{1}{2}\displaystyle\sum_{\substack{0\le k\le m-1\\ k\equiv m+1 \pmod 2}} ((m+k+3)a_{m,k+1}+(m-k+3)a_{m,k-1})\cos(kx)\\
\qquad\qquad +a_{m,m}\cos((m+1)x)+\frac{1}{2}(m+2)a_{m,0}\cos(x) & \text{if } m \text{ is even}.
\end{cases}
\end{aligned}$$

Therefore we obtain the following recursive formula for $a_{m,k}$:

$$2a_{m+1,k} = \begin{cases} (m+k+3)a_{m,k+1}+(m-k+3)a_{m,k-1} & \text{if } k\neq 1,\\ (m+k+3)a_{m,2}+2(m-k+3)a_{m,0} & \text{if } k=1.\end{cases}$$

Note that this also holds for $k=0$ and $k=m+1$ by putting $a_{m,-1}=a_{m,m+2}=0$. Then, (i) follows since $m-k+3\ge 3$, (iii) follows since $a_{m+1,m+1}=1$, and (ii) follows since the sum of the coefficients in the third expression of (2-24) equals

$$\frac{1}{2}\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} ((m+j+2)a_{m,j}+(m-j+2)a_{m,j}) = (m+2)\sum_{\substack{0\le j\le m\\ j\equiv m \pmod 2}} a_{m,j}. \qquad \square$$

For a real number $\nu>0$, we define the region

$$\bowtie_\nu := \Big\{z\in\mathbb{C}\;\Big|\;\operatorname{Im} z\ge 0, \operatorname{Im}\Big(\frac{z-\nu}{\gamma}\Big)\le 0\Big\}\cup\Big\{z\in\mathbb{C}\;\Big|\;\operatorname{Im} z\le 0, \operatorname{Im}\Big(\frac{z+\nu}{\gamma}\Big)\ge 0\Big\}.$$

See Figure 6.

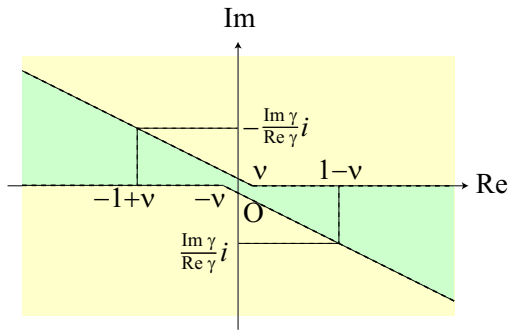


Figure 6: The green region is $\bowtie_\nu$.

Lemma 2.22 *There exist positive real numbers c and ε such that*

$$\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z+1)-\mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big)\right|<ce^{-\varepsilon N}$$

for any z in the region $\mathbb{C}\setminus\bowtie_\nu$ if N is sufficiently large.

Remark 2.23 The left-hand side is defined unless $\operatorname{Im} z=0$ or $z=s\gamma$ ($|s|\geq 1/(2N)$). Therefore if $z\notin\bowtie_\nu$, then the left-hand side is defined.

Proof Note that $\operatorname{Im} z\neq 0$ if $z\notin\bowtie_\nu$.

First, suppose that $\operatorname{Im} z>0$.

Since $\mathcal{L}_2(z)=\mathcal{L}_2(z+1)$ from (2-4), we will prove that $|\mathcal{L}_1\big((N/\gamma)z+\frac{1}{2}\big)|<ce^{-\varepsilon N}$ for some $c>0$ and $\varepsilon>0$. Note that $\operatorname{Im}(z/\gamma)>\operatorname{Im}(\nu/\gamma)=-\nu\operatorname{Im}\gamma/|\gamma|^2>0$ because $z\notin\bowtie_\nu$. So

$$\mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big)=\log(1+e^{2N\pi\sqrt{-1}z/\gamma})$$

from (2-3). Now since $\log(1+x)=\sum_{k=1}^{\infty}(-1)^{k-1}x^k/k$ for $|x|<1$, one has

$$|\log(1+e^a)|\leq\sum_{k=1}^{\infty}\frac{e^{k\operatorname{Re}a}}{k}<\sum_{k=1}^{\infty}e^{k\operatorname{Re}a}=\frac{e^{\operatorname{Re}a}}{1-e^{\operatorname{Re}a}}$$

if $\operatorname{Re}a<0$. Since $\operatorname{Re}(2N\pi\sqrt{-1}z/\gamma)=-2N\pi\operatorname{Im}(z/\gamma)<2N\pi\nu\operatorname{Im}\gamma/|\gamma|^2<0$, we have

$$|\log(1+e^{2N\pi\sqrt{-1}z/\gamma})|<\frac{e^{2N\pi\nu\operatorname{Im}\gamma/|\gamma|^2}}{1-e^{2N\pi\nu\operatorname{Im}\gamma/|\gamma|^2}}<ce^{-\varepsilon N}$$

where we put $\varepsilon:=-2\pi\nu\operatorname{Im}\gamma/|\gamma|^2>0$ and $c:=1/(1-e^{-\varepsilon})>0$.

Next, suppose that $\operatorname{Im} z<0$.

From Corollary 2.6, we have

$$\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z+1)=\frac{2N\pi\sqrt{-1}z}{\gamma}.$$

Since $z \notin \triangleright\!\!\triangleleft_\nu$, we have $\operatorname{Im}(z/\gamma) < -\operatorname{Im}(\nu/\gamma) = \nu \operatorname{Im}\gamma/|\gamma|^2 < 0$. Thus from (2-3) we obtain

$$\mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big) = \log(1+e^{-2N\pi\sqrt{-1}z/\gamma}) + \frac{2N\pi\sqrt{-1}z}{\gamma}.$$

Since $\operatorname{Re}(-2N\pi\sqrt{-1}z/\gamma) = 2N\pi \operatorname{Im}(z/\gamma) < 2N\pi\nu \operatorname{Im}\gamma/|\gamma|^2$, we finally have

$$\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z+1) - \mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big)\right| = |\log(1+e^{-2N\pi\sqrt{-1}z/\gamma})| < ce^{-\varepsilon N}$$

as above, completing the proof. □

The following lemma is similar to [24, Lemma 2.4] and the proof is omitted.

Lemma 2.24 *Let ν and M be positive real numbers. Then there exists a constant $c > 0$ such that*

$$\left|T_N(z) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| = \frac{c}{N}$$

for z in the region $\{z \in \mathbb{C} \mid \nu \le \operatorname{Re} z \le 1-\nu, |\operatorname{Im} z| \le M\}$ if N is sufficiently large, where c does not depend on z.

Now we can prove the following proposition:

Proposition 2.25 *Suppose that $\nu < \frac{1}{4}$. We have*

$$T_N(z) = \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z) + O(N^{-1})$$

as $N \to \infty$ in the region Ω_ν^*.

Proof We need to prove the proposition for z with $-1+\nu \le \operatorname{Re} z < \nu$ or $1-\nu < \operatorname{Re} z \le 2-\nu$.

If $z \in \Omega_\nu^*$ and $-1+\nu \le \operatorname{Re} z < -\nu$, we use (2-15). We have

$$\begin{aligned}
&\left|T_N(z) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| \\
&\quad = \left|T_N(z+1) + \mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| \\
&\quad \le \left|T_N(z+1) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z+1)\right| + \left|\mathcal{L}_1\Big(\frac{N}{\gamma}z+\frac{1}{2}\Big) - \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z) + \frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z+1)\right| \\
&\quad = O(1/N),
\end{aligned}$$

where we apply Lemmas 2.24 and 2.22, noting that we can apply Lemma 2.22 because $z \notin \triangleright\!\!\triangleleft_\nu$.

Similarly, if $z \in \Omega_\nu^*$ and $1+\nu < \operatorname{Re} z \le 2-\nu$, using (2-16), we have

$$\begin{aligned}
&\left|T_N(z)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| \\
&\quad= \left|T_N(z-1)-\mathcal{L}_1\Big(\frac{N}{\gamma}(z-1)+\frac{1}{2}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| \\
&\quad\le \left|T_N(z-1)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z-1)\right|+\left|-\mathcal{L}_1\Big(\frac{N}{\gamma}(z-1)+\frac{1}{2}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)+\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z-1)\right| \\
&\quad= O(1/N),
\end{aligned}$$

noting that we can apply Lemma 2.22 because $z-1 \notin \bowtie_\nu$.

If $z \in \Omega_\nu^*$ and $-\nu \le \operatorname{Re} z < \nu$, we put $m := \lfloor 2N\nu/\operatorname{Re}\gamma \rfloor + 1$. From Lemma 2.17 we have

$$\begin{aligned}
T_N(z) &= T_N\Big(z+\frac{\gamma}{N}\Big)+\mathcal{L}_1\Big(z+\frac{\gamma}{2N}\Big) = T_N\Big(z+\frac{2\gamma}{N}\Big)+\mathcal{L}_1\Big(z+\frac{3\gamma}{2N}\Big)+\mathcal{L}_1\Big(z+\frac{\gamma}{2N}\Big) \\
&= \cdots = T_N\Big(z+\frac{m\gamma}{N}\Big)+\sum_{j=1}^{m}\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big).
\end{aligned}$$

Now since $m \le 2N\nu/\operatorname{Re}\gamma + 1 < m+1$, we have $\nu < \operatorname{Re}(z+m\gamma/N) < 3\nu + \operatorname{Re}\gamma/N < 1-\nu$ if $N > \operatorname{Re}\gamma/(1-4\nu)$, and so we can apply Lemma 2.24 to $z+m\gamma/N$. We have

$$\begin{aligned}
&\left|T_N(z)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)\right| \\
&\qquad= \left|T_N\Big(z+\frac{m\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)+\sum_{j=1}^{m}\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big)\right| \\
&\qquad\le \left|T_N\Big(z+\frac{m\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{m\gamma}{N}\Big)\right| \\
&\qquad\qquad+\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{m\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)+\sum_{j=1}^{m}\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big)\right| \\
&\qquad= \left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{m\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)+\sum_{j=1}^{m}\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big)\right|+O(1/N).
\end{aligned}$$

Since

$$\mathcal{L}_2\Big(z+\frac{m\gamma}{N}\Big)-\mathcal{L}_2(z) = \sum_{j=1}^{m}\Big(\mathcal{L}_2\Big(z+\frac{j\gamma}{N}\Big)-\mathcal{L}_2\Big(z+\frac{(j-1)\gamma}{N}\Big)\Big),$$

we have

$$\begin{aligned}
&\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{m\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2(z)+\sum_{j=1}^{m}\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big)\right| \\
&\qquad\le \sum_{j=1}^{m}\left|\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{j\gamma}{N}\Big)-\frac{N}{2\pi\sqrt{-1}\gamma}\mathcal{L}_2\Big(z+\frac{(j-1)\gamma}{N}\Big)+\mathcal{L}_1\Big(z+\frac{(2j-1)\gamma}{2N}\Big)\right|.
\end{aligned} \tag{2-25}$$

We use Lemma 2.20 to conclude that each summand of the right-hand side of (2-25) is less than c/N^2 for $c>0$. Note that c is independent of j. Since $m=\lfloor 2N\nu/\operatorname{Re}\gamma\rfloor+1$, the right-hand side of (2-25) is less than

$$\frac{mc}{N^2}\le\Big(\frac{2N\nu}{\operatorname{Re}\gamma}+1\Big)\frac{c}{N^2}\le\frac{c'}{N}$$

if we put $c':=(2c\nu/\operatorname{Re}\gamma+1)$.

If $1-\nu<\operatorname{Re}z\le 1+\nu$, from Lemma 2.17 we have

$$\begin{aligned}T_N(z)&=T_N\Big(z-\frac{\gamma}{N}\Big)-\mathcal{L}_1\Big(z-\frac{\gamma}{2N}\Big)=T_N\Big(z-\frac{2\gamma}{N}\Big)-\mathcal{L}_1\Big(z-\frac{3\gamma}{2N}\Big)-\mathcal{L}_1\Big(z-\frac{\gamma}{2N}\Big)\\&=\cdots=T_N\Big(z-\frac{m\gamma}{N}\Big)-\sum_{j=1}^{m}\mathcal{L}_1\Big(z-\frac{(2j-1)\gamma}{2N}\Big),\end{aligned}$$

where we put $m:=\lfloor 2N\nu/\operatorname{Re}\gamma\rfloor+1$ as before. Since $\nu<\operatorname{Re}(z-m\gamma/N)<1-\nu$ as before, we can prove the proposition similarly. □

3 The colored Jones polynomial

In this section, we show several results following [24].

First of all, we recall the following formula due to Habiro [10, page 36, (1)] and Le [18, 1.2.2 Example, page 129] (see also [20, Theorem 5.1]):

$$\begin{aligned}J_N(\mathscr{E};q)&=\sum_{k=0}^{N-1}\prod_{l=1}^{k}(q^{(N-l)/2}-q^{-(N-l)/2})(q^{(N+l)/2}-q^{-(N+l)/2})\\&=\sum_{k=0}^{N-1}q^{-kN}\prod_{l=1}^{k}(1-q^{N-l})(1-q^{N+l}).\end{aligned}\tag{3-1}$$

For a positive integer p, we put $\xi:=\kappa+2p\pi\sqrt{-1}$, where $\kappa:=\operatorname{arccosh}\big(\frac{3}{2}\big)$. We will study the asymptotic behavior of

$$J_N(\mathscr{E};e^{\xi/N})=\sum_{k=0}^{N-1}\prod_{l=1}^{k}e^{-k\xi}(1-e^{(N-l)\xi/N})(1-e^{(N+l)\xi/N})$$

as $N\to\infty$.

We can express $J_N(\mathscr{E};e^{\xi/N})$ in terms of T_N, putting $\gamma:=\xi/(2\pi\sqrt{-1})$, similarly to [24, Section 3, (3.2)]. We have

$$J_N(\mathscr{E};e^{\xi/N})=(1-e^{-4pN\pi^2/\xi})\sum_{m=0}^{p-1}\Big(\beta_{p,m}\sum_{mN/p<k\le(m+1)N/p}\exp\Big(Nf_N\Big(\frac{2k+1}{2N}-\frac{m}{\gamma}\Big)\Big)\Big)\tag{3-2}$$

since $2\sinh\big(\frac{1}{2}\kappa\big)=1$, where we put

$$\beta_{p,m}:=e^{-4mpN\pi^2/\xi}\prod_{j=1}^{m}(1-e^{4(p-j)N\pi^2/\xi})(1-e^{4(p+j)N\pi^2/\xi}),\tag{3-3}$$

$$f_N(z):=\frac{1}{N}T_N(\gamma(1-z)-p+1)-\frac{1}{N}T_N(\gamma(1+z)-p)-\kappa z-\frac{2p\pi\sqrt{-1}}{\gamma}.\tag{3-4}$$

Lemma 3.1 *The function f_N is defined in the region*

$$\Theta_0 := \left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p} + \frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p} - \frac{1}{2N} \right\} \setminus (\underline{\nabla}_0^+ \cup \underline{\nabla}_0^- \cup \overline{\nabla}_0^+ \cup \overline{\nabla}_0^-),$$

where we put

$$\begin{aligned}
\underline{\nabla}_0^+ &:= \left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p} + \frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le 0,\ \mathrm{Re}(\xi z) \le \kappa,\ \mathrm{Im}\, z \le -\frac{2p\pi\kappa}{|\xi|^2} \right\},\\
\underline{\nabla}_0^- &:= \left\{ z \in \mathbb{C} \;\middle|\; \frac{1}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p} - \frac{1}{2N},\ \mathrm{Re}(\xi z) \ge \kappa,\ \mathrm{Im}\, z \ge \frac{2(1-p)\pi\kappa}{|\xi|^2} \right\},\\
\overline{\nabla}_0^+ &:= \left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p} + \frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le 0,\ \mathrm{Re}(\xi z) \le -\kappa,\ \mathrm{Im}\, z \le \frac{2p\pi\kappa}{|\xi|^2} \right\},\\
\overline{\nabla}_0^- &:= \left\{ z \in \mathbb{C} \;\middle|\; \frac{1}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p} - \frac{1}{2N},\ \mathrm{Re}(\xi z) \ge -\kappa,\ \mathrm{Im}\, z \ge \frac{2(p+1)\pi\kappa}{|\xi|^2} \right\}.
\end{aligned}$$

See Figure 7, where we put

$$\begin{aligned}
\overline{K} &:= \left\{ z \in \mathbb{C} \mid z = \frac{2\pi\sqrt{-1}}{\xi} t - 1,\ t \in \mathbb{R} \right\} = \{ z \in \mathbb{C} \mid \mathrm{Re}(\xi z) = -\kappa \},\\
\underline{K} &:= \left\{ z \in \mathbb{C} \mid z = \frac{2\pi\sqrt{-1}}{\xi} t + 1,\ t \in \mathbb{R} \right\} = \{ z \in \mathbb{C} \mid \mathrm{Re}(\xi z) = \kappa \},\\
L_s &:= \{ z \in \mathbb{C} \mid \mathrm{Im}(\xi z) = 2s\pi \}.
\end{aligned}$$

Proof Recall that the function T_N is defined in Ω; see (2-18).

Since $\gamma = \xi/(2\pi\sqrt{-1}) = (\kappa + 2p\pi\sqrt{-1})/(2\pi\sqrt{-1})$, we have $\mathrm{Re}\,\gamma = p$, $\mathrm{Im}\,\gamma = -\kappa/(2\pi)$, and

$$\mathrm{Re}(\gamma(1 \pm z)) = p \pm \frac{\mathrm{Im}(\xi z)}{2\pi}, \quad \mathrm{Im}(\gamma(1 \pm z)) = -\frac{\kappa}{2\pi} \mp \frac{\mathrm{Re}(\xi z)}{2\pi}.$$

Therefore

$$\begin{aligned}
-1 + \frac{p}{2N} < \mathrm{Re}(\gamma(1-z) - p + 1) < 2 - \frac{p}{2N} &\iff -\frac{1}{p} + \frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p} - \frac{1}{2N}\\
&\iff -1 + \frac{p}{2N} < \mathrm{Re}(\gamma(1+z) - p) < 2 - \frac{p}{2N}.
\end{aligned}$$

We can also see that the condition $\gamma(1-z) - p + 1 \in \Delta_0^+$ is equivalent to $z \in \underline{\nabla}_0^-$, that the condition $\gamma(1-z) - p + 1 \in \Delta_1^-$ is equivalent to $z \in \underline{\nabla}_0^+$, that the condition $\gamma(1+z) - p \in \Delta_0^+$ is equivalent to $z \in \overline{\nabla}_0^+$, and that the condition $\gamma(1+z) - p \in \Delta_1^-$ is equivalent to $z \in \overline{\nabla}_0^-$. □

We would like to approximate $f_N(z)$ by using $\mathcal{L}_2$. From Proposition 2.25 and (3-4), the series of functions $\{f_N(z)\}$ converges uniformly to

$$F(z) := \frac{1}{\xi}\left(\mathcal{L}_2\left(\frac{\xi(1-z)}{2\pi\sqrt{-1}} - p + 1 \right) - \mathcal{L}_2\left(\frac{\xi(1+z)}{2\pi\sqrt{-1}} - p \right) \right) - \kappa z + \frac{4p\pi^2}{\xi}$$

in the region

$$\left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p} + \frac{\nu}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} \le \frac{2}{p} - \frac{\nu}{p},\ |\mathrm{Re}(\xi z)| \le 2M\pi - \kappa \right\} \setminus (\underline{\nabla}_{0,\nu}^+ \cup \underline{\nabla}_{0,\nu}^- \cup \overline{\nabla}_{0,\nu}^+ \cup \overline{\nabla}_{0,\nu}^-), \tag{3-5}$$

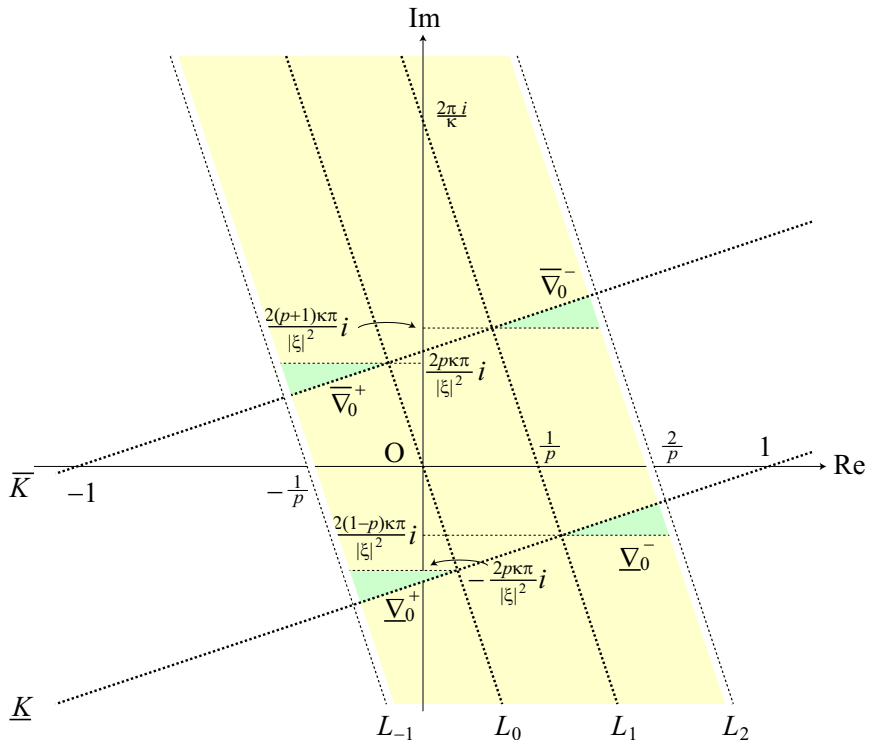


Figure 7: The function f_N is defined in the yellow region Θ_0. The green triangles are $\underline{\nabla}_0^+$, $\underline{\nabla}_0^-$, $\overline{\nabla}_0^+$, and $\overline{\nabla}_0^-$.

where we put

$$\underline{\nabla}_{0,\nu}^+ := \left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p}+\frac{\nu}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{\nu}{p},\ \mathrm{Re}(\xi z) < \kappa + 2\pi\nu,\ \mathrm{Im}\, z < -\frac{2(p-\nu)\pi\kappa}{|\xi|^2} \right\},$$

$$\underline{\nabla}_{0,\nu}^- := \left\{ z \in \mathbb{C} \;\middle|\; \frac{1}{p}-\frac{\nu}{p} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le \frac{2}{p}-\frac{\nu}{p},\ \mathrm{Re}(\xi z) > \kappa - 2\pi\nu,\ \mathrm{Im}\, z > \frac{2(1-p-\nu)\pi\kappa}{|\xi|^2} \right\},$$

$$\overline{\nabla}_{0,\nu}^+ := \left\{ z \in \mathbb{C} \;\middle|\; -\frac{1}{p}+\frac{\nu}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{\nu}{p},\ \mathrm{Re}(\xi z) < -\kappa + 2\pi\nu,\ \mathrm{Im}\, z < \frac{2(p+\nu)\pi\kappa}{|\xi|^2} \right\},$$

$$\overline{\nabla}_{0,\nu}^- := \left\{ z \in \mathbb{C} \;\middle|\; \frac{1}{p}-\frac{\nu}{p} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le \frac{2}{p}-\frac{\nu}{p},\ \mathrm{Re}(\xi z) > -\kappa - 2\pi\nu,\ \mathrm{Im}\, z > \frac{2(p+1-\nu)\pi\kappa}{|\xi|^2} \right\}.$$

Lemma 3.2 *The series of functions $\{f_N(z)\}$ uniformly converges to $F(z)$ in the region (3-5).*

Proof In a way similar to the proof of Lemma 3.1, we have

$$\begin{aligned} -1+\nu < \mathrm{Re}(\gamma(1-z)-p+1) < 2-\nu \;&\Longleftrightarrow\; -\frac{1}{p}+\frac{\nu}{p} < \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p}-\frac{\nu}{p} \\ &\Longleftrightarrow\; -1+\nu < \mathrm{Re}(\gamma(1+z)-p) < 2-\nu, \\ |\mathrm{Im}(\gamma(1\pm z))| \le M \;&\Longleftrightarrow\; \kappa - 2M\pi \le \mp\,\mathrm{Re}(\xi z) \le \kappa + 2M\pi, \end{aligned}$$

and

$$\gamma(1-z)-p+1\in\Delta^+_{0,\nu} \iff z\in\underline{\nabla}^-_{0,\nu},\quad \gamma(1-z)-p+1\in\Delta^-_{1,\nu} \iff z\in\underline{\nabla}^+_{0,\nu},$$
$$\gamma(1+z)-p\in\Delta^+_{0,\nu} \iff z\in\overline{\nabla}^+_{0,\nu},\qquad \gamma(1+z)-p\in\Delta^-_{1,\nu} \iff z\in\overline{\nabla}^-_{0,\nu}.$$

Then the lemma follows from Proposition 2.25. □

We can express $F(z)$ in terms of Li_2 for certain cases.

Lemma 3.3 *If z is in between $\overline{K}$ and $\underline{K}$, or between L_0 and L_1, then we have*

$$F(z)=\frac{1}{\xi}\mathrm{Li}_2(e^{-\xi(1+z)})-\frac{1}{\xi}\mathrm{Li}_2(e^{-\xi(1-z)})+\kappa z-2\pi\sqrt{-1}. \tag{3-6}$$

Moreover, if z is between L_0 and L_1, we also have

$$F(z)=\frac{1}{\xi}\mathrm{Li}_2(e^{\xi(1-z)})-\frac{1}{\xi}\mathrm{Li}_2(e^{\xi(1+z)})-\kappa z+\frac{4p\pi^2}{\xi}. \tag{3-7}$$

Proof Since $\mathrm{Im}(\xi(1\pm z)/(2\pi\sqrt{-1}))=(-1/2\pi)(\kappa\pm\mathrm{Re}(\xi z))$, we see that $\mathrm{Im}(\xi(1+z)/(2\pi\sqrt{-1}))<0$ and $\mathrm{Im}(\xi(1-z)/(2\pi\sqrt{-1}))<0$ if z is between $\overline{K}$ and $\underline{K}$. Thus, in this case, we have (3-6) from (2-6).

Next, we consider the case where z is between L_0 and L_1, that is, where $0<\mathrm{Im}(\xi z)<2\pi$.

We have $\mathrm{Re}(\xi(1-z)/(2\pi\sqrt{-1}))-p+1=1-\mathrm{Im}(\xi z)/(2\pi)$ and $\mathrm{Re}(\xi(1+z)/(2\pi\sqrt{-1}))-p=\mathrm{Im}(\xi z)/(2\pi)$, both of which are between 0 and 1. So, from Lemmas 2.3 and 2.5, we have (3-7).

Now we will show that (3-6) also holds in this case.

From (2-7), we have

$$\begin{aligned}\mathrm{Li}_2(e^{\xi(1-z)})&=-\mathrm{Li}_2(e^{-\xi(1-z)})-\tfrac{1}{6}\pi^2-\tfrac{1}{2}(\log(-e^{-\xi(1-z)}))^2\\&=-\mathrm{Li}_2(e^{-\xi(1-z)})-\tfrac{1}{6}\pi^2-\tfrac{1}{2}(-\xi(1-z)+(2p-1)\pi\sqrt{-1})^2\end{aligned}$$

since $\mathrm{Im}\,\xi(1-z)=2p\pi-\mathrm{Im}(\xi z)$, which is between $2(p-1)\pi$ and $2p\pi$ when $0<\mathrm{Im}(\xi z)<2\pi$, that is, when z is between L_0 and L_1. Similarly,

$$\begin{aligned}\mathrm{Li}_2(e^{\xi(1+z)})&=-\mathrm{Li}_2(e^{-\xi(1+z)})-\tfrac{1}{6}\pi^2-\tfrac{1}{2}(\log(-e^{-\xi(1+z)}))^2\\&=-\mathrm{Li}_2(e^{-\xi(1+z)})-\tfrac{1}{6}\pi^2-\tfrac{1}{2}(-\xi(1+z)+(2p+1)\pi\sqrt{-1})^2\end{aligned}$$

since $\mathrm{Im}\,\xi(1+z)=2p\pi+\mathrm{Im}(\xi z)$, which is between $2p\pi$ and $2(p+1)\pi$. Thus, from (3-7), we obtain (3-6), completing the proof. □

The derivatives of $F(z)$ are given as follows from Lemma 2.7:

$$F'(z)=\mathcal{L}_1\left(\frac{\xi(1-z)}{2\pi\sqrt{-1}}-p+1\right)+\mathcal{L}_1\left(\frac{\xi(1+z)}{2\pi\sqrt{-1}}-p\right)-\kappa, \tag{3-8}$$

$$F''(z)=\frac{\xi(e^{-\xi z}-e^{\xi z})}{3-e^{\xi z}-e^{-\xi z}}, \tag{3-9}$$

$$F^{(3)}(z)=\frac{\xi^2(4-3(e^{\xi z}+e^{-\xi z}))}{(3-e^{\xi z}-e^{-\xi z})^2}. \tag{3-10}$$

If z is between $\overline{K}$ and $\underline{K}$, or between L_0 and L_1, we have

$$F'(z) = \log(1-e^{-\kappa-\xi z}) + \log(1-e^{-\kappa+\xi z}) + \kappa = \log(3-e^{\xi z}-e^{-\xi z}) \tag{3-11}$$

from Lemma 3.3, where the second equality follows from the same reason as [24, (4.2)].

Put $\sigma_0 := 2\pi\sqrt{-1}/\xi = (2\pi/|\xi|^2)(2p\pi + \kappa\sqrt{-1})$. Since $\mathrm{Re}(\xi\sigma_0) = 0$ and $\mathrm{Im}(\xi\sigma_0) = 2\pi$, we conclude that σ_0 is on L_1 and between $\overline{K}$ and $\underline{K}$. From (3-6), (3-11), (3-9), and (3-10) we have

$$F(\sigma_0) = \frac{4p\pi^2}{\xi}, \quad F'(\sigma_0) = 0, \quad F''(\sigma_0) = 0, \quad F^{(3)}(\sigma_0) = -2\xi^2. \tag{3-12}$$

4 The Poisson summation formula

In (3-2), we put $\varphi_{m,N}(z) := f_N(z - 2m\pi\sqrt{-1}/\xi)$ for $m = 0, 1, 2, \dots, p-1$ so that

$$J_N(\mathscr{E}; e^{\xi/N}) = (1-e^{-4pN\pi^2/\xi}) \sum_{m=0}^{p-1} \left(\beta_{p,m} \sum_{mN/p<k\le(m+1)N/p} \exp\left(N\varphi_{m,N}\left(\frac{2k+1}{2N}\right)\right)\right). \tag{4-1}$$

Note that the function $\varphi_{m,N}(z)$ is defined in the region

$$\Theta_m := \left\{ z\in\mathbb{C} \;\middle|\; -\frac{1}{p}+\frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{2}{p}-\frac{1}{2N} \right\} \setminus (\underline{\nabla}_m^+ \cup \underline{\nabla}_m^- \cup \overline{\nabla}_m^+ \cup \overline{\nabla}_m^-)$$

from Lemma 3.1, where we put

$$\underline{\nabla}_m^+ := \left\{ z\in\mathbb{C} \;\middle|\; \frac{m-1}{p}+\frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le \frac{m}{p},\ \mathrm{Re}(\xi z)\le\kappa,\ \mathrm{Im}\, z \le \frac{2(m-p)\pi\kappa}{|\xi|^2} \right\},$$

$$\underline{\nabla}_m^- := \left\{ z\in\mathbb{C} \;\middle|\; \frac{m+1}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{m+2}{p}-\frac{1}{2N},\ \mathrm{Re}(\xi z)\ge\kappa,\ \mathrm{Im}\, z \ge \frac{2(m-p+1)\pi\kappa}{|\xi|^2} \right\},$$

$$\overline{\nabla}_m^+ := \left\{ z\in\mathbb{C} \;\middle|\; \frac{m-1}{p}+\frac{1}{2N} < \frac{\mathrm{Im}(\xi z)}{2p\pi} \le \frac{m}{p},\ \mathrm{Re}(\xi z)\le-\kappa,\ \mathrm{Im}\, z \le \frac{2(m+p)\pi\kappa}{|\xi|^2} \right\},$$

$$\overline{\nabla}_m^- := \left\{ z\in\mathbb{C} \;\middle|\; \frac{m+1}{p} \le \frac{\mathrm{Im}(\xi z)}{2p\pi} < \frac{m+2}{p}-\frac{1}{2N},\ \mathrm{Re}(\xi z)\ge-\kappa,\ \mathrm{Im}\, z \ge \frac{2(m+p+1)\pi\kappa}{|\xi|^2} \right\}.$$

We would like to show that the sum

$$\sum_{mN/p<k\le(m+1)N/p} \exp\left(N\varphi_{m,N}\left(\frac{2k+1}{2N}\right)\right)$$

is approximated by the integral

$$N\int_{m/p}^{(m+1)/p} e^{N\varphi_{m,N}(z)}\,dz.$$

To do that, we use the following proposition, known as the Poisson summation formula:

Proposition 4.1 *Let a and b be real numbers with $a < b$, and $\{\psi_N(z)\}_{N=1,2,3,\dots}$ be a series of holomorphic functions in a domain $D \subset \mathbb{C}$ containing the closed interval $[a,b]$. We assume that $\psi_N(z)$ uniformly converges to a holomorphic function $\psi(z)$ in D. We also assume that $\operatorname{Re}\psi(a) < 0$ and $\operatorname{Re}\psi(b) < 0$.*

Putting $R_+ := \{z \in D \mid \operatorname{Im} z \geq 0, \operatorname{Re}\psi(z) < 2\pi \operatorname{Im} z\}$ and $R_- := \{z \in D \mid \operatorname{Im} z \leq 0, \operatorname{Re}\psi(z) < -2\pi \operatorname{Im} z\}$, we also assume that there are paths $C_\pm$ connecting a and b such that $C_\pm \subset R_\pm$ and that $C_\pm$ is homotopic to $[a,b]$ in D with a and b fixed.

Then we have

$$\frac{1}{N}\sum_{a \leq k/N \leq b} e^{N\psi_N(k/N)} = \int_a^b e^{N\psi_N(z)}\,dz + O(e^{-\varepsilon N})$$

for some $\varepsilon > 0$ independent of N.

A proof, which is essentially the same as that of [30, Proposition 4.2], is given in Appendix A.

From Lemma 3.2, the series of functions $\{\varphi_{m,N}(z)\}$ uniformly converges to $\Phi_m(z) := F(z - 2m\pi\sqrt{-1}/\xi)$ in the region $\Theta^*_{m,\nu}$ defined as

$$\Theta^*_{m,\nu} := \{z \in \mathbb{C} \mid 2(m-1+\nu)\pi \leq \operatorname{Im}(\xi z) \leq 2(m+2-\nu)\pi,\ |\operatorname{Re}(\xi z)| \leq 2M\pi - \kappa\} \setminus (\nabla^+_{m,\nu} \cup \underline{\nabla}^-_{m,\nu} \cup \overline{\nabla}^+_{m,\nu} \cup \overline{\nabla}^-_{m,\nu}), \tag{4-2}$$

where we put

$$\nabla^+_{m,\nu} := \{z \in \mathbb{C} \mid 2(m-1+\nu)\pi \leq \operatorname{Im}(\xi z) < 2(m+\nu)\pi,\ \operatorname{Re}(\xi z) < \kappa + 2\pi\nu,\ \operatorname{Im} z < 2(\nu - p + m)\pi\kappa/|\xi|^2\},$$

$$\underline{\nabla}^-_{m,\nu} := \{z \in \mathbb{C} \mid 2(m+1-\nu)\pi < \operatorname{Im}(\xi z) \leq 2(m+2-\nu)\pi,\ \operatorname{Re}(\xi z) \geq \kappa - 2\pi\nu,\ \operatorname{Im} z > 2(1-p+m-\nu)\pi\kappa/|\xi|^2\},$$

$$\overline{\nabla}^+_{m,\nu} := \{z \in \mathbb{C} \mid 2(m-1+\nu)\pi \leq \operatorname{Im}(\xi z) < 2(m+\nu)\pi,\ \operatorname{Re}(\xi z) < -\kappa + 2\pi\nu,\ \operatorname{Im} z < 2(\nu + p + m)\pi\kappa/|\xi|^2\},$$

$$\overline{\nabla}^-_{m,\nu} := \{z \in \mathbb{C} \mid 2(m+1-\nu)\pi < \operatorname{Im}(\xi z) \leq 2(m+2-\nu)\pi,\ \operatorname{Re}(\xi z) > -\kappa - 2\pi\nu,\ \operatorname{Im} z > 2(p+m+1-\nu)\pi\kappa/|\xi|^2\},$$

and we always assume that N is sufficiently large. From (3-8)–(3-10), we have

$$\Phi'_m(z) = \mathcal{L}_1\left(\frac{\xi(1-z)}{2\pi\sqrt{-1}} + m - p + 1\right) + \mathcal{L}_1\left(\frac{\xi(1+z)}{2\pi\sqrt{-1}} - m - p\right) - \kappa, \tag{4-3}$$

$$\Phi''_m(z) = \frac{\xi(e^{-\xi z} - e^{\xi z})}{3 - e^{\xi z} - e^{-\xi z}}, \tag{4-4}$$

$$\Phi^{(3)}_m(z) = \frac{\xi^2(4 - 3(e^{\xi z} + e^{-\xi z}))}{(3 - e^{\xi z} - e^{-\xi z})^2}. \tag{4-5}$$

Since $z - 2m\pi\sqrt{-1}/\xi$ is between L_0 and L_1 ($\overline{K}$ and $\underline{K}$, respectively) if and only if z is between L_m and L_{m+1} ($\overline{K}$ and $\underline{K}$, respectively), from (3-11) we have

$$\Phi'_m(z) = \log(3 - e^{\xi z} - e^{-\xi z}) \tag{4-6}$$

when z is between $\overline{K}$ and $\underline{K}$, or L_m and L_{m+1}.

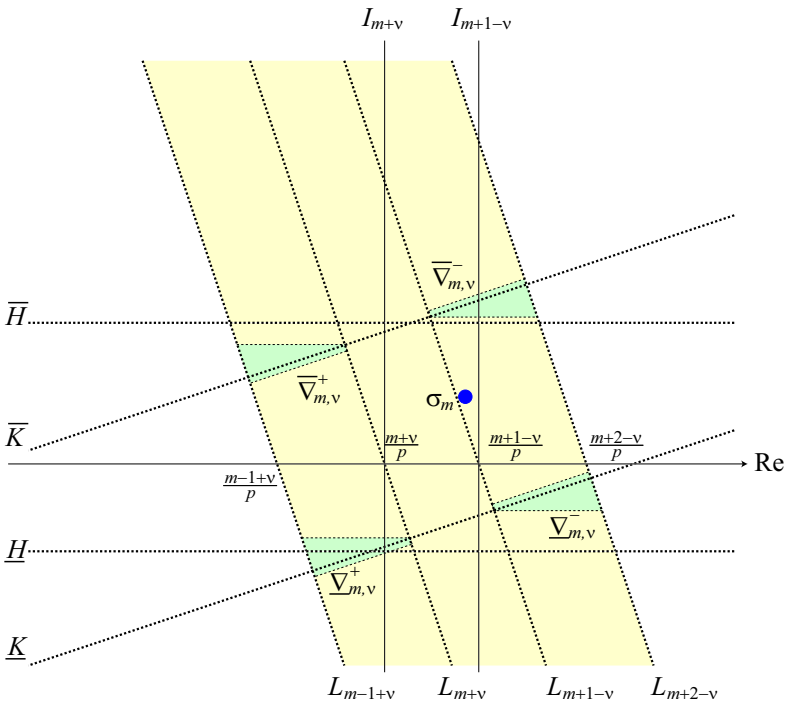


Figure 8: The yellow region is $\Theta^*_{m,\nu}$. The blue point is σ_m. The green trapezoids are $\underline{\nabla}^+_{m,\nu}$, $\underline{\nabla}^-_{m,\nu}$, $\overline{\nabla}^+_{m,\nu}$, and $\overline{\nabla}^-_{m,\nu}$.

We also put $\sigma_m := \sigma_0 + 2m\pi\sqrt{-1}/\xi = 2(m+1)\pi\sqrt{-1}/\xi = (2(m+1)\pi/|\xi|^2)(2p\pi + \kappa\sqrt{-1})$ so that

$$\Phi_m(\sigma_m) = \frac{4p\pi^2}{\xi}, \quad \Phi'_m(\sigma_m) = 0, \quad \Phi''_m(\sigma_m) = 0, \quad \Phi^{(3)}_m(\sigma_m) = -2\xi^2 \tag{4-7}$$

from (3-12). Since $\mathrm{Re}(\xi\sigma_m) = 0$ and $\mathrm{Im}(\xi\sigma_m) = 2(m+1)\pi$, we see that σ_m is between $\overline{K}$ and $\underline{K}$ and on the line L_{m+1}; see Figure 8.

Let I_s be the vertical line $\mathrm{Re}\, z = s/p$ for $s \in \mathbb{R}$.

For a small number $\chi > 0$, let $\Xi_{m,\chi}$ be the pentagonal region defined as

$$\Xi_{m,\chi} := \left\{ z \in \mathbb{C} \;\middle|\; \frac{m-\chi}{p} < \mathrm{Re}\, z < \frac{m+1+\chi}{p}, -\frac{2(m+1)\kappa\pi}{|\xi|^2} < \mathrm{Im}\, z < \frac{(p+m)\kappa}{2p^2\pi}, \right.$$
$$\left. \mathrm{Im}(\xi z) + \frac{(2\chi+1)|\xi|^2}{2(m+1)\kappa}\,\mathrm{Im}\, z > 2(m-\chi)\pi \right\}$$

when $m < p-1$, and

$$\Xi_{p-1,\chi} := \left\{ z \in \mathbb{C} \;\middle|\; \frac{p-1-\chi}{p} < \mathrm{Re}\, z < \frac{p+\chi}{p}, -\frac{2p\kappa\pi}{|\xi|^2} < \mathrm{Im}\, z < \frac{(2p-1)\kappa}{2p^2\pi}, \right.$$
$$\left. \mathrm{Im}(\xi z) + \frac{(2\chi+1)|\xi|^2}{2p\kappa}\,\mathrm{Im}\, z > 2(p-1-\chi)\pi \right\} \setminus \diamond_\nu,$$

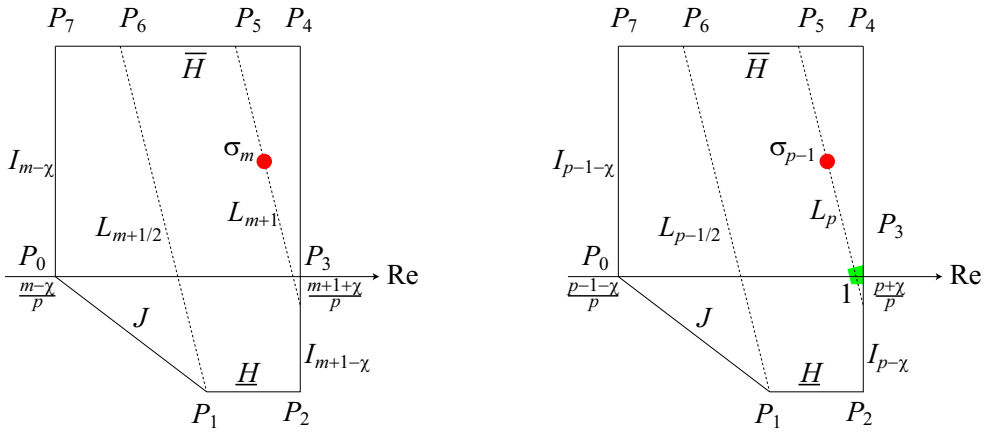


Figure 9: The region $\Xi_{m,\chi}$ when $m < p-1$ (left) and the region $\Xi_{p-1,\chi}$ (right), where the green quadrilateral indicates $\diamond_\nu$. Precisely speaking, the points P_0 and P_7 should be a little more to the right than indicated, and the points P_2, P_3, and P_4 should be a little more to the left than indicated.

where we put

$$\diamond_\nu := \{z \in \mathbb{C} \mid \operatorname{Re} z < 1 + \chi/p\} \cap \nabla^-_{p-1,\nu}.$$

Note that $\Xi_{m,\chi}$ $(m < p-1)$ is surrounded by $I_{m-\chi}$, $I_{m+1+\chi}$, $\overline{H}$, $\underline{H}$, and J, where $\overline{H}$ and $\underline{H}$ are the horizontal lines $\operatorname{Im} z = (p+m)\kappa/(2p^2\pi)$ and $\operatorname{Im} z = -\operatorname{Im}\sigma_m$, respectively, and J is the line connecting $(m-\chi)/p$ and $L_{m+1/2} \cap \underline{H}$, which is given as

$$J := \left\{ z \in \mathbb{C} \;\middle|\; \operatorname{Im}(\xi z) + \frac{(2\chi+1)|\xi|^2}{2(m+1)\kappa} \operatorname{Im} z = 2(m-\chi)\pi \right\}. \tag{4-8}$$

See Figure 9, left. Figure 9, right, indicates $\Xi_{p-1,\chi}$, where $\diamond_\nu$ is indicated by the green quadrilateral. Note that it is a neighborhood of the point 1.

Lemma 4.2 *If $\nu > 0$ is sufficiently small, then we can choose $\chi > 0$ so that $\Xi_{m,\chi}$ is included in $\Theta^*_{m,\nu}$ for $m = 0, 1, 2, \dots, p-1$.*

Proof First, $\Xi_{m,\chi}$ is in the rectangle surrounded by $I_{m-\chi}$, $I_{m+1+\chi}$, $\overline{H}$, and $\underline{H}$, with bottom left vertex

$$v_1 := (m-\chi)/p - (2(m+1)\kappa\pi/|\xi|^2)\sqrt{-1}$$

and top right vertex

$$v_2 := (m+1+\chi)/p + ((p+m)\kappa/2p^2\pi)\sqrt{-1}.$$

The vertices v_1 and v_2 are on the lines $L_{\operatorname{Im}(\xi v_1)/(2\pi)}$ and v_2 $L_{\operatorname{Im}(\xi v_2)/(2\pi)}$, respectively. Since

$$\frac{\operatorname{Im}(\xi v_1)}{2\pi} - (m-1+\nu) = (1-\chi-\nu) - \frac{(m+1)\kappa^2}{|\xi|^2}$$

and

$$(m+2-\nu)-\frac{\mathrm{Im}(\xi v_2)}{2\pi}=(1-\chi-\nu)-\frac{(p+m)\kappa^2}{4p^2\pi^2},$$

if

$$\nu+\chi<\min\left\{1-(m+1)\Big(\frac{\kappa}{|\xi|}\Big)^2,\,1-(p+m)\Big(\frac{\kappa}{2p\pi}\Big)^2\right\}=1-(p+m)\Big(\frac{\kappa}{2p\pi}\Big)^2, \tag{4-9}$$

then $\Xi_{m,\chi}$ is between the lines $L_{m-1+\nu}$ and $L_{m+2-\nu}$ for $m=0,1,2,\dots,p-1$.

So it remains to show that $\Xi_{m,\chi}$ excludes $\overline{\nabla}^+_{m,\nu}$, $\overline{\nabla}^-_{m,\nu}$, $\underline{\nabla}^+_{m,\nu}$, and $\underline{\nabla}^-_{m,\nu}$.

- The real part of the bottom right corner of $\overline{\nabla}^+_{m,\nu}$ is $(\kappa(2\pi\nu-\kappa)+4(m+\nu)p\pi^2)/|\xi|^2$, which is smaller than $(m-\chi)/p$ if

$$2p\pi(\kappa+2p\pi)\nu+|\xi|^2\chi<(p+m)\kappa^2. \tag{4-10}$$

So the trapezoid $\overline{\nabla}^+_{m,\nu}$ is to the right of $I_{m-\chi}$ if (4-10) holds.

- The difference between the imaginary parts of the bottom line of $\overline{\nabla}^-_{m,\nu}$ and $\overline{H}$ is

$$\frac{2(p+m+1-\nu)\kappa\pi}{|\xi|^2}-\frac{(p+m)\kappa}{2p^2\pi}=\frac{\kappa(4p^2(1-\nu)\pi^2-(p+m)\kappa^2)}{2p^2\pi|\xi|^2},$$

which is positive if

$$\nu<1-(p+m)\Big(\frac{\kappa}{2p\pi}\Big)^2. \tag{4-11}$$

So we conclude that $\overline{\nabla}^-_{m,\nu}$ is outside of $\Xi_{m,\chi}$ if (4-11) holds.

- To obtain a condition ensuring that $\underline{\nabla}^+_{m,\nu}$ is below J, it is enough to find a condition ensuring that the top right corner z_0 of the trapezoid is below J, since $L_{m+\nu}$ is steeper than J. Since $\mathrm{Im}\,z_0=2(\nu-p+m)\kappa\pi/|\xi|^2$ and z_0 is on $L_{m+\nu}$, the condition is

$$2(m+\nu)\pi+\frac{(2\chi+1)|\xi|^2}{2(m+1)\kappa}\frac{2(\nu-p+m)\kappa\pi}{|\xi|^2}<2(m-\chi)\pi$$

from (4-8). Therefore, if

$$2\nu\chi+(2m+3)\nu+2(2m-p+1)\chi<p-m, \tag{4-12}$$

the trapezoid $\underline{\nabla}^+_{m,\nu}$ is out of $\Xi_{m,\chi}$.

- The real part of the top left corner of $\underline{\nabla}^-_{m,\nu}$ is $(\kappa(\kappa-2\pi\nu)+4(m+1-\nu)p\pi^2)/|\xi|^2$, which is bigger than $(m+1+\chi)/p$ if

$$2p\pi(\kappa+2p\pi)\nu+|\xi|^2\chi<(p-m-1)\kappa^2. \tag{4-13}$$

So the trapezoid $\underline{\nabla}^-_{m,\nu}$ is outside of $\Xi_{m,\chi}$ if (4-13) holds.

From (4-10)–(4-13), we conclude that if $m < p-1$ and

$$\nu < \min\left\{ \frac{(p+m)\kappa}{2p\pi(\kappa+2p\pi)}, 1-(p+m)\Big(\frac{\kappa}{2p\pi}\Big)^2, \frac{p-m}{2m+3}, \frac{(p-m-1)\kappa^2}{2p\pi(\kappa+2p\pi)} \right\},$$

then we can choose $\chi > 0$ so that $\Xi_{m,\chi}$ is included in $\Theta^*_{m,\nu}$.

If $m = p-1$, then the real part of the top left corner of $\underline{\nabla}^-_{p-1,\nu}$ is $1-2\pi\nu(\kappa+2p\pi)/|\xi|^2$, which is slightly to the left of $1+\chi/p$. Its imaginary part is $2\pi\nu(2p\pi-\kappa)/|\xi|^2$, which is slightly above the real axis. The bottom left corner of $\underline{\nabla}^-_{p-1,\nu}$ is $1-4p\nu\pi^2/|\xi|^2$, which is slightly smaller than 1. Its imaginary part is $--2\nu\kappa\pi/|\xi|^2$, which is below the real axis. So if we exclude $\underline{\nabla}^-_{p-1,\nu}$, the rest is included in $\Theta^*_{p-1,\nu}$; see Figure 9, right. □

We will show that the assumption of Proposition 4.1 holds for the function $\psi_N(z) := \varphi_{m,N}(z) - \varphi_{m,N}(\sigma_m)$, the domain $D := \Xi_{m,\chi}$, and the numbers $a := m/p$ and $b := (m+1)/p$, with small $\chi > 0$. Note that the series of functions $\{\psi_N(z)\} := \{\varphi_{m,N}(z) - \varphi_{m,N}(\sigma_m)\}$ uniformly converges to $\psi(z) := \Phi_m(z) - \Phi_m(\sigma_m)$ in $\Xi_{m,\chi}$ for sufficiently small $\chi > 0$.

From now we will study properties of $\Phi_m(z)$ in the region $\Xi_{m,\chi}$ as if $\chi = 0$, taking care of the case where $\chi > 0$ if necessary.

Let $P_0, P_1, \dots, P_7$ be points defined as follows, which are already indicated in Figure 9:

$$P_0 := I_m \cap \text{real axis}, \quad P_1 := L_{m+1/2} \cap \underline{H}, \quad P_2 := I_{m+1} \cap \underline{H}, \quad P_3 := I_{m+1} \cap \text{real axis},$$
$$P_4 := I_{m+1} \cap \overline{H}, \quad P_5 := L_{m+1/2} \cap \overline{H}, \quad P_6 := I_m \cap \overline{H}.$$

Their coordinates are given as follows:

$$P_0 := \frac{m}{p}, \qquad P_1 := \frac{m+\frac{1}{2}}{p} + \frac{\operatorname{Im}\sigma_m}{2p\pi}\bar{\xi}, \qquad P_2 := \frac{m+1}{p} - \operatorname{Im}\sigma_m\sqrt{-1},$$
$$P_3 := \frac{m+1}{p}, \qquad P_4 := \frac{m+1}{p} + \frac{(p+m)\kappa}{2p^2\pi}\sqrt{-1}, \quad P_5 := \frac{m+1}{p} - \frac{(p+m)\kappa}{4p^3\pi^2}\bar{\xi},$$
$$P_6 := \frac{m+\frac{1}{2}}{p} - \frac{(p+m)\kappa}{4p^3\pi^2}\bar{\xi}, \quad P_7 := \frac{m}{p} + \frac{(p+m)\kappa}{2p^2\pi}\sqrt{-1}.$$

Lemma 4.3 *We have the following inequalities*:

$$\operatorname{Re} P_6 < \operatorname{Re} P_1 < \operatorname{Re} P_5.$$

Proof It is clear that $\operatorname{Re} P_6 < \operatorname{Re} P_1$, and so we will show the other inequality.

Since $\operatorname{Im}\sigma_m = 2(m+1)\kappa\pi/|\xi|^2$ and $\kappa > 1$, we have

$$\begin{aligned}\operatorname{Re} P_5 - \operatorname{Re} P_1 &= \frac{m+1}{p} - \frac{(p+m)\kappa^2}{4p^3\pi^2} - \left(\frac{2m+1}{2p} + \frac{(m+1)\kappa^2}{p|\xi|^2}\right) \\ &> \frac{1}{2p} - \frac{p+m}{4p^3\pi^2} - \frac{m+1}{4p^3\pi^2} \geq \frac{1}{2p} - \frac{3p-1}{4p^3\pi^2} > \frac{1}{2p} - \frac{3}{4p^2\pi^2} > 0,\end{aligned}$$

proving the inequality $\operatorname{Re} P_5 > \operatorname{Re} P_1$. □

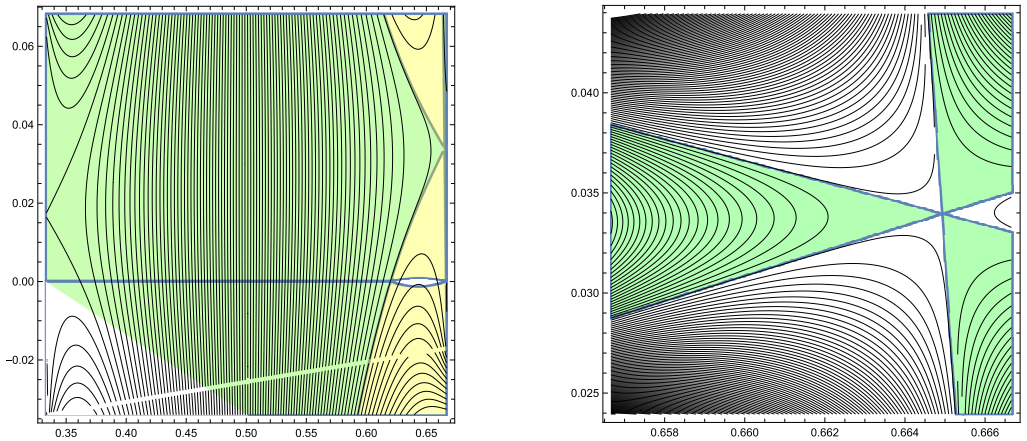


Figure 10: The left picture shows a contour plot of $\operatorname{Re}\Phi_1(z)$ in $\Xi_{1,0}$ for $p=3$, where $R_\pm$ are indicated by yellow and green and W_1^- is indicated by dark green. The right picture shows a contour plot of $\operatorname{Re}\Phi_1(z)$ in a neighborhood of σ_1, where W_1^- is indicated by green.

We put

$$W_m^+ := \{z \in \Xi_{m,\chi} \mid \operatorname{Re}\Phi_m(z) > \operatorname{Re}\Phi_m(\sigma_m)\}, \quad W_m^- := \{z \in \Xi_{m,\chi} \mid \operatorname{Re}\Phi_m(z) < \operatorname{Re}\Phi_m(\sigma_m)\}$$

for $m = 0, 1, 2, \dots, p-1$. Recall that in this case, $R_\pm$ in Proposition 4.1 becomes

$$R_+ := \{z \in \Xi_{m,\chi} \mid \operatorname{Im} z \geq 0, \operatorname{Re}\Phi_m(z) - \operatorname{Re}\Phi_m(\sigma) < 2\pi \operatorname{Im} z\},$$
$$R_- := \{z \in \Xi_{m,\chi} \mid \operatorname{Im} z \leq 0, \operatorname{Re}\Phi_m(z) - \operatorname{Re}\Phi_m(\sigma) < -2\pi \operatorname{Im} z\}.$$

In fact, we will show the following lemma, whose proof will be given later.

Lemma 4.4 *The following hold for $m = 0, 1, \dots, p-2$:*

(i) *The points m/p and $(m+1)/p$ are in W_m^-.*

(ii) *There is a path C_+ in R_+ connecting m/p and $(m+1)/p$.*

(iii) *There is a path C_- in R_- connecting m/p and $(m+1)/p$.*

When $m = p-1$, there exists $\delta > 0$ such that the following hold:

(i′) *The points $1-1/p$ and $1-\delta$ are in W_{p-1}^-.*

(ii′) *There is a path C_+ in R_+ connecting $1-1/p$ and $1-\delta$.*

(iii′) *There is a path C_- in R_- connecting $1-1/p$ and $1-\delta$.*

Note that since $\Xi_{m,\chi}$ is simply connected, both C_+ and C_- are homotopic to the segment $[m/p, (m+1)/p]$ ($[1-1/p, 1-\delta]$ if $m = p-1$) in $\Xi_{m,\chi}$ keeping the boundary points fixed.

See Figure 10.

To prove the lemma above, we study the behavior of $\operatorname{Re}\Phi_m$ in $\Xi_{m,0}$ more precisely.

We divide $\Xi_{m,0}$ into six parts by the three lines L_{m+1}, $L_{m+1/2}$, and K_σ, where we put

$$K_\sigma : \operatorname{Re}(\xi z) = 0.$$

We can see that σ_m is just the intersection of L_{m+1} and K_σ.

We also introduce the four points

$$P_{34} := I_{m+1} \cap K_\sigma, \quad P_{70} := I_m \cap K_\sigma,$$

with coordinates

$$P_{34} := \frac{(m+1)\bar{\xi}\sqrt{-1}}{2p^2\pi} = \frac{m+1}{p} + \frac{(m+1)\kappa}{2p^2\pi}\sqrt{-1}, \quad P_{70} := \frac{m\bar{\xi}\sqrt{-1}}{2p^2\pi} = \frac{m}{p} + \frac{m\kappa}{2p^2\pi}\sqrt{-1}.$$

Note that P_{34} is between P_3 and P_4 (when $p = 1$, P_{34} coincides with P_4), and that P_{70} is between P_7 and P_0 (when $m = 0$, P_{70} coincides with P_0).

As in the proof of Lemma 5.2 in [24], we can prove the following lemma:

Lemma 4.5 *Write $z = x + y\sqrt{-1}$ for $z \in \Xi_{m,\chi}$ with $x, y \in \mathbb{R}$. Then we have*:

- $(\partial \operatorname{Re}\Phi_m/\partial y)(z) > 0$ *if and only if*
  - $\operatorname{Re}(\xi z) > 0$ *and* $2k\pi < \operatorname{Im}(\xi z) < (2k+1)\pi$ *for some integer* k, *or*
  - $\operatorname{Re}(\xi z) < 0$ *and* $(2l-1)\pi < \operatorname{Im}(\xi z) < 2l\pi$ *for some integer* l.
- $(\partial \operatorname{Re}\Phi_m/\partial y)(z) < 0$ *if and only if*
  - $\operatorname{Re}(\xi z) < 0$ *and* $2k\pi < \operatorname{Im}(\xi z) < (2k+1)\pi$ *for some integer* k, *or*
  - *or* $\operatorname{Re}(\xi z) > 0$ *and* $(2l-1)\pi < \operatorname{Im}(\xi z) < 2l\pi$ *for some integer* l.

See Figure 11.

Proof From (4-6), we have

$$\frac{\partial \operatorname{Re}\Phi_m(z)}{\partial y} = -\arg(3 - 2\cosh(\xi z)).$$

The right-hand side is positive (negative, respectively) if and only if $\operatorname{Im}(3-2\cosh(\xi z))$ is negative (positive, respectively). Since $\operatorname{Im}(3 - 2\cosh(\xi z)) = -2\sinh(\operatorname{Re}(\xi z))\sin(\operatorname{Im}(\xi z))$, $\partial \operatorname{Re}\Phi_m(z)/\partial y$ is positive (negative, respectively) if and only if $\operatorname{Re}(\xi z) > 0$ and $2k\pi < \operatorname{Im}(\xi z) < (2k+1)\pi$ for some integer k, or $\operatorname{Re}(\xi z) < 0$ and $(2l-1)\pi < \operatorname{Im}(\xi z) < 2l\pi$ for some integer l ($\operatorname{Re}(\xi z) < 0$ and $2k\pi < \operatorname{Im}(\xi z) < (2k+1)\pi$ for some integer k, or $\operatorname{Re}(\xi z) > 0$ and $(2l-1)\pi < \operatorname{Im}(\xi z) < 2l\pi$ for some integer l, respectively). □

Lemma 4.6 *Let z be a point on the segment $\overline{P_{70}P_{34}}$. If $z \neq \sigma_m$ is between σ_m and P_{70}, then $z \in W_m^-$. Moreover, if $z \neq \sigma_m$ is between σ_m and P_{34}, then $z \in W_m^+$.*

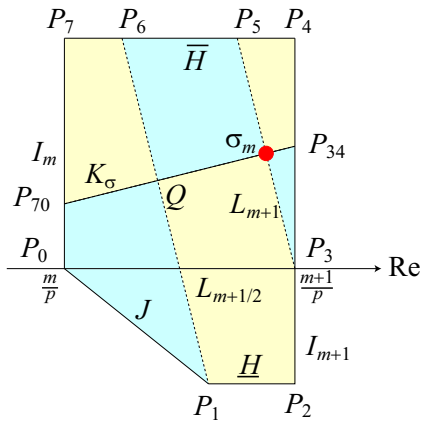


Figure 11: In the cyan (yellow, respectively) region, $\operatorname{Re}\Phi_m(z)$ is increasing (decreasing, respectively) with respect to $\operatorname{Im} z$.

Proof The segment $\overline{P_{70}P_{34}} \subset K_\sigma$ is parametrized as $(2\pi\sqrt{-1}/\xi)t$ for $m + m\kappa^2/(4p^2\pi^2) \le t \le m + 1 + (m+1)\kappa^2/(4p^2\pi^2)$. From (4-6) we have

$$\frac{d}{dt}\operatorname{Re}\Phi_m\left(\frac{2\pi\sqrt{-1}}{\xi}t\right) = \operatorname{Re}\left(\frac{2\pi\sqrt{-1}}{\xi}\log\left(3-2\cosh(2\pi\sqrt{-1}t)\right)\right) = \frac{4p\pi^2}{|\xi|^2}\log(3-2\cos 2\pi t) \ge 0,$$

and the equality holds only when $t = m+1$, that is, $z = \sigma_m$. So we conclude that if $z \in \overline{P_{70}\sigma_m} \setminus \{\sigma_m\}$, then $\operatorname{Re}\Phi_m(z) < \operatorname{Re}\Phi_m(\sigma_m)$, and that if $z \in \overline{\sigma_m P_{34}} \setminus \{\sigma_m\}$, then $\operatorname{Re}\Phi_m(z) > \operatorname{Re}\Phi_m(\sigma_m)$, as required. □

We can prove a similar result for $\overline{P_3P_5}$.

Lemma 4.7 *Let z be a point on the segment $\overline{P_3P_5}$. If $z \ne \sigma_m$ is between σ_m and P_3, then $z \in W_m^-$. Moreover, if $z \ne \sigma_m$ is between σ_m and P_5, then $z \in W_m^+$.*

Proof A point on $\overline{P_3P_5}$ is parametrized as $(m+1)/p - ((p+m)\kappa/(4p^3\pi^2))\bar{\xi}t$ for $0 \le t \le 1$. From (4-6) we have

$$\begin{aligned}\frac{d}{dt}\operatorname{Re}\Phi_m\left(\frac{m+1}{p} - \frac{(p+m)\kappa}{4p^3\pi^2}\bar{\xi}t\right) &= -\operatorname{Re}\left[\frac{(p+m)\kappa}{4p^3\pi^2}\bar{\xi}\log\left(3-2\cosh\left(\frac{(m+1)\xi}{p} - \frac{(p+m)\kappa}{4p^3\pi^2}|\xi|^2 t\right)\right)\right] \\ &= \frac{-(p+m)\kappa^2}{4p^3\pi^2}\log\left(3-2\cosh\left(\frac{(m+1)\kappa}{p} - \frac{(p+m)\kappa}{4p^3\pi^2}|\xi|^2 t\right)\right) \ge 0,\end{aligned}$$

where the equality holds when $t = 4(m+1)p^2\pi^2/((p+m)|\xi|^2)$, which shows that $\operatorname{Re}\Phi_m(z) < \operatorname{Re}\Phi_m(\sigma_m)$ if $z \in \overline{P_3\sigma_m} \setminus \{\sigma_m\}$ and that $\operatorname{Re}\Phi_m(z) > \operatorname{Re}\Phi_m(\sigma_m)$ if $z \in \overline{\sigma_m P_5} \setminus \{\sigma_m\}$, completing the proof. □

So far we have found two directions $\overrightarrow{\sigma_m P_{70}}$ and $\overrightarrow{\sigma_m P_3}$ that go down valleys, and two directions $\overrightarrow{\sigma_m P_{34}}$ and $\overrightarrow{\sigma_m P_5}$ that go up hills. Since the function $\Phi_m(z)$ is of the form $\Phi_m(\sigma_m) - \frac{1}{3}\xi^2 z^3 + \cdots$ from (4-7), that is, σ_m is a saddle point of order two, there should be another pair of valley and hill.

Lemma 4.8 *Let G be the line segment in $\Xi_{m,0}$ that bisects the angle $\angle P_{34}\sigma_m P_5$. If $z \in G \setminus \{\sigma_m\}$ is on the same side of P_{34} and P_5, then $z \in W_m^-$. If $z \in G \setminus \{\sigma_m\}$ is on the opposite side of P_{34} and P_5, and close enough to σ_m, then $z \in W_m^+$.*

Proof Since the vector $\overrightarrow{\sigma_m P_{34}}$ has the same direction as $\sqrt{-1}/\xi$ and the vector $\overrightarrow{\sigma_m P_5}$ has the same direction as $-1/\xi$, the bisector is parametrized as $\sigma_m + (\sqrt{-1}-1)t/\xi$ with $t \in \mathbb{R}$. Note that if $t > 0$, it goes to the top right, and that if $t < 0$, it goes to the bottom left.

From (4-6), we have

$$\frac{d}{dt}\operatorname{Re}\Phi_m\Big(\sigma_m + \frac{\sqrt{-1}-1}{\xi}t\Big) = \operatorname{Re}\Big(\frac{\sqrt{-1}-1}{\xi}\log(3-2\cosh((\sqrt{-1}-1)t))\Big),$$

and so $(d/dt)\operatorname{Re}\Phi_m\big(\sigma_m + ((\sqrt{-1}-1)/\xi)t\big) = 0$ when $t = 0$. Thus, it is sufficient to show that the second derivative of $\operatorname{Re}\Phi_m(\sigma_m + (\sqrt{-1}-1)t/\xi)$ is positive when $t < 0$ and $|t|$ is small, and that it is negative when $t > 0$ and $\sigma_m + (\sqrt{-1}-1)t/\xi \in \Xi_{m,0}$.

From (4-4), we have

$$\begin{aligned}\frac{d^2}{dt^2}\operatorname{Re}\Phi_m\left(\sigma_m + \frac{\sqrt{-1}-1}{\xi}t\right) &= \operatorname{Re}\left(\frac{(\sqrt{-1}-1)^2}{\xi^2}\frac{\xi(e^{-(\sqrt{-1}-1)t} - e^{(\sqrt{-1}-1)t})}{3-e^{(\sqrt{-1}-1)t} - e^{-(\sqrt{-1}-1)t}}\right)\\ &= \frac{1}{|\xi|^2}\operatorname{Re}((-4p\pi - 2\kappa\sqrt{-1})\lambda(t)) = \frac{1}{|\xi|^2}(-4p\pi\operatorname{Re}\lambda(t) + 2\kappa\operatorname{Im}\lambda(t)),\end{aligned}$$

where we put

$$\lambda(t) := \frac{e^{-(\sqrt{-1}-1)t} - e^{(\sqrt{-1}-1)t}}{3 - e^{(\sqrt{-1}-1)t} - e^{-(\sqrt{-1}-1)t}}.$$

We have

$$\begin{aligned}\lambda(t) &= \frac{2\sinh t\cos t - 2\sqrt{-1}\cosh t\sin t}{3 - 2\cosh t\cos t + 2\sqrt{-1}\sinh t\sin t}\\ &= \frac{2(\sinh t\cos t - \sqrt{-1}\cosh t\sin t)}{(3-2\cosh t\cos t)^2 + 4\sinh^2 t\sin^2 t}(3 - 2\cosh t\cos t - 2\sqrt{-1}\sinh t\sin t)\\ &= \frac{2\sinh t(3\cos t - 2\cosh t) + 2\sqrt{-1}\sin t(2\cos t - 3\cosh t)}{(3-2\cosh t\cos t)^2 + 4\sinh^2 t\sin^2 t}.\end{aligned}$$

Therefore if t is negative and $|t|$ is small enough, then $\operatorname{Re}\lambda(t) < 0$ and $\operatorname{Im}\lambda(t) > 0$, and so in this case $(d^2/dt^2)\operatorname{Re}\Phi_m\big(\sigma_m + ((\sqrt{-1}-1)/\xi)t\big) > 0$.

Next, we consider the case where $t > 0$.

Since $\operatorname{Re}(\sigma_m + (\sqrt{-1}-1)t/\xi) = (1/|\xi|^2)(4(m+1)p\pi^2 + (2p\pi - \kappa)t)$, a point in G that is between σ_m and I_{m+1} is parametrized as $\sigma_m + (\sqrt{-1}-1)t/\xi$ with $0 < t < (m+1)\kappa^2/(p(2p\pi - \kappa))$. Since $(m+1)\kappa^2/(p(2p\pi-\kappa)) \le \kappa^2/(2p\pi - \kappa) \le \kappa^2/(2\pi - \kappa)$, it is sufficient to prove

$$(d^2/dt^2)\operatorname{Re}\Phi_m(\sigma_m + ((\sqrt{-1}-1)/\xi)t) < 0$$

for $0 < t < \kappa^2/(2\pi - \kappa)$.

Since $3\cos(\kappa^2/(2\pi-\kappa))-2\cosh(\kappa^2/(2\pi-\kappa))=0.924\ldots$, and the function $3\cos t-2\cosh t$ is monotonically decreasing when $t>0$, we see that $\operatorname{Re}\lambda(t)>0$ for $0<t<\kappa^2/(2\pi-\kappa)$. We can easily see that $\operatorname{Im}\lambda(t)<0$ for $t>0$, and so we conclude that $(d^2/dt^2)\operatorname{Re}\Phi(\sigma_m+((\sqrt{-1}-1)/\xi)t)<0$. □

Remark 4.9 The imaginary part of the intersection of G with I_{m+1} is $(m+1)\kappa/(p(2p\pi-\kappa))$, which is smaller than the imaginary part of $\overline{H}$ when $p>1$. This is because

$$\begin{aligned}\frac{(p+m)\kappa}{2p^2\pi}-\frac{(m+1)\kappa}{p(2p\pi-\kappa)}&=\frac{(2p\pi(p-1)-(p+m)\kappa)\kappa}{2p^2\pi(2p\pi-\kappa)}\\&>\frac{(2p\pi(p-1)-(2p-1))\kappa}{2p^2\pi(2p\pi-\kappa)}=\frac{((2p\pi-1)(p-1)-p)\kappa}{2p^2\pi(2p\pi-\kappa)},\end{aligned}$$

which is positive when $p>1$, where we use the inequalities $\kappa<1$ and $m\le p-1$. So G intersects with the segment $\overline{P_4P_{34}}$.

If $p=1$, G intersects with the segment $\overline{P_4P_5}$.

Note that G does not intersect with $L_{m+1/2}$ in $\Xi_{m,0}$. This is because the intersection between G and $L_{m+1/2}$ is $(\pi+(2m+1)\pi\sqrt{-1})/\xi$, whose imaginary part is less than $-\operatorname{Im}\sigma_m$.

There are more line segments that are included in W_m^-.

Lemma 4.10 *The line segments $\overline{P_6P_1}$, $\overline{P_0P_{70}}$, and $\overline{P_0P_1}$ are in W_m^-.*

Proof A point on the segment $\overline{P_6P_1}$ is parametrized as

$$\frac{m+\frac12}{p}+\frac{\bar\xi}{2p\pi}t\quad\text{where}\quad-\frac{(p+m)\kappa}{2p^2\pi}\le t\le\operatorname{Im}\sigma_m.$$

We have

$$\begin{aligned}\frac{d}{dt}\operatorname{Re}\Phi_m\left(\frac{m+\frac12}{p}+\frac{\bar\xi}{2p\pi}t\right)&=\operatorname{Re}\left(\frac{\bar\xi}{2p\pi}\log\left(3-2\cosh\left(\frac{m+\frac12}{p}\xi+\frac{|\xi|^2t}{2p\pi}\right)\right)\right)\\&=\frac{\kappa}{2p\pi}\log\left(3+2\cosh\left(\frac{(m+\frac12)\kappa}{p}+\frac{|\xi|^2t}{2p\pi}\right)\right)>0.\end{aligned}$$

From Lemma 4.11 below we know that $\operatorname{Re}\Phi_m(P_1)<\operatorname{Re}\Phi_m(\sigma_m)$. It follows that $\overline{P_6P_1}\subset W_m^-$.

From Lemma 4.5, $\operatorname{Re}\Phi_m(z)$ is increasing with respect to $\operatorname{Im}z$ in the quadrilateral $P_{70}P_0P_1Q$, where Q is the crossing between K_σ and $L_{m+1/2}$. Since the upper segments $\overline{P_{70}Q}$ and $\overline{QP_1}$ are in W_m^-, so are the lower segments $\overline{P_{70}P_0}$ and $\overline{P_0P_1}$. □

Lemma 4.11 *The point P_1 is in W_m^-.*

Proof The following proof is similar to that of [24, Lemma 5.3].

Since P_1 is on $L_{m+1/2}$, we have $\operatorname{Im}(\xi(P_1-2m\pi\sqrt{-1}/\xi))=\pi$ and so $P_1-2m\pi\sqrt{-1}/\xi$ is on $L_{1/2}$. So from (3-6) we have

$$\xi\Phi_m(P_1)-\xi\Phi_m(\sigma_m)=\operatorname{Li}_2(-e^{-\kappa-(4m+3)\kappa/(2p)})-\operatorname{Li}_2(-e^{-\kappa+(4m+3)\kappa/(2p)})+\frac{(4m+3)\kappa^2}{2p}-\kappa\pi\sqrt{-1}.$$

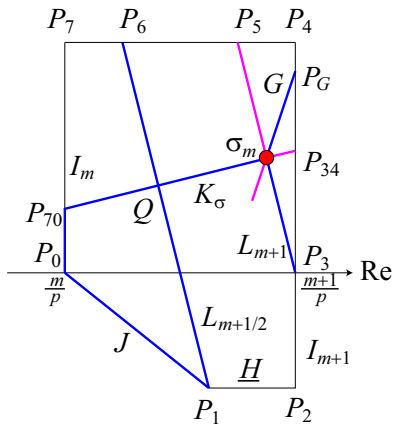


Figure 12: The blue (magenta, respectively) lines are included in W_m^- (W_m^+, respectively).

Its real part is

$$\mathrm{Li}_2(-e^{-\kappa-(4m+3)\kappa/(2p)})-\mathrm{Li}_2(-e^{-\kappa+(4m+3)\kappa/(2p)})+\frac{(4m+3)\kappa^2}{2p}$$

and its imaginary part is $-\kappa\pi$.

Therefore we have

$$\begin{aligned}(4\text{-}14)\quad &\frac{|\xi|^2}{\kappa}(\mathrm{Re}\,\Phi_m(P_1)-\mathrm{Re}\,\Phi_m(\sigma_m))\\ &=\mathrm{Re}(\xi\Phi_m(P_1)-\xi\Phi_m(\sigma_m))+\frac{2p\pi}{\kappa}\,\mathrm{Im}(\xi\Phi_m(P_1)-\xi\Phi_m(\sigma_m))\\ &=\mathrm{Li}_2(-e^{-\kappa-(4m+3)\kappa/(2p)})-\mathrm{Li}_2(-e^{-\kappa+(4m+3)\kappa/(2p)})+\frac{(4m+3)\kappa^2}{2p}-2p\pi^2,\end{aligned}$$

which is increasing with respect to m, fixing p. When $m=p-1$, (4-14) equals

$$(4\text{-}15)\qquad \mathrm{Li}_2(-e^{-\kappa-(4p-1)\kappa/(2p)})-\mathrm{Li}_2(-e^{-\kappa+(4p-1)\kappa/(2p)})+\frac{(4p-1)\kappa^2}{2p}-2p\pi^2.$$

Its derivative with respect to p is

$$\frac{\kappa}{2p^2}\log\Big(3+2\cosh\Big(\kappa(2-\frac{1}{2p})\Big)\Big)-2\pi^2,$$

which is less than $\log(3+2\cosh(2\kappa))-2\pi^2=\log(10)-2\pi^2<0$. Since (4-15) equals $-17.2195\ldots$ when $p=1$, we conclude that (4-14) is negative, proving the lemma. □

Remark 4.12 One can also show that the polygonal line $\overline{P_{70}P_7P_6}$ is in W_m^-.

The results in Lemmas 4.6–4.8 and 4.10 are summarized in Figure 12.

Proof of Lemma 4.4 First, suppose that $m<p-1$.

(i) Since $m/p=P_0$ and $(m+1)/p=P_3$, it follows from Figure 12 that these points are in W_m^-.

(ii) Consider the polygonal line $C_+ := \overline{P_0P_{70}\sigma_m P_3}$. From Figure 12, it is in W_m^- and in the upper half plane $\{z \in \mathbb{C} \mid \operatorname{Im} z \geq 0\}$. So it is contained in R_+.

(iii) From Figure 12, the line segment J is in W_m^- and in the lower half plane $\{z \in \mathbb{C} \mid \operatorname{Im} z \leq 0\}$. This implies that $J \subset R_-$.

We will show that the segments $\overline{P_1P_2}$ and $\overline{P_2P_3}$ are also in R_-.

We first show that $\overline{P_1P_2} \subset R_-$, that is, $\operatorname{Re}\Phi_m(z) - \operatorname{Re}\Phi_m(\sigma_m) < -2\pi \operatorname{Im} z$, if $z \in \overline{P_1P_2}$. From the proof of Lemma 4.5, $-\pi < \partial \operatorname{Re}\Phi_m(z)/\partial y < 0$ if $z = x + y\sqrt{-1}$ is in the pentagonal region $QP_1P_2P_3\sigma_m$. We also know that if $z \in \overline{Q\sigma_m P_3}$, then $\operatorname{Re}\Phi_m(z) - \operatorname{Re}\Phi_m(\sigma_m) \leq 0$. Since the difference between the imaginary part of the point on $\overline{Q\sigma_m P_3}$ and that of the point on $\overline{P_1P_2}$ is less than or equal to $2\operatorname{Im}\sigma_m$, it follows that for $z \in \overline{P_1P_2}$, we have $\operatorname{Re}\Phi_m(z) - \operatorname{Re}\Phi_m(\sigma_m) < \pi(2\operatorname{Im}\sigma_m) = -2\pi \operatorname{Im} z$.

Next we show $\overline{P_2P_3} \subset R_-$. Consider $r(y) := \operatorname{Re}\Phi_m((m+1)/p + y\sqrt{-1}) - \operatorname{Re}\Phi_m(\sigma_m) + 2\pi y$. Since $(d/dy)r(y) = (\partial/\partial y)\operatorname{Re}\Phi_m((m+1)/p + y\sqrt{-1}) + 2\pi > 0$ and $r(0) < 0$ from the argument above, we conclude that $r(y) < 0$ if $y \geq -\operatorname{Im}\sigma_m$. So if $z \in \overline{P_2P_3}$, then $z \in R_-$.

Therefore, we can put $C_- := \overline{P_0P_1P_2P_3} \subset R_-$.

Next, we consider the case where $m = p-1$. Here we can push P_3 slightly to the left to avoid $\diamond_v$. Accordingly, we move the segments $\overline{\sigma_m P_3}$ and $\overline{P_2P_3}$ slightly. □

Therefore we can apply Proposition 4.1 to the series of functions $\psi_N(z) = \varphi_{m,N}(z) - \varphi_{m,N}(\sigma_m)$. We conclude that

$$\frac{1}{N}e^{-N\varphi_{m,N}(\sigma_m)} \sum_{m/p \leq k/N \leq (m+1)/p} e^{N\varphi_{m,N}(k/N)} = e^{-N\varphi_{m,N}(\sigma_m)} \int_{m/p}^{(m+1)/p} e^{N\varphi_{m,N}(z)}\, dz + O(e^{-\varepsilon_m N}) \tag{4-16}$$

for $\varepsilon_m > 0$ if $m < p-1$, and

$$\frac{1}{N}e^{-N\varphi_{p-1,N}(\sigma_m)} \sum_{(p-1)/p \leq k/N \leq 1-\delta} e^{N\varphi_{p-1,N}(k/N)} = e^{-N\varphi_{p-1,N}(\sigma_{p-1})} \int_{(p-1)/p}^{1-\delta} e^{N\varphi_{p-1,N}(z)}\, dz + O(e^{-\varepsilon_{p-1}N}) \tag{4-17}$$

for $\varepsilon_{p-1} > 0$.

5 The saddle point method of order two

We would like to know the asymptotic behavior of the integrals appearing in the right-hand sides of (4-16) and (4-17) by using the saddle point method of order two.

To describe it, let us consider a holomorphic function $\eta(z)$ in a domain $D \ni O$ with $\eta(0) = \eta'(0) = \eta''(0) = 0$ and $\eta^{(3)}(0) \neq 0$, where O is the origin of the complex plane. Write $\eta^{(3)}(0) = 6re^{\theta\sqrt{-1}}$ with $r > 0$ and $-\pi < \theta \leq \pi$. Then $\eta(z)$ is of the form $\eta(z) = re^{\theta\sqrt{-1}}z^3 g(z)$, where $g(z)$ is holomorphic with $g(0) = 1$. The origin is called a saddle point of $\operatorname{Re}\eta(z)$ of order two. We put $V := \{z \in D \mid \operatorname{Re}\eta(z) < 0\}$.

There exists a small disk $\widehat{D} \subset D$ centered at O, where we can define a cubic root $g^{1/3}(z)$ of $g(z)$ such that $g^{1/3}(0) = 1$. Put $G(z) := zg^{1/3}(z)$ in $\widehat{D} \subset D$. We can choose $\widehat{D}$ so that G gives a bijection from $\widehat{D}$ to $E := G(\widehat{D})$ from the inverse function theorem because $G'(0) = 1$. Since $re^{\theta\sqrt{-1}}G(z)^3 = \eta(z)$, the function G also gives a bijection from the region $V \cap \widehat{D}$ to the region $U := \{w \in E \mid \operatorname{Re}(re^{\theta\sqrt{-1}}w^3) < 0\}$.

The region U splits into the three connected components (valleys) U_1, U_2, and U_3. Therefore the region $V \cap \widehat{D}$ also splits into three valleys $V_k := G^{-1}(U_k)$, for $k = 1, 2, 3$, of $\operatorname{Re}\eta(z)$.

Remark 5.1 Since $G'(0) = 1$, and U_k contains the ray $\{w \in E \mid w = se^{((2k-1)\pi-\theta)\sqrt{-1}/3}, s > 0\}$ as a bisector, V_k also contains a segment $\{z \in \widehat{D} \mid z = te^{((2k-1)\pi-\theta)\sqrt{-1}/3}$ for $t > 0$ small$\}$.

The following is the statement of the saddle point method of order two:

Proposition 5.2 *Let $\eta(z)$ be a holomorphic function in a domain $D \ni O$ with $\eta(0) = \eta'(0) = \eta''(0) = 0$ and $\eta^{(3)}(0) \neq 0$. Write $\eta^{(3)}(0) = 6re^{\theta\sqrt{-1}}$ with $r > 0$ and $-\pi < \theta \leq \pi$. Put $V := \{z \in D \mid \operatorname{Re}\eta(z) < 0\}$ and define V_k for $k = 1, 2, 3$ as above. Let $C \subset D$ be a path from a to b with $a, b \in V$.*

We assume that there exist paths $P_k \subset V \cup \{O\}$ from a to O and $P_{k+1} \subset V \cup \{O\}$ from O to b such that

(i) $(P_k \cap \widehat{D}) \setminus \{O\} \subset V_k$,

(ii) $(P_{k+1} \cap \widehat{D}) \setminus \{O\} \subset V_{k+1}$, *and*

(iii) *the path $P_k \cup P_{k+1}$ is homotopic to C in D keeping a and b fixed,*

where $\widehat{D} \in O$ is a disk as above.

Let $\{h_N(z)\}$ be a series of holomorphic functions in D that uniformly converges to a holomorphic function $h(z)$ with $h(0) \neq 0$. We also assume that $|h_N(z)|$ is bounded irrelevant to z or N. Then

$$\int_C h_N(z)e^{N\eta(z)}\,dz = \frac{h(0)\Gamma(1/3)\sqrt{-1}}{\sqrt{3}r^{1/3}N^{1/3}}\omega^k e^{-\theta\sqrt{-1}/3}(1 + O(N^{-1/3})) \tag{5-1}$$

as $N \to \infty$, where $\omega := e^{2\pi\sqrt{-1}/3}$.

The proposition may be well known to experts, but we give a proof in Appendix B because the author is not an expert and could not find appropriate references.

We will apply Proposition 5.2 to

- $\eta(z) := \Phi_m(z + \sigma_m) - \Phi_m(\sigma_m)$,
- $D := \{z \in \mathbb{C} \mid z + \sigma_m \in \Xi_{m,\chi}\}$,

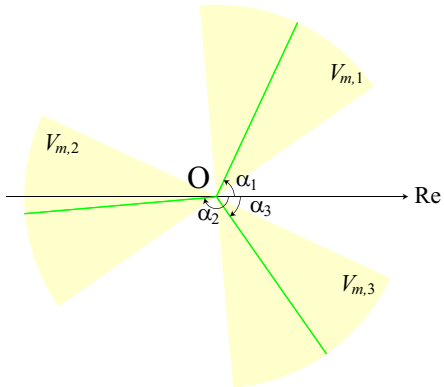


Figure 13: The yellow regions indicates the valleys $V_{m,1}$, $V_{m,2}$, and $V_{m,3}$.

- $h_N(z) := \exp[N(\varphi_{m,N}(z+\sigma_m) - \Phi_m(z+\sigma_m))]$, and
- $C := [m/p - \sigma_m, (m+1)/p - \sigma_m]$ for $m < p-1$, and $C := [(p-1)/p, 1-\delta]$ for $m = p-1$, where δ is a positive small number (see Lemma 4.4).

Note that $\eta(0) = \eta'(0) = \eta''(0) = 0$, $\eta^{(3)}(0) = -2\xi^2 \neq 0$, $h(z) := \lim_{N\to\infty} h_N(z) = 1$, and that V is equal to the region $\{z \in \mathbb{C} \mid z + \sigma_m \in W_m^-\}$.

Since $\eta(z) = -\frac{1}{3}\xi^2 z^3 + \cdots$, we can define a holomorphic function $g(z) := -3\eta(z)/(\xi^2 z^3)$ so that $g(0) = 1$. Put $G(z) := zg^{1/3}(z)$ as above. Let $\widehat{D} \subset D$ be a small disk centered at 0 such that the function $G(z)$ is a bijection. Then the region V splits into three valleys $V_{m,1}$, $V_{m,2}$, and $V_{m,3}$. From Remark 5.1, the argument of the bisector of $V_{m,k}$ is given by $(2k-1)\frac{1}{3}\pi - \frac{1}{3}\theta \pmod{2\pi}$ for $k = 1, 2, 3$, where $\theta := \arg(-2\xi^2) = -\pi + 2\arctan(2p\pi/\kappa)$. So the valley $V_{m,k}$ is approximated by the small sector

$$\{z \in \mathbb{C} \mid z = te^{\tau\sqrt{-1}} \text{ and } |\tau - \alpha_k| < \tfrac{1}{6}\pi \text{ for } t > 0 \text{ small}\},$$

where we put

$$\text{(5-2)}\quad \alpha_1 := -\tfrac{2}{3}\arctan(2p\pi/\kappa) + \tfrac{2}{3}\pi, \quad \alpha_2 := -\tfrac{2}{3}\arctan(2p\pi/\kappa) - \tfrac{2}{3}\pi, \quad \alpha_3 := -\tfrac{2}{3}\arctan(2p\pi/\kappa).$$

Note that since $\frac{1}{4}\pi < \arctan(2p\pi/\kappa) < \frac{1}{2}\pi$, we have $\frac{1}{3}\pi < \alpha_1 < \frac{1}{2}\pi$, $-\pi < \alpha_2 < -\frac{5}{6}\pi$, and $-\frac{1}{3}\pi < \alpha_3 < -\frac{1}{6}\pi$; see Figure 13.

Remark 5.3 Denote by P_G the intersection between G and the boundary of $\Xi_{m,0}$, as in Figure 12. Note that $P_G \subset I_{m+1}$ if $m < p-1$ and $P_G \subset \overline{H}$ if $p = 1$ from Remark 4.9. The arguments of $\overrightarrow{\sigma_m P_G}$, $\overrightarrow{\sigma_m P_{70}}$, and $\overrightarrow{\sigma_m P_3}$ are

$$\text{(5-3)}\qquad \beta_1 := -\arctan(2p\pi/\kappa) + \tfrac{3}{4}\pi, \quad \beta_2 := -\arctan(2p\pi/\kappa) - \tfrac{1}{2}\pi, \quad \beta_3 := -\arctan(2p\pi/\kappa),$$

respectively, because the vector $\overrightarrow{\sigma_m P_{70}}$ has the same direction as $-\sqrt{-1}/\xi$, the vector $\overrightarrow{\sigma_m P_3}$ has the same direction as $1/\xi$, and G is their bisection.

Since $\frac{1}{4}\pi < \arg(2p\pi/\kappa) < \frac{1}{2}\pi$, we can see

$$\alpha_1 - \beta_1 = -\tfrac{1}{12}\pi + \tfrac{1}{3}\arctan(2p\pi/\kappa), \quad \beta_2 - \alpha_2 = \tfrac{1}{6}\pi - \tfrac{1}{3}\arctan(2p\pi/\kappa), \quad \alpha_3 - \beta_3 = \tfrac{1}{3}\arctan(2p\pi/\kappa),$$

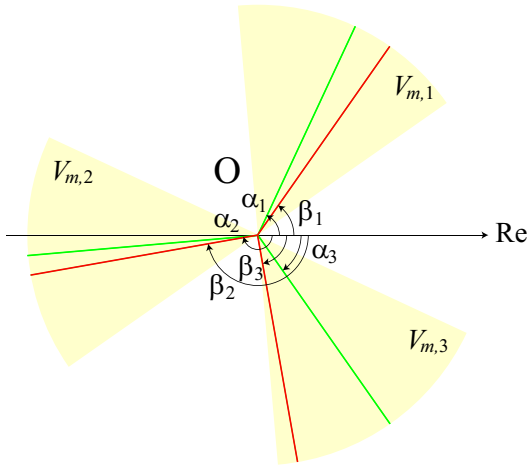


Figure 14: The yellow regions indicate the valleys.

and

$$\alpha_2 < \beta_2 < \beta_3 < \alpha_3 < \beta_1 < \alpha_1,$$

where α_k for $k = 1, 2, 3$ are given in (5-2). We also conclude that $|\alpha_k - \beta_k| < \frac{1}{6}\pi$, that is, $\overline{\sigma_m P_G}$ is in the valley $V_{m,1}$, $\overline{\sigma_m P_{50}}$ is in the valley $V_{m,2}$, and $\overline{\sigma_m P_3}$ is in the valley $V_{m,3}$; see Figure 14.

We need to show that the assumption of Proposition 5.2 holds, that is, we will show the following lemma:

Lemma 5.4 *First suppose that $m = 0, 1, 2, \dots, p-2$. If a disk $\widetilde{D} \subset \Xi_{m,\chi}$ centered at σ_m is small enough, then the following hold*:

(i) *There exists a path $\rho_2 \subset W_m^- \cup \{\sigma_m\}$ connecting m/p and σ_m such that $(\rho_2 \cap \widetilde{D}) \setminus \{\sigma_m\} \subset V_{m,2}$.*

(ii) *There exists a path $\rho_3 \subset W_m^- \cup \{\sigma_m\}$ connecting σ_m and $(m+1)/p$ such that $(\rho_3 \cap \widetilde{D}) \setminus \{\sigma_m\} \subset V_{m,3}$.*

Next, suppose that $m = p-1$. If a disk $\widetilde{D} \subset \Xi_{p-1,\chi}$ centered at σ_{p-1} is small enough, then the following hold:

(i′) *There exists a path $\rho_2 \subset W_{p-1}^- \cup \{\sigma_{p-1}\}$ connecting $1-1/p$ and σ_{p-1} such that $(\rho_2 \cap \widetilde{D}) \setminus \{\sigma_{p-1}\} \subset V_{p-1,2}$.*

(ii′) *There exists a path $\rho_3 \subset W_m^{\nu} \cup \{\sigma_{p-1}\}$ connecting σ_{p-1} and $1-\delta$ such that $(\rho_3 \cap \widehat{D}) \setminus \{\sigma_{p-1}\} \subset V_{p-1,3}$,*

Again, since $\Xi_{m,\chi}$ is simply connected, the path $\rho_2 \cup \rho_3$ is homotopic to the interval $[m/p, (m+1)/p]$ ($[1-1/p, 1-\delta]$, respectively) if $m < p-1$ ($m = p-1$, respectively).

Proof The proof is essentially the same for both cases $m < p-1$ and $m = p-1$.

(i) The path $\rho_2 := \overline{P_0 P_{70} \sigma_m}$ is a required one for $m = 0, 1, 2, \dots, p-1$.

(ii) When $m < p-1$, consider the path $\rho_3 := \overline{\sigma_m P_3}$, and when $m = p-1$ push it a little more to the left near the point 1. □

If $m < p-1$, we apply Proposition 5.2 to

$$\eta(z) = \Phi_m(z+\sigma_m) - \Phi_m(\sigma_m), \qquad h_N(z) = \exp[N(\varphi_{m,N}(z+\sigma_m) - \Phi_m(z+\sigma_m))],$$
$$C := [m/p - \sigma_m, (m+1)/p - \sigma_m], \qquad k = 2.$$

Noting that $h_N(z)$ converges to 1 and $\eta^{(3)}(0) = -2\xi^2 = 2|\xi|^2 e^{\theta\sqrt{-1}}$ with $\theta = -\pi + 2\arctan(2p\pi/\kappa)$ from the argument above, we have

$$\begin{aligned}\int_{m/p}^{(m+1)/p} & e^{N(\varphi_{m,N}(z)-\Phi_m(\sigma_m))}\,dz \\ &= \int_C e^{N(\varphi_{m,N}(z+\sigma_m)-\Phi_m(\sigma_m))}\,dz = \int_C h_N(z)e^{N\eta(z)}\,dz \\ &= \frac{\Gamma\left(\frac{1}{3}\right)\sqrt{-1}}{\sqrt{3}\left(\frac{1}{3}|\xi|^2\right)^{1/3}N^{1/3}}\omega^2 e^{\pi\sqrt{-1}/3 - 2\arctan(2p\pi/\kappa)\sqrt{-1}/3}(1+O(N^{-1/3})) \\ &= \frac{\Gamma\left(\frac{1}{3}\right)\sqrt{-1}}{3^{1/6}|\xi|^{2/3}N^{1/3}}e^{-(\pi+2\arctan(2p\pi/\kappa))\sqrt{-1}/3}(1+O(N^{-1/3}))\end{aligned}$$

as $N \to \infty$. Similarly, if $m = p-1$, putting $C := [1-1/p-\sigma_m, 1-\sigma_m-\delta]$, we have

$$\int_{1-1/p}^{1-\delta} e^{N(\varphi_{p-1,N}(z)-\Phi_{p-1}(\sigma_{p-1}))}\,dz = \frac{\Gamma\left(\frac{1}{3}\right)\sqrt{-1}}{3^{1/6}|\xi|^{2/3}N^{1/3}}e^{-(\pi+2\arctan(2p\pi/\kappa))\sqrt{-1}/3}(1+O(N^{-1/3}))$$

as $N \to \infty$. Since $\Phi_m(\sigma_m) = 4p\pi^2/\xi$ from (3-12), we conclude

$$\int_{m/p}^{(m+1)/p} e^{N\varphi_{m,N}(z)}\,dz = \frac{\Gamma\left(\frac{1}{3}\right)\sqrt{-1}}{3^{1/6}|\xi|^{2/3}N^{1/3}}e^{-(\pi+2\arctan(2p\pi/\kappa))\sqrt{-1}/3}e^{4p\pi^2N/\xi}(1+O(N^{-1/3}))$$

if $m < p-1$, and

$$\int_{1-1/p}^{1-\delta} e^{N\varphi_{p-1,N}(z)}\,dz = \frac{\Gamma\left(\frac{1}{3}\right)\sqrt{-1}}{3^{1/6}|\xi|^{2/3}N^{1/3}}e^{-(\pi+2\arctan(2p\pi/\kappa))\sqrt{-1}/3}e^{4p\pi^2N/\xi}(1+O(N^{-1/3})).$$

Since $\varphi_{m,N}(\sigma_m) = f_N(\sigma_0)$ converges to $F(\sigma_0) = 4p\pi^2/\xi$ as $N \to \infty$ from (3-12), together with (4-16) and (4-17), we finally have

$$\sum_{m/p \le k/N \le (m+1)/p} e^{N\varphi_{m,N}(k/N)} = \frac{\Gamma\left(\frac{1}{3}\right)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\left(\frac{N}{\xi}\right)^{2/3}e^{4p\pi^2N/\xi}(1+O(N^{-1/3})) \tag{5-4}$$

if $m < p-1$, and

$$\sum_{1-1/p \le k/N \le 1-\delta} e^{N\varphi_{p-1,N}(k/N)} = \frac{\Gamma\left(\frac{1}{3}\right)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\left(\frac{N}{\xi}\right)^{2/3}e^{4p\pi^2N/\xi}(1+O(N^{-1/3})) \tag{5-5}$$

because $\mathrm{Re}(4p\pi^2/\xi) > 0$, where we define $\xi^{2/3}$ to be $|\xi|^{2/3}e^{2\arctan(2p\pi/\kappa)\sqrt{-1}/3}$.

It remains to obtain the asymptotic behavior of $\sum_{1-1/p\le k/N<1} e^{N\varphi_{p-1,N}(k/N)}$ instead of the sum for $1-1/p\le k/N\le 1-\delta$. To do that, we need to estimate the sum $\sum_{1-\delta<k/N<1} e^{N\varphi_{p-1,N}(k/N)}$. We use the following lemma, which corresponds to [24, Lemma 6.1].

Lemma 5.5 *For any ε, there exists $\delta'>0$ such that*

$$\operatorname{Re}\varphi_{p-1,N}\Big(\frac{2k+1}{2N}\Big)<\operatorname{Re}\Phi_{p-1}(\sigma_{p-1})-\varepsilon$$

for sufficiently large N, if $1-\delta'<k/N<1$.

Since a proof is similar to that of [24, Lemma 6.1], we omit it.

From Lemma 5.5, we conclude that

$$\sum_{1-\delta<k/N<1}\exp\Big(N\varphi_{p-1,N}\Big(\frac{2k+1}{2N}\Big)\Big)$$

is of order $O(e^{N(\operatorname{Re}\Phi_{p-1}(\sigma_{p-1})-\varepsilon)})$ if $\delta'<\delta$. Since $\Phi_{p-1}(\sigma_{p-1})=4p\pi^2\sqrt{-1}/\xi$ from (4-7), we have

$$\sum_{1-1/p\le k/N<1} e^{N\Phi_{p-1,N}(k/N)}=\frac{\Gamma\big(\frac13\big)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\Big(\frac{N}{\xi}\Big)^{2/3}e^{4p\pi^2N/\xi}(1+O(N^{-1/3}))$$

from (5-5). Together with (4-1) and (5-4), we have

$$J_N(\mathscr{E};e^{\xi/N})=(1-e^{-4pN\pi^2/\xi})\frac{\Gamma\big(\frac13\big)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\Big(\frac{N}{\xi}\Big)^{2/3}e^{4p\pi^2N/\xi}\Big(\sum_{m=0}^{p-1}\beta_{p,m}\Big)(1+O(N^{-1/3})). \tag{5-6}$$

Now from (3-3) and (3-1), the sum in the parentheses is just $J_p(\mathscr{E};e^{4\pi^2N/\xi})$. Therefore we finally have

$$J_N(\mathscr{E};e^{\xi/N})=J_p(\mathscr{E};e^{4\pi^2N/\xi})\frac{\Gamma\big(\frac13\big)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\Big(\frac{N}{\xi}\Big)^{2/3}\exp\Big(\frac{2\kappa\pi\sqrt{-1}}{\xi}N\Big)(1+O(N^{-1/3})),$$

where we replace $e^{4p\pi^2N/\xi}$ with $e^{(4p\pi^2N)/\xi+2N\pi\sqrt{-1}}=e^{2N\kappa\pi\sqrt{-1}/\xi}$ on purpose; see Section 6. Note that we choose the argument of $\xi^{2/3}$ as $\frac23\arctan(2p\pi/\kappa)$, which is between $\frac16\pi$ and $\frac13\pi$.

Proof of Corollary 1.9 Since the figure-eight knot is amphicheiral, that is, it is equivalent to its mirror image, we have $J_N(\mathscr{E};q^{-1})=J_N(\mathscr{E};q)$. It follows that $J_N(\mathscr{E};e^{\xi'/N})=J_N(\mathscr{E};e^{-\xi'/N})=J_N(\mathscr{E};e^{\bar\xi/N})=\overline{J_N(\mathscr{E};e^{\xi/N})}$, where $\bar\xi$ is the complex conjugate. So we obtain

$$\begin{aligned}J_N(\mathscr{E};e^{\xi'/N})&\underset{N\to\infty}{\sim}\overline{J_p(\mathscr{E};e^{4\pi^2N/\xi})}\frac{\Gamma\big(\frac13\big)e^{-\pi\sqrt{-1}/6}}{3^{1/6}}\Big(\frac{N}{\bar\xi}\Big)^{2/3}\exp\Big(\frac{-2\kappa\pi\sqrt{-1}}{\bar\xi}N\Big)\\&=J_p(\mathscr{E};e^{4\pi^2N/\xi'})\frac{\Gamma\big(\frac13\big)e^{\pi\sqrt{-1}/6}}{3^{1/6}}\Big(\frac{N}{\xi'}\Big)^{2/3}\exp\Big(\frac{S_{-\kappa}(\mathscr{E})}{\xi'}N\Big),\end{aligned}$$

where $(\xi')^{1/3}:=|\xi'|^{1/3}e^{-\arctan(2p\pi/\kappa)\sqrt{-1}/3}e^{-\pi\sqrt{-1}/3}$. The last equality follows since $e^{-2\kappa\pi\sqrt{-1}N/\bar\xi}=e^{2\kappa\pi\sqrt{-1}N/\xi'}=e^{2\kappa\pi\sqrt{-1}N/\xi'+4N\pi\sqrt{-1}}=e^{(-2\kappa\pi\sqrt{-1}-8pN\pi^2)N/\xi'}=e^{(S_{-\kappa}(\mathscr{E})-8pN\pi^2)/\xi'}$ and the Chern–Simons invariant is defined modulo an integer multiple of π^2 (see Section 6). □

6 The Chern–Simons invariant

In this section, we show a relation between $S_\kappa(E) = 2\kappa\pi\sqrt{-1}$ appearing in Theorem 1.8 and the Chern–Simons invariant. For the definition of the Chern–Simons invariant of a representation from the fundamental group of a three-manifold with toric boundary to $\mathrm{SL}(2;\mathbb{C})$, we refer the readers to [17].

Let W be the three-manifold obtained from S^3 by removing the open tubular neighborhood of a knot $K \subset S^3$. We denote by $X(W)$ the $\mathrm{SL}(2;\mathbb{C})$ character variety, that is, the set of characters of representations from $\pi_1(W)$ to $\mathrm{SL}(2;\mathbb{C})$. Let $E(\partial W)$ be the quotient space $(\mathrm{Hom}(\pi_1(\partial W),\mathbb{C})\times\mathbb{C}^\times)/G$, where $\mathbb{C}^\times := \mathbb{C}\setminus\{0\}$ and $G := \langle x, y, b \mid xy = yx,\ bxbx = byby = b^2 = 1\rangle$ acts on $\mathrm{Hom}(\pi_1(\partial W),\mathbb{C})\times\mathbb{C}^\times$ as follows:

$$x\cdot(\alpha,\beta;z) := \big(\alpha+\tfrac{1}{2},\beta;z\exp(-4\pi\sqrt{-1}\beta)\big),\quad y\cdot(\alpha,\beta;z) := \big(\alpha,\beta+\tfrac{1}{2};z\exp(4\pi\sqrt{-1}\alpha)\big),\tag{6-1}$$
$$b\cdot(\alpha,\beta) := (-\alpha,-\beta;z).$$

Here we fix a generator $(\mu^*,\lambda^*)\in\mathrm{Hom}(\pi_1(\partial W);\mathbb{C})\cong\mathbb{C}^2$ for a meridian μ (the homotopy class of the loop that goes around K) and a preferred longitude λ (the homotopy class of the loop that goes along K so that its linking number with K is zero). Then the projection $p\colon E(\partial W)\to X(\partial W)$ sending $[\alpha,\beta;z]$ to $[\alpha,\beta]$ becomes a $\mathbb{C}^\times$-bundle, where the square brackets mean the equivalence class.

The $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant of W defines a lift $\mathrm{cs}_W\colon X(W)\to E(\partial W)$ of $X(W)\xrightarrow{i^*}X(\partial W)$, that is, $p\circ c_W = i^*$ holds, where i^* is induced by the inclusion map $i\colon\partial W\to W$:

$$\begin{array}{ccc} & & E(\partial W)\\ & \overset{\mathrm{cs}_W}{\nearrow} & \big\downarrow p\\ X(W) & \xrightarrow{\ i^*\ } & X(\partial W)\end{array}$$

For a representation ρ, we have $\mathrm{cs}_W([\rho]) = [u/(4\pi\sqrt{-1}), v/(4\pi\sqrt{-1}); \exp((2/(\pi\sqrt{-1}))\,\mathrm{CS}_{u,v}(\rho))]$ if

$$\rho(\mu) = \begin{pmatrix} e^{u/2} & * \\ 0 & e^{-u/2}\end{pmatrix}\quad\text{and}\quad \rho(\lambda) = \begin{pmatrix} e^{v/2} & * \\ 0 & e^{-v/2}\end{pmatrix},$$

up to conjugation, where $[\rho]\in X(W)$ means the equivalence class, and $\mathrm{CS}_{u,v}(\rho)$ is the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant of ρ associated with (u,v). Note that $\mathrm{CS}_{u,v}(\rho)$ is defined modulo π^2, and depends on the choice of branches of logarithms of $e^{u/2}$ and $e^{v/2}$.

Now, we calculate the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant of the figure-eight knot. See also [29, Section 5.2] for calculation about the figure-eight knot complement.

By using generators x and y as indicated in Figure 15, the fundamental group $G_{\mathscr{E}} := \pi_1(S^3\setminus\mathscr{E})$ has a presentation $\langle x, y \mid \omega x = y\omega\rangle$, where $\omega := xy^{-1}x^{-1}y$. We choose (the homotopy class of) x as the meridian μ, and (the homotopy class of) l depicted in Figure 15 as the preferred longitude λ. The loop l presents the element $x\omega^{-1}\overleftarrow{\omega}^{-1}x^{-1}\in G_{\mathscr{E}}$, where $\overleftarrow{\omega} := yx^{-1}y^{-1}x$ is the word obtained from ω by

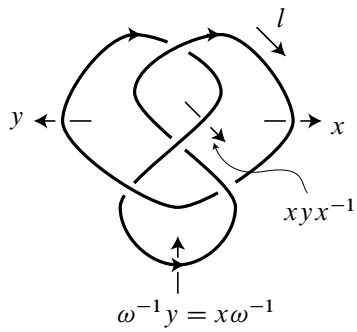


Figure 15: The figure-eight knot $\mathscr{E}$ and generators of $G_{\mathscr{E}} := \pi_1(S^3 \setminus \mathscr{E})$.

reading backward. Due to [36] (see also [22, Section 3]), for a real number u with $0 \le u \le \kappa$ we consider the nonabelian representation $\rho_u : G_{\mathscr{E}} \to \mathrm{SL}(2;\mathbb{C})$ sending x and y to

$$\begin{pmatrix} e^{u/2} & 1 \\ 0 & e^{-u/2} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} e^{u/2} & 0 \\ d & e^{-u/2} \end{pmatrix},$$

respectively, where d is given as

$$d := \tfrac{3}{2} - \cosh u + \tfrac{1}{2}\sqrt{(2\cosh(u)+1)(2\cosh(u)-3)}.$$

The preferred longitude is sent to

$$\begin{pmatrix} e^{v(u)/2} & * \\ 0 & e^{-v(u)/2} \end{pmatrix},$$

where

$$v(u) := 2\log(\cosh(2u) - \cosh(u) - 1 - \sinh(u)\sqrt{(2\cosh(u)+1)(2\cosh(u)-3)}) + 2\pi\sqrt{-1}.$$

Here we add $2\pi\sqrt{-1}$ so that $v(0) = 0$.

It is well known [39] that when $u = 0$, the irreducible representation ρ_0 induces a complete hyperbolic structure in $S^3 \setminus \mathscr{E}$, and when $0 < u < \kappa$, ρ_u is irreducible and induces an incomplete hyperbolic structure. When $u = \kappa$, the representation ρ_κ becomes reducible (and nonabelian), and the hyperbolic structure collapses. In fact, in this case, both x and y are sent to upper triangular matrices, and so every element of $G_{\mathscr{E}}$ is sent to an upper triangular matrix, which means that ρ_κ is reducible. This kind of reducible and nonabelian representation is called affine, and corresponds to the zeroes of the Alexander polynomial; see [3; 16, Exercise 11.2; 35; 40, 2.4.3. Corollary].

Now, we calculate the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant $\mathrm{CS}_{\kappa,v(\kappa)}(\rho_\kappa)$ associated with $(\kappa, v(\kappa)) = (\kappa, 2\pi\sqrt{-1})$; see [17] for details.

Since the Chern–Simons invariant of a representation is determined by its character, and ρ_κ shares the same character (trace) with the abelian representation $\rho_\kappa^{\mathrm{abel}}$ sending $\mu := x$ to the diagonal matrix

$$\begin{pmatrix} e^{\kappa/2} & 0 \\ 0 & e^{-\kappa/2} \end{pmatrix}$$

and $\lambda := l$ to the identity matrix, it can be easily seen that $\mathrm{cs}_W(\rho_\kappa^{\mathrm{abel}}) = [\kappa/(4\pi\sqrt{-1}), 0; 1]$, where we put $W := S^3 \setminus N(\mathscr{E})$ with $N(\mathscr{E})$ is the open tubular neighborhood of $\mathscr{E}$ in S^3. Since we have

$$\left[\frac{\kappa}{4\pi\sqrt{-1}}, 0; 1\right] = \left[\frac{\kappa}{4\pi\sqrt{-1}}, \frac{1}{2}; e^{\kappa}\right]$$

from (6-1), we conclude that $\mathrm{CS}_{\kappa, 2\pi\sqrt{-1}}(\rho_\kappa) = \frac{1}{2}\kappa\pi\sqrt{-1}$. Note that here we change the pair $(\kappa, 0)$ to $(\kappa, 2\pi\sqrt{-1})$.

As in [23], if we define

$$S_u(\mathscr{E}) := \mathrm{CS}_{u, v(u)}(\rho_u) + \pi\sqrt{-1}u + \tfrac{1}{4}uv(u) \tag{6-2}$$

for $0 \le u \le \kappa$, then $S_\kappa(\mathscr{E}) = 2\kappa\pi\sqrt{-1}$ when $(u, v(u)) = (\kappa, 2\pi\sqrt{-1})$.

Similarly, $\mathrm{CS}_{-\kappa, 2\pi\sqrt{-1}}(\rho_{-\kappa}) = -\frac{1}{2}\kappa\pi\sqrt{-1}$, and $S_{-u}(\mathscr{E}) = -2\kappa\pi\sqrt{-1}$.

Appendix A Proof of the Poisson summation formula

In this appendix, we give a proof of the Poisson summation formula following [30, Proposition 4.2].

Proof of Proposition 4.1 Let $\varepsilon > 0$ be small enough that

$$\operatorname{Re}\psi(a) < -\varepsilon, \qquad \operatorname{Re}\psi(b) < -\varepsilon,$$
$$\operatorname{Re}\psi(z) - 2\pi \operatorname{Im} z < -\varepsilon \quad \text{if } z \in C_+, \qquad \operatorname{Re}\psi(z) + 2\pi \operatorname{Im} z < -\varepsilon \quad \text{if } z \in C_-.$$

Then for sufficiently large N, the following also hold:

(i) $\operatorname{Re}\psi_N(a) < -\varepsilon$,

(ii) $\operatorname{Re}\psi_N(b) < -\varepsilon$,

(iii) $\operatorname{Re}\psi_N(z) - 2\pi \operatorname{Im} z < -\varepsilon$ if $z \in C_+$,

(iv) $\operatorname{Re}\psi_N(z) + 2\pi \operatorname{Im} z < -\varepsilon$ if $z \in C_-$.

Moreover, there exists $\delta > 0$ such that $\operatorname{Re}\psi_N(t) < -\varepsilon$ if $t \in [a, a+\delta]$ or $t \in [b-\delta, b]$ from (i) and (ii) for such N.

Let $\beta\colon \mathbb{R} \to [0,1]$ be a C^∞-function such that

$$\beta(t) = \begin{cases} 1 & \text{if } t \in [a+\delta, b-\delta], \\ 0 & \text{if } t < a \text{ or } t > b. \end{cases}$$

We also assume that $\beta(t)$ is in the Schwartz space $\mathcal{S}(\mathbb{R})$, that is, $\sup_{x\in\mathbb{R}}|x^m f^{(n)}(x)| < \infty$ for any nonnegative integers m and n. Put $\Psi_N(x) := \beta(x/N)e^{N\psi_N(x/N)}$.

We have

$$\left|\sum_{a \le k/N < a+\delta} e^{N\psi_N(k/N)}\right| \le \sum_{a \le k/N < a+\delta} e^{N\operatorname{Re}\psi_N(k/N)} < \delta N e^{-\varepsilon N}, \tag{A-1}$$

where the second inequality follows since $\operatorname{Re}\psi_N(k/N) < -\varepsilon$ when $a \le k/N \le a+\delta$. Similarly we have

$$\Big|\sum_{b-\delta\le k/N<b} e^{N\psi_N(k/N)}\Big| < \delta N e^{-\varepsilon N}. \tag{A-2}$$

We also have

$$\Big|\sum_{k/N<a+\delta} \Psi_N(k)\Big| \le \sum_{a\le k/N<a+\delta} \beta(k/N) e^{N\operatorname{Re}\psi_N(k/N)} < \delta N e^{-\varepsilon N} \tag{A-3}$$

and

$$\Big|\sum_{k/N>b-\delta} \Psi_N(k)\Big| < \delta N e^{-\varepsilon N}. \tag{A-4}$$

Since $\Psi_N(k) = e^{N\psi_N(k/N)}$ if $a+\delta \le k/N \le b-\delta$, we have

$$\begin{aligned}\Big|\sum_{k\in\mathbb{Z}} \Psi_N(k) - \sum_{a\le k/N\le b} e^{N\psi_N(k/N)}\Big| &\le \Big|\sum_{k/N<a+\delta} \Psi_N(k)\Big| + \Big|\sum_{a\le k/N<a+\delta} e^{N\psi_N(k/N)}\Big| + \Big|\sum_{b-\delta<k/N\le b} \Psi_N(k)\Big| \\ &\quad + \Big|\sum_{k/N>b-\delta} e^{N\psi_N(k/N)}\Big| \\ &< 4\delta N e^{-\varepsilon N}\end{aligned} \tag{A-5}$$

from (A-1)–(A-4).

Since $\Psi_N(t)$ is also in $\mathcal{S}(\mathbb{R})$, we can apply the Poisson summation formula (see eg [38, Theorem 3.1]):

$$\sum_{k\in\mathbb{Z}} \Psi_N(k) = \sum_{l\in\mathbb{Z}} \widehat{\Psi}_N(l), \tag{A-6}$$

where $\widehat{\Psi}_N$ is the Fourier transform of Ψ_N, that is, $\widehat{\Psi}_N(l) := \int_{-\infty}^{\infty} \Psi_N(t) e^{-2l\pi\sqrt{-1}t}\, dt$.

Putting $s := t/N$,

$$\widehat{\Psi}_N(l) = N\int_{-\infty}^{\infty} \beta(s) e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\, ds. \tag{A-7}$$

From the properties of $\beta(s)$, we have

$$\begin{aligned}\Big|\frac{1}{N}\widehat{\Psi}_N(0) - \int_a^b e^{N\psi_N(s)}\, ds\Big| &\le \Big|\int_a^{a+\delta} (\beta(s)-1) e^{N\psi_N(s)}\, ds\Big| + \Big|\int_{b-\delta}^{b} (\beta(s)-1) e^{N\psi_N(s)}\, ds\Big| \\ &\le \int_a^{a+\delta} (1-\beta(s)) e^{N\operatorname{Re}\psi_N(s)}\, ds + \int_{b-\delta}^{b} (1-\beta(s)) e^{N\operatorname{Re}\psi_N(s)}\, ds \\ &< 2\delta e^{-\varepsilon N}.\end{aligned} \tag{A-8}$$

Therefore

$$\text{(A-9)}\quad \Big|\frac{1}{N}\sum_{a\le k/N\le b} e^{N\psi_N(k/N)}-\int_a^b e^{N\psi_N(s)}\,ds\Big|$$
$$\le \Big|\frac{1}{N}\sum_{a\le k/N\le b} e^{N\psi_N(k/N)}-\frac{1}{N}\sum_{l\in\mathbb{Z}}\hat{\Psi}_N(l)\Big|+\Big|\frac{1}{N}\sum_{l\in\mathbb{Z}}\hat{\Psi}_N(l)-\int_a^b e^{N\psi_N(s)}\,ds\Big|$$
$$\le \Big|\frac{1}{N}\sum_{a\le k/N\le b} e^{N\psi_N(k/N)}-\frac{1}{N}\sum_{k\in\mathbb{Z}}\Psi_N(k)\Big|+\Big|\frac{1}{N}\hat{\Psi}_N(0)-\int_a^b e^{N\psi_N(s)}\,ds\Big|+\frac{1}{N}\sum_{\substack{l\in\mathbb{Z}\\ l\ne 0}}|\hat{\Psi}_N(l)|$$
$$< \frac{1}{N}\sum_{\substack{l\in\mathbb{Z}\\ l\ne 0}}|\hat{\Psi}_N(l)|+6\delta e^{-\varepsilon N},$$

where the first inequality follows from (A-6), and the second from (A-5) and (A-8). Next we calculate $\hat{\Psi}_N(l)$ for $l\ne 0$. Integrating the right-hand side of (A-7) by parts twice, we have

$$\hat{\Psi}_N(l)=\frac{1}{2l\pi\sqrt{-1}}\int_{-\infty}^{\infty}\frac{d}{ds}(\beta(s)e^{N\psi_N(s)})e^{-2l\pi\sqrt{-1}Ns}\,ds$$
$$=-\frac{1}{4l^2\pi^2N}\int_{-\infty}^{\infty}\frac{d^2}{ds^2}(\beta(s)e^{N\psi_N(s)})e^{-2l\pi\sqrt{-1}Ns}\,ds.$$

Putting

$$B_N(s):=\beta''(s)+2N\beta'(s)\psi_N'(s)+N\beta(s)\psi_N''(s)+N^2\beta(s)(\psi'(s))^2,$$
$$\widetilde{B}_N(s):=N\psi_N''(s)+N^2(\psi_N'(s))^2,$$

we have

$$-4l^2\pi^2N\hat{\Psi}_N(l)$$
$$=\int_a^b B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$$
$$=\int_{a+\delta}^{b-\delta}\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds+\int_a^{a+\delta}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$$
$$+\int_{b-\delta}^{b}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$$
$$=\int_a^b\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds-\int_a^{a+\delta}\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$$
$$-\int_{b-\delta}^{b}\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds+\int_a^{a+\delta}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$$
$$+\int_{b-\delta}^{b}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds,$$

where the second equality follows because $B_N(s)=\widetilde{B}_N(s)$ when $s\in[a+\delta,b-\delta]$. So we have

$$\text{(A-10)}\quad \Big|4l^2\pi^2N\hat{\Psi}_N(l)+\int_a^b\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\Big|$$
$$\le\Big|\int_a^{a+\delta}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\Big|+\Big|\int_{b-\delta}^{b}B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\Big|$$
$$+\Big|\int_a^{a+\delta}\widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\Big|+\Big|\int_{b-\delta}^{b}\widetilde{B}_N(s)e^{N(\psi_N(s)-2k\pi\sqrt{-1}s)}\,ds\Big|.$$

Since $\operatorname{Re}\psi_N(s) < -\varepsilon$ if $a \le s \le a+\delta$, we have

$$\text{(A-11)}\qquad \left|\int_a^{a+\delta} B_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\right| \le \int_a^{a+\delta}|B_N(s)|e^{N\operatorname{Re}\psi_N(s)}\,ds < \delta e^{-\varepsilon N}\max_{s\in[a,a+\delta]}|B_N(s)| \le K_a N^2 e^{-\varepsilon N},$$

where we put

$$K_a := \max_{s\in[a,a+\delta]}|\beta''(s)| + \max_{s\in[a,a+\delta]}|2\beta'(s)\psi_N'(s)| + \max_{s\in[a,a+\delta]}|\beta(s)\psi_N''(x)| + \max_{s\in[a,a+\delta]}|\beta(s)(\psi_N'(s))^2|$$
$$\ge \max_{s\in[a,a+\delta]}\left|\frac{\beta''(s)}{N^2} + \frac{2\beta'(s)\psi_N'(s)}{N} + \frac{\beta(s)\psi_N''(s)}{N} + \beta(s)(\psi_N'(s))^2\right| = \max_{s\in[a,a+\delta]}|B_N(s)|\frac{1}{N^2}.$$

Similarly, putting

$$K_b := \max_{s\in[b-\delta,b]}|\beta''(s)| + \max_{s\in[b-\delta,b]}|2\beta'(s)\psi_N'(s)| + \max_{s\in[b-\delta,b]}|\beta(s)\psi_N''(x)| + \max_{s\in[b-\delta,b]}|\beta(s)(\psi_N'(s))^2|,$$
$$\widetilde{K}_a := \max_{s\in[a,a+\delta]}|\psi_N''(s)| + \max_{s\in[a,a+\delta]}|(\psi_N'(s))^2|, \quad \widetilde{K}_b := \max_{s\in[b-\delta,b]}|\psi_N''(s)| + \max_{s\in[b-\delta,b]}|(\psi_N'(s))^2|,$$

we have

$$\text{(A-12)}\qquad \left|\int_{b-\delta}^{b} B_N(s)e^{N(\psi_N(s)-2k\pi\sqrt{-1}s)}\,ds\right| < K_b N^2 e^{-\varepsilon N},$$

$$\text{(A-13)}\qquad \left|\int_{a}^{a+\delta} \widetilde{B}_N(s)e^{N(\psi_N(s)-2k\pi\sqrt{-1}s)}\,ds\right| < \widetilde{K}_a N^2 e^{-\varepsilon N},$$

$$\text{(A-14)}\qquad \left|\int_{b-\delta}^{b} \widetilde{B}_N(s)e^{N(\psi_N(s)-2k\pi\sqrt{-1}s)}\,ds\right| < \widetilde{K}_b N^2 e^{-\varepsilon N}.$$

Therefore

$$\text{(A-15)}\qquad |\widehat{\Psi}_N(l)| < \frac{1}{4l^2\pi^2 N}\left|\int_a^b \widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds\right| + \frac{KN}{4l^2\pi^2}e^{-\varepsilon N}$$

from (A-11)–(A-14), where we put $K := K_a + K_b + \widetilde{K}_a + \widetilde{K}_b$.

To evaluate $\int_a^b \widetilde{B}_N(s)e^{N(\psi_N(s)-2l\pi\sqrt{-1}s)}\,ds$, we consider the paths $C_\pm \subset R_\pm$. Note that $\widetilde{B}_N$ is defined in D.

By replacing the path $[a,b]$ with $C_\pm$, we have

$$\text{(A-16)}\qquad \left|\int_a^b \widetilde{B}_N(s)e^{N(\psi_N(s)-2\pi\sqrt{-1}ls)}\,ds\right| = \left|\int_{C_\pm} \widetilde{B}_N(z)e^{N(\psi_N(z)-2\pi\sqrt{-1}lz)}\,dz\right|$$
$$\le \int_{C_\pm}|\widetilde{B}_N(z)|e^{N(\operatorname{Re}\psi_N(z)+2l\pi\operatorname{Im}z)}\,|dz|$$
$$\le \max_{z\in C_\pm}|\widetilde{B}_N(z)|\int_{C_\pm} e^{N(\operatorname{Re}\psi_N(z)+2l\pi\operatorname{Im}z)}\,|dz|$$
$$\le K_\pm N^2\int_{C_\pm} e^{N(\operatorname{Re}\psi_N(z)+2l\pi\operatorname{Im}z)}\,|dz|,$$

where we put

$$K_\pm := \max_{z\in C_\pm}|\psi_N''(z)| + \max_{z\in C_\pm}|(\psi_N'(z))^2| \ge \max_{z\in C_\pm}\left|\frac{\psi_N''(z)}{N} + (\psi_N'(z))^2\right| = \max_{z\in C_\pm}|\widetilde{B}_N(z)|\frac{1}{N^2}.$$

If $l \geq 1$, we use C_-. Since $C_- \subset R_-$, we have $\operatorname{Im} z \leq 0$ and $\operatorname{Re} \psi_N(z) + 2\pi \operatorname{Im} z < -\varepsilon$ from (iv). So from (A-16), we have

$$\left| \int_a^b \widetilde{B}_N(s) e^{N(\psi_N(s) - 2\pi\sqrt{-1}ls)}\, ds \right| < \widetilde{K}_- N^2 e^{-\varepsilon N}, \tag{A-17}$$

where $\widetilde{K}_- := K_-(\text{length of } C_-)$.

Similarly, if $l \leq -1$, putting $\widetilde{K}_+ := K_+(\text{length of } C_+)$, we have

$$\left| \int_a^b \widetilde{B}_N(s) e^{N(\psi_N(s) - 2\pi\sqrt{-1}ls)}\, ds \right| < \widetilde{K}_+ N^2 e^{-\varepsilon N} \tag{A-18}$$

from (iii).

Therefore, from (A-15)–(A-18), we have

$$\begin{aligned} \left| \sum_{l \in \mathbb{Z}, l \neq 0} \hat{\Psi}_N(l) \right| &< \sum_{l=1}^{\infty} \left(\frac{\widetilde{K}_- N}{4l^2\pi^2} e^{-\varepsilon N} + \frac{KN}{4l^2\pi^2} e^{-\varepsilon N} \right) + \sum_{l=1}^{\infty} \left(\frac{\widetilde{K}_+ N}{4l^2\pi^2} e^{-\varepsilon N} + \frac{KN}{4l^2\pi^2} e^{-\varepsilon N} \right) \\ &= \left(\frac{\widetilde{K}_-}{24} + \frac{\widetilde{K}_+}{24} + \frac{K}{12} \right) N e^{-\varepsilon N}, \end{aligned}$$

since $\sum_{l=1}^{\infty} 1/l^2 = \frac{1}{6}\pi^2$.

From (A-9), we finally have

$$\left| \frac{1}{N} \sum_{a \leq k/N \leq b} e^{N\psi(k/N)} - \int_a^b e^{N\psi_N(s)}\, ds \right| < \left(6\delta + \frac{\widetilde{K}_-}{24} + \frac{\widetilde{K}_+}{24} + \frac{K}{12} \right) e^{-\varepsilon N},$$

proving the proposition. □

Appendix B Proof of the saddle point method of order two

In this appendix, we give a proof of Proposition 5.2.

Let $c := re^{\theta\sqrt{-1}}$ be a complex number with $r > 0$ and $-\pi < \theta \leq \pi$, and put $U := \{z \in \mathbb{C} \mid \operatorname{Re}(cz^3) < 0\}$. If we write $z := se^{\tau\sqrt{-1}}$ with $s > 0$ and $\tau \in \mathbb{R}$, then since $cz^3 = rs^3 e^{(\theta+3\tau)\sqrt{-1}}$, the region U has three connected components U_k for $k = 1, 2, 3$:

$$U_k := \left\{ w \in \mathbb{C} \mid w = se^{\tau\sqrt{-1}},\ s > 0,\ \left|\tau + \tfrac{1}{3}\theta - (2k-1)\tfrac{1}{3}\pi\right| < \tfrac{1}{6}\pi \right\}. \tag{B-1}$$

Note that U_k for $k = 1, 2, 3$ is obtained from U_{k-1} by the $\frac{2}{3}\pi$-rotation around the origin O, where U_0 means U_3. The origin O is a saddle point of order two for the function $\operatorname{Re}(cz^3)$, and the regions U_k are called valleys.

First of all, we study the asymptotic behavior of the integral $\int_C h_N(z) e^{Ncz^3}\, dz$ as $N \to \infty$, where C is a path starting at the origin and going into a valley, and $h_N(z)$ is a holomorphic function depending on N. The next lemma follows from the techniques described in [42, II.4]:

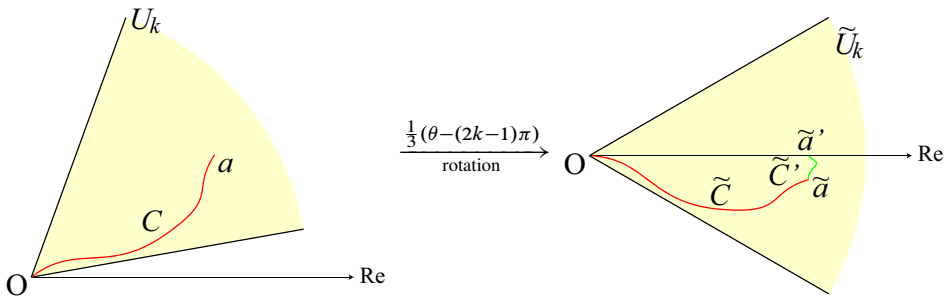


Figure 16: The yellow regions are U_k and $\widetilde{U}_k$, the red curves are C and $\widetilde{C}$, and the green curve is $\widetilde{C}'$.

Lemma B.1 *Let D be an open bounded region in $\mathbb{C}$ containing O, $h_N(z)$ be a holomorphic function in D depending on a positive integer N, and U_k be as above. We assume that $h_N(z)$ uniformly converges to a holomorphic function $h(z)$ with $h(0)\neq 0$ and that $|h_N(z)|$ is bounded irrelevant to z or N. We also assume that $U_k\cap D$ is connected and simply connected for each k. For a point $a\in U_k\cap D$, let $C\subset(U_k\cap D)\cup\{O\}$ be a path from O to a. Then we have*

$$\int_C h_N(z)e^{Ncz^3}\,dz=\frac{e^{((2k-1)\pi-\theta)\sqrt{-1}/3}h(0)\Gamma\left(\frac{1}{3}\right)}{3r^{1/3}N^{1/3}}(1+O(N^{-1/3}))$$

as $N\to\infty$, where $\Gamma(x):=\int_0^\infty t^{x-1}e^{-t}\,dt$ is the gamma function.

Proof Let $\widetilde{U}_k$ be the region obtained from U_k by the $\frac{1}{3}(\theta-(2k-1)\pi)$-rotation around O, that is,

$$\widetilde{U}_k:=\left\{w\in\mathbb{C}\mid w=se^{\tau\sqrt{-1}},\ s>0,\ |\tau|<\tfrac{1}{6}\pi\right\}. \tag{B-2}$$

The same rotation sends D to $\widetilde{D}$, C to $\widetilde{C}\subset(\widetilde{U}_k\cap\widetilde{D})\cup\{O\}$, and a to $\tilde{a}:=e^{(\theta-(2k-1)\pi)\sqrt{-1}/3}a\in\widetilde{U}_k\cap\widetilde{D}$; see Figure 16.

Putting

$$w:=e^{(\theta-(2k-1)\pi)\sqrt{-1}/3}z$$

and

$$\tilde{h}_N(w):=h_N(e^{((2k-1)\pi-\theta)\sqrt{-1}/3}w),$$

we have

$$\int_C h_N(z)e^{Ncz^3}\,dz=e^{((2k-1)\pi-\theta)\sqrt{-1}/3}\int_{\widetilde{C}}\tilde{h}_N(w)e^{-Nrw^3}\,dw. \tag{B-3}$$

Since $\widetilde{U}_k\cap\widetilde{D}$ is connected, we can choose $\tilde{a}'>0$ in $\mathbb{R}\cap\widetilde{U}_k\cap\widetilde{D}$ and connect $\tilde{a}$ to $\tilde{a}'$ by a path $\widetilde{C}'\subset\widetilde{U}_k\cap\widetilde{D}$. Now the function $\tilde{h}_N$ is defined in $\widetilde{D}$, and we will extend $\tilde{h}_N|_{\widetilde{U}_k\cap\widetilde{D}\cap\mathbb{R}}$ to a C^∞ function $h_N^*(t)$ for any $t\geq 0$. Here we assume the following:

(i) $h_N^*(t)$ is bounded.

(ii) $h_N^*(t)$ converges uniformly to a C^∞ function $h^*(t)$.

(iii) $h_N^*(t)=\tilde{h}_N(t)$ and $h^*(t)=\tilde{h}(t):=h(e^{((2k-1)\pi-\theta)\sqrt{-1}/3}t)$ for $t\in\widetilde{U}_k\cap\widetilde{D}\cap\mathbb{R}$.

Then since $\widetilde{U}_k \cap \widetilde{D}$ is simply connected, by Cauchy's theorem we have

$$\int_{\widetilde{C}} \tilde{h}_N(w) e^{-Nrw^3}\, dw = I_1 - I_2 - I_3, \tag{B-4}$$

where we put

$$I_1 := \int_0^\infty h_N^*(w) e^{-Nrw^3}\, dw, \quad I_2 := \int_{\tilde{a}'}^\infty h_N^*(w) e^{-Nrw^3}\, dw, \quad I_3 := \int_{\widetilde{C}'} \tilde{h}_N(w) e^{-Nrw^3}\, dw.$$

We use Watson's lemma [41] to evaluate I_1. Putting $t := w^3$, we have

$$I_1 = \int_0^\infty h_N^*(t^{1/3}) \frac{1}{3t^{2/3}} e^{-Nrt}\, dt.$$

Since $h_N^*(s)$ uniformly converges to an analytic function $h^*(s)$ in $\widetilde{D} \cap \mathbb{R}$, we conclude that

$$h_N^*(s) = h^*(s) + \frac{g_N(s)}{N}$$

with $|g_N(s)| < c$, where c is a constant independent of s. Since $h^*(0) = h(0)$, $h_N^*(s)$ is of the form

$$h_N^*(s) = h(0) + \frac{g_N(s)}{N} + \sum_{j=1}^\infty b_j s^j$$

near 0, where $b_j := (1/j!)(d^j/ds^j)h(0)$. So we have

$$h_N^*(t^{1/3}) \frac{1}{3t^{2/3}} = \tfrac{1}{3} h(0) t^{-2/3} + \frac{g_N(t^{1/3})}{3t^{2/3}N} + \sum_{j=1}^\infty \tfrac{1}{3} b_j t^{(j-2)/3}.$$

Since $|g_N(s)| < c$,

$$\left| \int_0^\infty \frac{g_N(t^{1/3})}{3t^{2/3}N} e^{-Nrt}\, dt \right| < \frac{c}{3N} \int_0^\infty t^{-2/3} e^{-Nrt}\, dt = \frac{c\Gamma\left(\frac{1}{3}\right)}{3r^{1/3}N^{4/3}}.$$

Therefore from Watson's lemma [41, page 133] (see also [42, page 20]), we have

$$I_1 = \frac{h(0)\Gamma\left(\frac{1}{3}\right)}{3(rN)^{1/3}} + \sum_{j=1}^\infty \tfrac{1}{3} b_j \Gamma\left(\tfrac{1}{3}(j+1)\right)(rN)^{-(j+1)/3} + O(N^{-4/3}) = \frac{h(0)\Gamma\left(\frac{1}{3}\right)}{3(rN)^{1/3}} + O(N^{-2/3}) \tag{B-5}$$

as $N \to \infty$.

As for I_2, since $|h_N^*(w)| < M$ if $w \in \mathbb{R}$ for some $M > 0$, we have

$$|I_2| \le \int_{\tilde{a}'}^\infty |h_N^*(w)| e^{-rN\tilde{a}'^2 w}\, dw = \frac{M e^{-\tilde{a}'^3 rN}}{\tilde{a}'^2 rN} < M_1 e^{-\varepsilon_1 N} \tag{B-6}$$

if $N > 1$, where we put $M_1 := M/(r\tilde{a}'^2)$ and $\varepsilon_1 := r\tilde{a}'^3 > 0$.

As for I_3, we note that if $w \in \widetilde{C}' \subset \widetilde{U}_k$, then $\operatorname{Re} w^3 > \varepsilon_2$ for some $\varepsilon_2 > 0$, since $|\arg(w^3)| < \frac{1}{2}\pi$ from (B-2). So

$$|I_3| < \max_{w \in \widetilde{C}'} |\tilde{h}_N(w)| \int_{\widetilde{C}'} e^{-Nr\varepsilon_2}\, dw \le M_2 e^{-r\varepsilon_2 N}, \tag{B-7}$$

where we put $M_2 := \max_{w\in\widetilde{C}'}|\tilde{h}_N(w)|(\text{length of } \widetilde{C}')$.

From (B-4), (B-6), and (B-7), we have

$$\left|\int_{\widetilde{C}} \tilde{h}_N(z)e^{-Nrw^3}\,dw - I_1\right| \le |I_2| + |I_3| = O(e^{-\varepsilon_3 N}),$$

with $\varepsilon_3 := \min\{\varepsilon_1, r\varepsilon_2\}$. Therefore from (B-3) and (B-5) we finally have

$$\int_C h_N(z)e^{Ncz^3}\,dz = \frac{e^{((2k-1)\pi-\theta)\sqrt{-1}/3}h(0)\Gamma\left(\frac{1}{3}\right)}{3r^{1/3}N^{1/3}}(1+O(N^{-1/3})). \qquad \square$$

Corollary B.2 *Let $c := re^{\theta\sqrt{-1}}$, D, $h_N(z)$, $h(z)$, and U_k be as in Lemma B.1. Let $C \subset D$ be a path from $a_k \in U_k \cap D$ to $a_{k+1} \in U_{k+1} \cap D$, where U_4 means U_1. We also assume that there exist paths C_k from a_k to O and C_{k+1} from O to a_{k+1} with the following properties:*

(i) $C_k \setminus \{O\} \subset U_k \cap D$.

(ii) $C_{k+1} \setminus \{O\} \subset U_{k+1} \cap D$.

(iii) *The path $C_k \cup C_{k+1}$ is homotopic to C in D keeping a_k and a_{k+1} fixed.*

Then

$$\int_C h_N(z)e^{Ncz^3}\,dz = \frac{h(0)\Gamma\left(\frac{1}{3}\right)}{\sqrt{3}r^{1/3}N^{1/3}}\sqrt{-1}\omega^k e^{-\theta\sqrt{-1}/3}(1+O(N^{-1/3})),$$

where we put $\omega := e^{2\pi\sqrt{-1}/3}$.

Proof By Cauchy's theorem, $\int_C h_N(z)e^{Ncz^3}\,dz = \int_{C_k\cup C_{k+1}} h_N(z)e^{Ncz^3}\,dz$. Then from Lemma B.1 we have

$$\begin{aligned}\int_{C_k\cup C_{k+1}} h_N(z)e^{Ncz^3}\,dz &= \frac{e^{-\theta\sqrt{-1}/3}h(0)\Gamma\left(\frac{1}{3}\right)}{3r^{1/3}N^{1/3}}(e^{(2k+1)\pi\sqrt{-1}/3} - e^{(2k-1)\pi\sqrt{-1}/3})(1+O(N^{-1/3}))\\ &= \frac{e^{-\theta\sqrt{-1}/3}h(0)\Gamma\left(\frac{1}{3}\right)}{\sqrt{3}r^{1/3}N^{1/3}}\sqrt{-1}\omega^k(1+O(N^{-1/3})),\end{aligned}$$

completing the proof. $\square$

Proof of Proposition 5.2 We use Cauchy's theorem to study the integral $\int_{P_k\cup P_{k+1}} h_N(z)e^{N\eta(z)}\,dz$. Since any point on P_k or P_{k+1} outside $\widehat{D}$ satisfies the inequality $\operatorname{Re}\eta(z) < -\varepsilon$ for some $\varepsilon > 0$, the integrals along P_k and P_{k+1} outside $\widehat{D}$ are of order $O(e^{-\varepsilon N})$. So it is enough to show that the integral $\int_P h_N(z)e^{N\eta(z)}\,dz$ equals the right-hand side of (5-1), where we put $P := (-P_k \cup P_{k+1}) \cap \widehat{D}$.

Define the function G so that $\eta(z) = re^{\theta\sqrt{-1}}G(z)^3$ and G is a bijection from $\widehat{D}$ to $E := G(\widehat{D})$, as described at the beginning of Section 5. Let $\widehat{P}$ be the image of P by G, and a_k and a_{k+1} be the endpoints of $\widehat{P}$ with $a_k \in V_k$ and $a_{k+1} \in V_{k+1}$. Putting $w := G(z)$ and $c := re^{\theta\sqrt{-1}}$, we have

$$\int_P h_N(z)e^{N\eta(z)}\,dz = \int_{\widehat{P}} \gamma_N(w)e^{Ncw^3}\,dw,$$

since $\eta(z) = re^{\theta\sqrt{-1}}G(z)^3$, where $\gamma_N(w) := h_N(G^{-1}(w))(dG^{-1}(w)/dw)$. Since $(d/dz)G(0) = 1$ and $\gamma_N(w)$ converges to $\gamma(w) := h(G^{-1}(w))(dG^{-1}(w)/dw)$, we have $\gamma(0) = h(0)$. So from Corollary B.2, we conclude

$$\int_{\widehat{P}} h_N(z)e^{N\eta(z)}\,dz = \frac{h(0)\Gamma\left(\frac{1}{3}\right)}{\sqrt{3}r^{1/3}N^{1/3}}\sqrt{-1}\omega^k e^{-\theta\sqrt{-1}/3}(1+O(N^{-1/3})),$$

completing the proof. □

Appendix C Some computer calculations on the stevedore knot

Theorem 1.8 says that the colored Jones polynomial of the figure-eight knot $\mathscr{E}$ evaluated at $(2\pi\sqrt{-1}+\kappa)/N$ grows exponentially with growth rate determined by the Chern–Simons invariant of an affine representation associated with the pair $(\kappa, 2\pi\sqrt{-1})$, where $e^{\kappa} = \frac{1}{2}(3+\sqrt{5})$ is a zero of the Alexander polynomial $\Delta(\mathscr{E};t) = -t+3-t^{-1}$. Corollary 1.9 says that the same is true for $-\kappa$.

In this appendix, we use the computer programs Mathematica and PARI/GP [33] to study the asymptotic behavior of $J_N(\mathscr{S}; e^{(2\pi\sqrt{-1}\pm\tilde{\kappa})/N})$ for the stevedore knot $\mathscr{S}$ with $\tilde{\kappa} := \log 2$, expecting a similar asymptotic behavior as $\mathscr{E}$. Note that $e^{\pm\tilde{\kappa}} = 2^{\pm 1}$ annihilates the Alexander polynomial $\Delta(\mathscr{S};t) := -2t+5-2t^{-1}$ of $\mathscr{S}$.

The stevedore knot $\mathscr{S}$ is the mirror image of the 6_1 knot in Rolfsen's book [37] (see also the knot atlas [2]) as depicted in Figure 17. Note that in KnotInfo [19] it is denoted by 6_1.

Due to [20, Theorem 5.1], we obtain

$$J_N(\mathscr{S};q) = \sum_{k=0}^{N-1} q^{-k(N+k+1)} \prod_{a=1}^{k}((1-q^{N+a})(1-q^{N-a})) \sum_{l=0}^{k} q^{l(k+1)} \frac{\prod_{b=l+1}^{k}(1-q^b)}{\prod_{c=1}^{k-l}(1-q^c)}.$$

Put $J_N^{\pm} := J_N(\mathscr{S}; e^{(2\pi\sqrt{-1}\pm\tilde{\kappa})/N})$. By using PARI/GP [33], we calculate $(2\pi\sqrt{-1}\pm\tilde{\kappa})\log(J_{N+1}^{\pm}/J_N^{\pm})$ for $N = 2,3,4,\dots,200$, and plot them by using Mathematica in Figures 18 and 19. The plots indicate that J_N^{+} grows like $\exp\bigl((S_+/(2\pi\sqrt{-1}+\tilde{\kappa}))N\bigr)$(polynomial in N) with

$$S_+ := -6.485 + 5.697\sqrt{-1}, \tag{C-1}$$

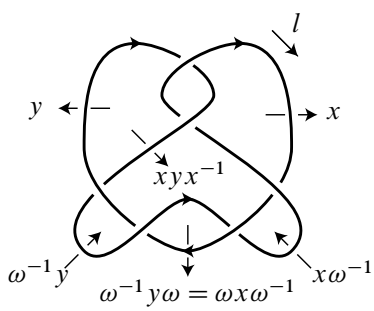


Figure 17: The stevedore knot.

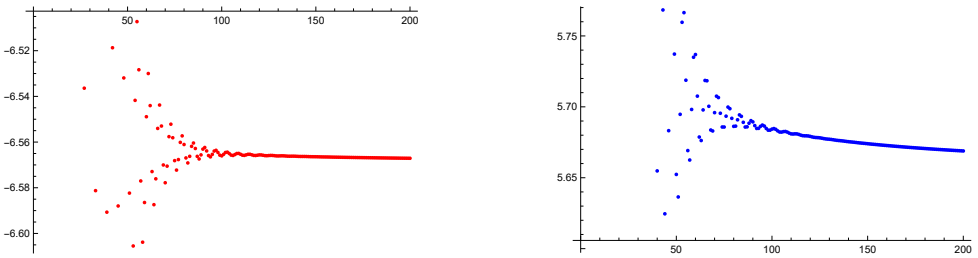


Figure 18: Plots of the real (left) and imaginary (right) parts of $(2\pi\sqrt{-1}+\tilde{\kappa})\log(J^+_{N+1}/J^+_N)$ with $N=2,3,4,\dots,200$.

and that J^-_N grows like $\exp\big((S_-/(2\pi\sqrt{-1}-\tilde{\kappa}))N\big)$(polynomial in N) with

(C-2) $$S_- := -0.06880 + 8.747\sqrt{-1}.$$

Here we use Mathematica again to find the constants $S_\pm$ such that $S_\pm + c_{\pm,1}N^{-1} + c_{\pm,2}N^{-2}$ best fits the data. Note that the constants $S_\pm$ are defined modulo integral multiples of $2\pi\sqrt{-1}(2\pi\sqrt{-1}\pm\tilde{\kappa})$, and that they may also be defined modulo integral multiples of π^2 because of the definition of the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant (see Section 6).

From Theorem 1.8, we expect that $S_\pm = \pm 2\tilde{\kappa}\pi\sqrt{-1}$. However, since $\pm 2\tilde{\kappa}\pi\sqrt{-1} = \pm 4.355\sqrt{-1}$, neither S_+ nor S_- fits with $\pm 2\tilde{\kappa}\pi\sqrt{-1}$ even modulo $2\pi\sqrt{-1}(\pm 2\tilde{\kappa}+2\pi\sqrt{-1})$ or π^2.

Now let us seek for other interpretations of $S_\pm$.

Let x and y be elements in the fundamental group $G_{\mathscr{S}} := \pi_1(S^3\setminus\mathscr{S})$ as indicated in Figure 17. Then the group $G_{\mathscr{S}}$ has the presentation

$$G_{\mathscr{S}} = \langle x, y \mid \omega^2 x = y\omega^2\rangle,$$

where we put $\omega := xy^{-1}x^{-1}y$ as in the case of the figure-eight knot. The preferred longitude l is given as $x^3\omega^{-2}\overleftarrow{\omega}^{-2}x^{-3}$, where $\overleftarrow{\omega} := yx^{-1}y^{-1}x$, as before.

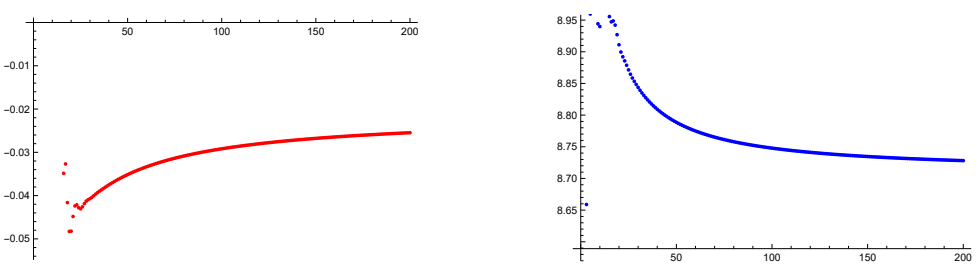


Figure 19: Plots of the real (left) and imaginary (right) parts of $(2\pi\sqrt{-1}-\tilde{\kappa})\log(J^-_{N+1}/J^-_N)$ with $N=2,3,4,\dots,200$.

Let $\rho\colon G_{\mathscr{S}} \to \mathrm{SL}(2;\mathbb{C})$ be a nonabelian representation. Due to R Riley, it is of the form

$$\rho(x) = \begin{pmatrix} m^{1/2} & 1 \\ 0 & m^{-1/2} \end{pmatrix}, \quad \rho(y) = \begin{pmatrix} m^{1/2} & 0 \\ d & m^{-1/2} \end{pmatrix}$$

up to conjugation, for some $m \neq 0$ and d.

From the relation $\omega^2 x = y\omega^2$, d and m should satisfy the following equation, known as Riley's equation:

$$d^4 + (2(m+m^{-1})-5)d^3 + ((m^2+m^{-2})-6(m+m^{-1})+13)d^2 \\ -(m^2+m^{-2}-7(m+m^{-1})+14)d - (2(m+m^{-1})-5) = 0.$$

We call the left-hand side of this equation the Riley polynomial.

If $(m,d) = (1, 0.1049 + 1.552\sqrt{-1})$, then ρ is the holonomy representation of $G_{\mathscr{S}}$ and defines the complete hyperbolic structure of $S^3 \setminus \mathscr{S}$. If $(m,d) = (2,0)$ or $\left(\frac{1}{2}, 0\right)$, then ρ gives an affine representation.

Let us consider irreducible representations corresponding to $1 \leq m \leq 2$.

The Riley polynomial is a quartic equation with respect to d, and there are four solutions, $d_1(m)$, $d_2(m)$, $d_3(m)$, and $d_4(m)$. To describe them we introduce the following functions. Let $D(m)$ be the discriminant of the Riley polynomial with respect to d, that is,

$$D(m) := 5(m^6+m^{-6}) - 32(m^5+m^{-5}) + 56(m^4+m^{-4}) - 118(m^3+m^{-3}) + 124(m^2+m^{-2}) \\ +32(m+m^{-1}) + 123.$$

We also put

$$A(m) := 4B(m)C(m)^{-1/3} + 4C(m)^{1/3} + 3(2(m+m^{-1})-5)^2 - 8(m^2+m^{-2}-6(m+m^{-1})+13),$$

where

$$B(m) := m^4 + m^{-4} - 6(m^3+m^{-3}) + 5(m^2+m^{-2}) + 3(m+m^{-1}) + 9,$$
$$C(m) := \tfrac{3}{2}\sqrt{3}\sqrt{-D(m)} + m^6 + m^{-6} - 9(m^5+m^{-5}) + 21(m^4+m^{-4}) - \tfrac{9}{2}(m^3+m^{-3}) + 6(m^2+m^{-2}) \\ -27(m+m^{-1}) - \tfrac{31}{2}.$$

We also put

$$J_{\pm}(m) := \pm 3\sqrt{3}(2(m+m^{-1})+1)A(m)^{-1/2} - 2B(m)C(m)^{-1/3} - 2C(m)^{1/3} \\ -8(m^2+m^{-2}-6(m+m^{-1})+13) + 3(2(m+m^{-1})-5)^2.$$

Now define the following four functions for $1 \leq m \leq 2$:

$$d_1(m) := \tfrac{-1}{12}(6(m+m^{-1}) - 15 + \sqrt{3}\sqrt{A(m)} + \sqrt{6}\sqrt{J_-(m)}),$$
$$d_2(m) := \tfrac{-1}{12}(6(m+m^{-1}) - 15 + \sqrt{3}\sqrt{A(m)} - \sqrt{6}\sqrt{J_-(m)}),$$
$$d_3(m) := \tfrac{-1}{12}(6(m+m^{-1}) - 15 - \sqrt{3}\sqrt{A(m)} + \sqrt{6}\sqrt{J_+(m)}),$$
$$d_4(m) := \tfrac{-1}{12}(6(m+m^{-1}) - 15 - \sqrt{3}\sqrt{A(m)} - \sqrt{6}\sqrt{J_+(m)}).$$

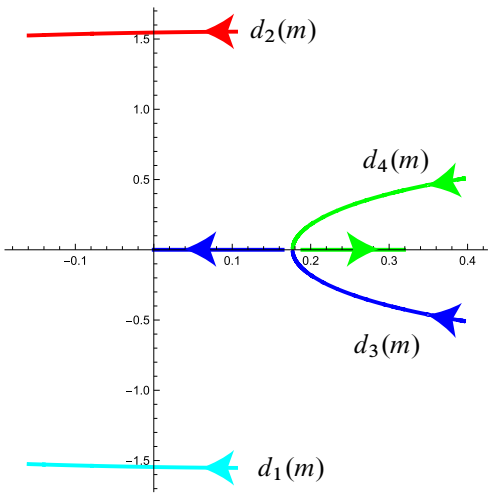


Figure 20: The cyan, red, blue, and green curves indicate $d_1(m)$, $d_2(m)$, $d_3(m)$, and $d_4(m)$, respectively. The arrows indicate the directions of increase with respect to m.

Note that

- $A(m)$, $B(m)$, $D(m)$, $J_+(m)$, and $J_-(m)$ are in $\mathbb{R}$,
- $A(m) > 0$, $B(m) > 0$, and $J_-(m) < 0$ for $1 \le m \le 2$,
- $D(m) > 0$ for $1 \le m < m_0$, $D(m_0) = 0$, and $D(m) < 0$ for $m_0 < m \le 2$, where $m_0 = 1.950$ is the unique solution to the equation $D(m) = 0$ between 1 and 2,
- $J_+(m) < 0$ for $1 \le m < m_0$, $J_+(m_0) = 0$, and $J_+(m) > 0$ for $m_0 < m \le 2$,
- $\operatorname{Im} C(m) = 0$ and $\operatorname{Re} C(m) > 0$ for $m_0 \le m \le 2$, and $\operatorname{Im} C(m) > 0$ for $1 \le m < m_0$,

which are checked by Mathematica (the author does not have proofs).

We plot, by using Mathematica, the complex-valued functions $d_i(m)$ (where $i = 1, 2, 3, 4$) for $1 \le m \le 2$ on the complex plane as in Figure 20. The following facts are also suggested by Mathematica (see Figure 20):

- $d_2(m) = \overline{d_1(m)}$ and $d_4(m) = \overline{d_3(m)}$ for $1 \le m \le 2$.
- $d_2(1) = 0.1049 + 1.552\sqrt{-1}$ and $d_2(2) = -0.1595 + 1.525\sqrt{-1}$.
- $d_3(m) \in \mathbb{R}$ and $d_4(m) \in \mathbb{R}$ for $m_0 \le m \le 2$.
- $d_3(m_0) = d_4(m_0) = 0.1770$, $d_3(2) = 0$, and $d_4(2) = 0.3189$.
- $d_3(1) = 0.3951 - 0.5068\sqrt{-1}$.

Therefore, for each i, $d_i(m)$ gives an irreducible representation $\rho_m : G_{\mathscr{S}} \to \mathrm{SL}(2;\mathbb{C})$ except for $d_3(2)$, and if $m \ne m_0$ they are mutually distinct.

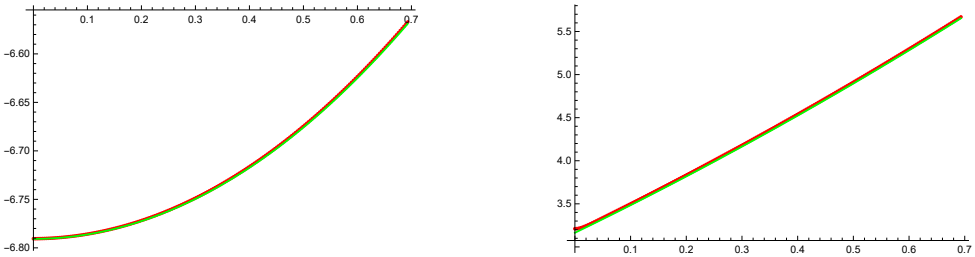


Figure 21: The left picture shows the plots of the real parts of $(2\pi\sqrt{-1}+u)\log(J_{201}(u)/J_{200}(u))$ (red) and $S_u(\mathscr{S})$ (green) for $0 \le u \le \log 2$, and the right picture shows the plots of the imaginary parts of $(u+2\pi\sqrt{-1})\log(J_{201}(u)/J_{200}(u))$ (red) and $S_u(\mathscr{S})$ (green) for $0 \le u \le \log 2$, where we put $J_N(u) := J_N(\mathscr{S}; e^{(u+2\pi\sqrt{-1})/N})$ and we use PARI/GP and Mathematica.

If we write $\rho_{d_i(m)}$ for the irreducible representation corresponding to $d_i(m)$, then we have the following:

- $\rho_{d_i(1)}$ is a parabolic representation for $i = 1, 2, 3, 4$.
- $\rho_{d_3(2)}$ is an affine representation since $d_3(2) = 0$.
- $\rho_{d_2(1)}$ is the holonomy representation, and $\rho_{d_1(1)}$ gives the holonomy representation for the mirror image of $\mathscr{S}$, because

$$\rho_{d_2(1)}(l) = \begin{pmatrix} -1-1.827-2.565\sqrt{-1} \\ 0 \quad\quad -1 \end{pmatrix}, \quad \rho_{d_1(1)}(l) = \begin{pmatrix} -1-1.827+2.565\sqrt{-1} \\ 0 \quad\quad -1 \end{pmatrix}.$$

Let $\lambda(m)$ be the $(1,1)$-entry of $\rho_{d_2(m)}(l)$, and put $v(u) := 2\log\lambda(e^{u/2})$, where we choose the logarithm branch so that $v(0) = 0$. Then the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant of $\rho_{d_2(e^{u/2})}$ associated with $(u, v(u))$ is given as

$$\mathrm{CS}_{u,v(u)}(\rho_{d_2(e^{u/2})}) = \mathrm{cv}(S^3 \setminus \mathscr{S}) + \frac{1}{2}\int_0^u v(s)\,ds - \frac{1}{4}uv(u),$$

where $\mathrm{cv}(S^3 \setminus \mathscr{S}) = -6.791 + 3.164\sqrt{-1}$ is the complex volume, which is defined to be

$$\sqrt{-1}\,\mathrm{Vol}(S^3 \setminus \mathscr{S}) - 2\pi^2\,\mathrm{CS}^{\mathrm{SO}(3)}(S^3 \setminus \mathscr{S}) \pmod{\pi^2},$$

with $\mathrm{CS}^{\mathrm{SO}(3)}$ the $\mathrm{SO}(3)$ Chern–Simons invariant of the Levi-Civita connection. Here the complex volume and $\mathrm{SO}(3)$ Chern–Simons invariant are taken from KnotInfo, where $\mathrm{Vol}(S^3 \setminus \mathscr{S}) = 3.163963229$ and $\mathrm{CS}^{\mathrm{SO}(3)}(S^3 \setminus \mathscr{S}) = 0.155977017$. Observe that $-0.155977017\pi^2 + \pi^2 = 6.79074$. Note that $\mathrm{cv}(S^3 \setminus \mathscr{S})$ coincides with the $\mathrm{SL}(2;\mathbb{C})$ Chern–Simons invariant $\mathrm{CS}_{(0,0)}(\rho_2(1))$; see [29, Chapter 5].

Putting

$$S_u(\mathscr{S}) := \mathrm{CS}_{u,v(u)}(\rho_{d_2(e^{u/2})}) + u\pi\sqrt{-1} + \tfrac{1}{4}uv(u), \tag{C-3}$$

the graphs depicted in Figure 21 indicate that

$$J_N(\mathscr{S}; e^{(u+2\pi\sqrt{-1})/N}) \underset{N\to\infty}{\sim} (\text{polynomial in } N)\exp\left(\frac{S_u(\mathscr{S})}{u+2\pi\sqrt{-1}}N\right)$$

for $0 \le u \le \tilde{\kappa} = \log 2$. When $u = \tilde{\kappa}$, Mathematica calculates $S_{\tilde{\kappa}}(\mathscr{S}) = -6.569 + 5.653\sqrt{-1}$, which is close to S_+ appearing in (C-1).

Note that the case $u = \tilde{\kappa}$ does not correspond to an affine representation. This also suggests that for $0 < u \le \tilde{\kappa}$ the representation $d_2(e^{u/2})$ induces an incomplete hyperbolic structure of $S^3 \setminus \mathscr{S}$, but the author does not know whether it is correct or not. The author does not know either any topological/geometric interpretation about the asymptotic behavior of $J_N(\mathscr{S}; e^{(u+2\pi\sqrt{-1})/N})$ for $u < 0$.

Compare this with Theorem 1.8 and Corollary 1.9, where $S_{\pm\kappa}(\mathscr{E}) = \pm 2\kappa\pi\sqrt{-1}$ are the Chern–Simons invariants of affine representations, which correspond to the fact that when $u = \pm\kappa$ the hyperbolic structure collapses.

References

[1] **J E Andersen**, **S K Hansen**, *Asymptotics of the quantum invariants for surgeries on the figure 8 knot*, J. Knot Theory Ramifications 15 (2006) 479–548 MR

[2] **D Bar-Natan**, **S Morrison**, et al., *The knot atlas*, online resource (2005) Available at `http://katlas.org/`

[3] **G Burde**, *Darstellungen von Knotengruppen*, Math. Ann. 173 (1967) 24–33 MR

[4] **T Dimofte**, **S Gukov**, *Quantum field theory and the volume conjecture*, 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) 41–67 MR

[5] **T Dimofte**, **S Gukov**, **J Lenells**, **D Zagier**, *Exact results for perturbative Chern–Simons theory with complex gauge group*, Commun. Number Theory Phys. 3 (2009) 363–443 MR

[6] **L D Faddeev**, *Discrete Heisenberg–Weyl group and modular group*, Lett. Math. Phys. 34 (1995) 249–254 MR

[7] **M Gromov**, *Volume and bounded cohomology*, Inst. Hautes Études Sci. Publ. Math. 56 (1982) 5–99 MR

[8] **S Gukov**, *Three-dimensional quantum gravity, Chern–Simons theory, and the A-polynomial*, Comm. Math. Phys. 255 (2005) 577–627 MR

[9] **S Gukov**, **H Murakami**, $\mathrm{SL}(2,\mathbb{C})$ *Chern–Simons theory and the asymptotic behavior of the colored Jones polynomial*, from "Modular forms and string duality" (N Yui, H Verrill, C F Doran, editors), Fields Inst. Commun. 54, Amer. Math. Soc., Providence, RI (2008) 261–277 MR

[10] **K Habiro**, *On the colored Jones polynomials of some simple links*, from "Recent progress towards the volume conjecture" (H Murakami, editor), Sūrikaisekikenkyūsho Kōkyūroku 1172, RIMS, Kyoto (2000) 34–43 MR

[11] **K Hikami**, **H Murakami**, *Colored Jones polynomials with polynomial growth*, Commun. Contemp. Math. 10 (2008) 815–834 MR

[12] **W H Jaco**, **P B Shalen**, *Seifert fibered spaces in 3-manifolds*, Mem. Amer. Math. Soc. 220, Amer. Math. Soc., Providence, RI (1979) MR

[13] **K Johannson**, *Homotopy equivalences of 3-manifolds with boundaries*, Lecture Notes in Math. 761, Springer (1979) MR

[14] **R M Kashaev**, *A link invariant from quantum dilogarithm*, Modern Phys. Lett. A 10 (1995) 1409–1418 MR

[15] **R M Kashaev**, *The hyperbolic volume of knots from the quantum dilogarithm*, Lett. Math. Phys. 39 (1997) 269–275 MR

[16] **L H Kauffman**, *On knots*, Annals of Math. Studies 115, Princeton Univ. Press (1987) MR

[17] **P Kirk**, **E Klassen**, *Chern–Simons invariants of 3-manifolds decomposed along tori and the circle bundle over the representation space of T^2*, Comm. Math. Phys. 153 (1993) 521–557 MR

[18] **T T Q Le**, *Quantum invariants of 3-manifolds: integrality, splitting, and perturbative expansion*, from "Proceedings of the Pacific Institute for the Mathematical Sciences Workshop "Invariants of Three-Manifolds"" (J Bryden, editor), volume 127 (2003) 125–152 MR

[19] **C Livingston**, **A H Moore**, *KnotInfo*: *table of knot invariants*, electronic reference (2023) Available at http://knotinfo.org

[20] **G Masbaum**, *Skein-theoretical derivation of some formulas of Habiro*, Algebr. Geom. Topol. 3 (2003) 537–556 MR

[21] **H Murakami**, *The colored Jones polynomials and the Alexander polynomial of the figure-eight knot*, JP J. Geom. Topol. 7 (2007) 249–269 MR

[22] **H Murakami**, *An introduction to the volume conjecture and its generalizations*, Acta Math. Vietnam. 33 (2008) 219–253 MR

[23] **H Murakami**, *The coloured Jones polynomial, the Chern–Simons invariant, and the Reidemeister torsion of the figure-eight knot*, J. Topol. 6 (2013) 193–216 MR

[24] **H Murakami**, *The colored Jones polynomial of the figure-eight knot and a quantum modularity*, Canad. J. Math. 76 (2024) 519–554 MR

[25] **H Murakami**, **J Murakami**, *The colored Jones polynomials and the simplicial volume of a knot*, Acta Math. 186 (2001) 85–104 MR

[26] **H Murakami**, **J Murakami**, **M Okamoto**, **T Takata**, **Y Yokota**, *Kashaev's conjecture and the Chern–Simons invariants of knots and links*, Experiment. Math. 11 (2002) 427–435 MR

[27] **H Murakami**, **A T Tran**, *On the asymptotic behavior of the colored Jones polynomial of the figure-eight knot associated with a real number*, from "Low dimensional topology and number theory" (M Morishita, H Nakamura, J Ueki, editors), Springer Proc. Math. Stat. 456, Springer (2025) 175–209 MR

[28] **H Murakami**, **Y Yokota**, *The colored Jones polynomials of the figure-eight knot and its Dehn surgery spaces*, J. Reine Angew. Math. 607 (2007) 47–68 MR

[29] **H Murakami**, **Y Yokota**, *Volume conjecture for knots*, SpringerBriefs in Mathematical Physics 30, Springer (2018) MR

[30] **T Ohtsuki**, *On the asymptotic expansion of the Kashaev invariant of the 5_2 knot*, Quantum Topol. 7 (2016) 669–735 MR

[31] **T Ohtsuki**, *On the asymptotic expansions of the Kashaev invariant of hyperbolic knots with seven crossings*, Internat. J. Math. 28 (2017) art. id. 1750096 MR

[32] **T Ohtsuki**, **Y Yokota**, *On the asymptotic expansions of the Kashaev invariant of the knots with 6 crossings*, Math. Proc. Cambridge Philos. Soc. 165 (2018) 287–339 MR

[33] *PARI/GP*, Univ. Bordeaux (2023) (computer algebra system) version 2.15.3 Available at http://pari.math.u-bordeaux.fr/

[34] **J Porti**, *Torsion de Reidemeister pour les variétés hyperboliques*, Mem. Amer. Math. Soc. 612, Amer. Math. Soc., Providence, RI (1997) MR

[35] **G de Rham**, *Introduction aux polynômes d'un nœud*, Enseign. Math. 13 (1967) 187–194 MR

[36] **R Riley**, *Nonabelian representations of 2-bridge knot groups*, Quart. J. Math. Oxford Ser. 35 (1984) 191–208 MR

[37] **D Rolfsen**, *Knots and links*, Math. Lecture Series 7, Publish or Perish, Houston, TX (1990) MR

[38] **E M Stein**, **R Shakarchi**, *Fourier analysis*: *an introduction*, Princeton Lectures in Analysis 1, Princeton Univ. Press (2003) MR

[39] **W P Thurston**, *The geometry and topology of three-manifolds*, *IV*, Amer. Math. Soc., Providence, RI (2022) MR

[40] **L T K Tkhang**, *Varieties of representations and their subvarieties of cohomology jumps for knot groups*, Mat. Sb. 184 (1993) 57–82 MR In Russian; translated in Russian Acad. Sci. Sb. Math. 78 (1994) 187–209

[41] **G N Watson**, *The Harmonic Functions Associated with the Parabolic Cylinder*, Proc. London Math. Soc. 17 (1918) 116–148 MR

[42] **R Wong**, *Asymptotic approximations of integrals*, Academic, Boston, MA (1989) MR

Graduate School of Information Sciences, Tohoku University
Sendai, Japan

hitoshi@tohoku.ac.jp

Received: 23 July 2023 Revised: 3 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3585–3601
DOI: 10.2140/agt.2025.25.3585
Published: 1 October 2025

The Goldman bracket characterizes homeomorphisms between noncompact surfaces

SUMANTA DAS
SIDDHARTHA GADGIL
AJAY KUMAR NAIR

We show that a homotopy equivalence between two noncompact orientable surfaces is homotopic to a homeomorphism if and only if it preserves the Goldman bracket, provided our surfaces are neither the plane nor the punctured plane.

57K20

1 Introduction

All manifolds are assumed to be second countable and Hausdorff. A surface is a two-dimensional manifold. Throughout this note, all surfaces will be assumed to be connected and orientable. We say a surface is of *finite-type* if its fundamental group is finitely generated; otherwise, we say it is of *infinite-type*.

A fundamental question in topology is whether homotopy equivalent closed n-manifolds are necessarily homeomorphic, and whether every homotopy equivalence is homotopic to a homeomorphism. More generally, we may ask what additional structures or conditions characterize homotopy equivalences that are homotopic to homeomorphisms. For $n \geq 3$, the answer to these questions is negative in general — for example, the Lens spaces $L(7,1)$ and $L(7,2)$ are homotopy equivalent but not homeomorphic.

In the case of closed surfaces, the classical *Dehn–Nielsen–Baer theorem* (see Stillwell [5]) says that every homotopy equivalence is homotopic to a homeomorphism. However, the corresponding result does not hold for compact two-manifolds with boundary. For example, the torus with one hole and the sphere with three holes are homotopy equivalent but not homeomorphic.

Similarly, homotopy equivalence does not imply homeomorphism for noncompact surfaces without boundary. For example, the once-punctured torus and the thrice-punctured sphere are homotopy equivalent but not homeomorphic. Indeed up to homotopy equivalence, there is precisely one (connected) surface of infinite type, but up to homeomorphism, there are $2^{\aleph_0}$ many infinite-type surfaces; see Das [1, Proposition 3.1.11].

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

In this note, we show that there is a simple and natural characterization of when a homotopy equivalence $f\colon \Sigma' \to \Sigma$ between noncompact surfaces without boundary is homotopic to a homeomorphism in terms of the *Goldman bracket*, which is a Lie Algebra structure associated to a surface. More precisely, if $\hat{\pi}(\Sigma)$ denotes the set of free homotopy classes of closed curves on a noncompact surface Σ, then the Goldman bracket (whose definition we recall in Section 2.6) is a bilinear map

$$[\cdot,\cdot]\colon \mathbb{Z}[\hat{\pi}(\Sigma)] \times \mathbb{Z}[\hat{\pi}(\Sigma)] \to \mathbb{Z}[\hat{\pi}(\Sigma)],$$

which is skew-symmetric and satisfies the Jacobi identity. Our main result is the following.

Theorem 1.1 *A homotopy equivalence $f\colon \Sigma' \to \Sigma$ between two noncompact oriented surfaces without boundary is homotopic to an orientation-preserving homeomorphism if and only if it commutes with the Goldman bracket, ie for all $x', y' \in \mathbb{Z}[\hat{\pi}(\Sigma')]$, we have,*

$$[f_*(x'), f_*(y')] = f_*([x', y']), \tag{1}$$

where $f_*\colon \mathbb{Z}[\hat{\pi}(\Sigma')] \to \mathbb{Z}[\hat{\pi}(\Sigma)]$ is the function induced by f, provided Σ is not homeomorphic to the plane or the cylinder $S^1 \times \mathbb{R}$.

The analogous result for compact, connected, oriented two-dimensional manifolds with boundary was proved by Gadgil [2].

Our methods are based on the relation between the Goldman bracket and the *geometric intersection number* of curves on a surface. The same methods prove a related characterization in terms of intersection numbers for homotopy equivalences.

Recall that, for a 2-manifold M with or without boundary, the geometric intersection number I_M is defined as follows. Let $x, y \in \hat{\pi}(M)$. Suppose $\alpha\colon \mathcal{S}^1 \to M$ and $\beta\colon \mathcal{S}^1 \to M$ are smooth immersions representing x and y, respectively. We say that α and β are in *general position* if they intersect transversally so that if $p \in \alpha(\mathcal{S}^1) \cap \beta(\mathcal{S}^1)$, then $\alpha^{-1}(p)$ and $\beta^{-1}(p)$ are both singletons. We sometimes simply say α and β intersect in double points.

We define the geometric intersection number as

$$I_M(x, y) := \min\{|\alpha \cap \beta| : \alpha \in x,\ \beta \in y,\ \alpha \text{ and } \beta \text{ in general position}\}.$$

Theorem 1.2 *Let $f\colon \Sigma' \to \Sigma$ be a homotopy equivalence between two noncompact surfaces without boundary, where Σ is not homeomorphic to the plane or a cylinder. Then the following are equivalent:*

(1) *f is homotopic to a homeomorphism.*

(2) *$I_\Sigma(f_*(x'), f_*(y')) = I_{\Sigma'}(x', y')$ for all $x', y' \in \hat{\pi}(\Sigma')$.*

(3) *For all $x', y' \in \hat{\pi}(\Sigma')$, $I_\Sigma(f_*(x'), f_*(y')) = 0 \iff I_{\Sigma'}(x', y') = 0$.*

1.1 Outline of the proof of Theorem 1.1

As orientation-preserving homeomorphisms preserve the Goldman bracket by definition, it is easy to see that if $f\colon \Sigma' \to \Sigma$ is homotopic to an orientation-preserving homeomorphism, then it commutes with the Goldman bracket.

Conversely, suppose $f\colon \Sigma' \to \Sigma$ commutes with the Goldman bracket, we construct a *proper* map $g\colon \Sigma' \to \Sigma$ such that f is homotopic to g. By Das [1, Theorem 1], g is in turn (properly) homotopic to a homeomorphism.

To construct g, we pick an exhaustion $K_1 \subset K_2 \subset \cdots \subset K_n \subset \cdots$ of Σ (see Section 2.2) by compact subsurfaces (in this outline, we suppress a few technical conditions on exhaustions and base-points). We construct (as we sketch below) an exhaustion $K'_1 \subset K'_2 \subset \cdots \subset K'_n \subset \cdots$ of Σ' by compact subsurfaces together with maps $g_i\colon K'_i \to \Sigma$ homotopic to the restrictions $f|_{K'_i}$ such that for all i, g_{i+1} is an extension of g_i, ie $g_{i+1}|_{K'_i} = g_i$. Furthermore, we ensure that $g_n(K'_n \setminus K'_i) \subset \Sigma \setminus K_i$ whenever $i \leq n$.

To construct K'_1 and the map $g_1\colon K'_1 \to \Sigma$, we consider a system of simple closed curves $\alpha_1, \alpha_2, \ldots, \alpha_{k_1}$ that *fill* K_1 (see Section 2.4). Thus, if γ is a closed curve that can be homotoped to be disjoint from all α_i, then γ is homotopic to a closed curve in $\Sigma \setminus K_1$. As f is a homotopy equivalence, there exist closed curves $\alpha'_1, \alpha'_2, \ldots, \alpha'_{k_1}$ such that $f(\alpha'_i)$ is homotopic to α_i for all i. We pick a compact subsurface K'_1 of Σ' that contains all the curves α'_i.

The key observation is that for any curve $\gamma' \subset \Sigma' \setminus K'_1$, $f(\gamma')$ is homotopic to a curve in $\Sigma \setminus K_1$. This follows from the properties of the Goldman bracket. Namely, the Goldman bracket of γ' with each α'_i is zero, and hence the Goldman bracket of $f(\gamma')$ with each α_i is zero. As the α_i are simple closed curves, by a theorem of Goldman, it follows that, for each i, $f(\gamma')$ is homotopic to a curve disjoint from α_i. As the α_i fill K_1, it follows that $f(\gamma')$ is homotopic to a curve in $\Sigma \setminus K_1$.

As the boundary components of K'_1 are homotopic to curves in $\Sigma' \setminus K'_1$, this lets us define a map $g_1\colon K'_1 \to \Sigma$ that is homotopic to $f|_{K'_1}$ and so that $g_1(\partial K'_1) \subset \Sigma \setminus K_1$. To proceed inductively, we prove a refinement of the above key observation. Namely, for each component V' of $\overline{\Sigma' \setminus K'_1}$, there is a component V of $\overline{\Sigma \setminus K_1}$ such that for any closed curve $\gamma' \subset V'$, $f(\gamma')$ is homotopic to a curve in V. This lets us make a construction similar to the above for each component V' of $\overline{\Sigma' \setminus K'_1}$ to obtain a subsurface in V and an extension of the map to this subsurface. Taking the union of these subsurfaces, we get K'_2 and an extension g_2 of g_1 to K'_2 so that $g_2(\partial K'_2) \subset \Sigma \setminus K_2$. We proceed inductively to obtain K'_n and maps $g_n\colon K'_n \to \Sigma$ with $g_n(\partial K'_n) \subset \Sigma \setminus K_n$ for all n.

Finally, the limit of the maps g_i gives us a proper map $g\colon \Sigma' \to \Sigma$ homotopic to f as each $g_i\colon K'_i \to \Sigma$ homotopic to the restriction $f|_{K'_i}$.

Acknowledgements

Das and Nair are supported by fellowships from the National Board for Higher Mathematics. We are grateful to the referee for a careful reading and comments and suggestions which helped considerably in improving the manuscript.

2 Background

A *surface* Σ is a connected, orientable two-dimensional manifold with or without boundary.

2.1 Subsurfaces

A (compact) *subsurface* F of a surface Σ is an embedded submanifold (in general with boundary) of codimension zero. Thus, $F \subset \Sigma$ is a subset homeomorphic to a compact surface with boundary so that $\partial F \subset \Sigma$ is a collection of disjoint simple closed curves. Further, if F is connected, closure of a component of $\Sigma \setminus \partial F$ is F.

We assume that all subsurfaces are connected.

2.2 Exhaustions

An *exhaustion* of a noncompact surface Σ is a collection of subsurfaces K_i, $i = 1, 2, \ldots$ such that

- $K_i \subset \text{int}(K_{i+1})$ for all i,
- $\bigcup_i K_i = \Sigma$.

2.3 Intersections and geodesics

A very useful classical result on surfaces relates pairwise and mutual intersections. This can be conveniently formulated in terms of hyperbolic metrics on surfaces (one can define this purely topologically, but the definition is a bit more complicated).

Recall that except sphere and torus, every compact surface Σ without boundary admits a complete hyperbolic metric. Similarly, every compact surface Σ with boundary admits a hyperbolic metric with geodesic boundary if it has negative Euler characteristic. Assume that such a metric has been fixed. Then the following holds.

Lemma 2.1 *Let Σ be a surface with a hyperbolic metric.*

(1) *Every nontrivial homotopy class of curves in Σ contains a unique geodesic representative.*

(2) *Every homotopically nontrivial simple closed curve α in Σ is ambient isotopic to a simple geodesic.*

(3) *If two homotopy classes of curves in Σ have representatives that are disjoint, then their geodesic representatives are disjoint.*

It follows that for a finite collection of homotopy classes of curves $x_1, x_2, \ldots, x_n$ in Σ, if every pair x_i and x_j have disjoint representatives, then there exists a family of mutually disjoint representatives of $x_1, x_2, \ldots, x_n$.

In particular, if $F \subset \Sigma$ is a subsurface, then the geodesics homotopic to the components of ∂F are pairwise disjoint, and the closure of a component of $\Sigma \setminus \partial F$ is a hyperbolic surface with geodesic boundary which is isotopic to F. We can thus assume that subsurfaces have geodesic boundary.

2.4 Filling curves

Let F be a compact surface, possibly with boundary. A homotopically nontrivial closed curve $\alpha \subset F$ is said to be *peripheral* if α is (freely) homotopic to a curve with the image in ∂F.

Fix a hyperbolic metric on F with geodesic boundary. We say that a collection of geodesics $\alpha_1, \alpha_2, \dots, \alpha_n$ in F *fills* F if every component of $F \setminus \bigcup_i(\alpha_i)$ is either an open disc or a punctured disc with boundary contained in ∂F. By the results recalled in Lemma 2.1, if a homotopically nontrivial curve γ is homotopic to curves disjoint from each α_i, $1 \le i \le n$, then γ is peripheral.

Similarly, if $F \subset \Sigma$ is a compact subsurface with geodesic boundary and the geodesics $\alpha_1, \alpha_2, \dots, \alpha_n$ in F fill F, then every homotopically nontrivial curve γ in Σ that is disjoint from each α_i is homotopic to a curve in $\Sigma \setminus \operatorname{int}(F)$, and hence to a curve in $\Sigma \setminus F$.

2.5 Splitting surfaces

Let Σ be a surface, and γ be a simple closed curve on Σ so that $\Sigma \setminus \gamma$ has two components. Observe that a curve α in a component V of $\Sigma \setminus \gamma$ is peripheral in V if it is homotopic in Σ to a curve in a regular neighbourhood of γ.

Using the results recalled in Lemma 2.1, we deduce the following.

Lemma 2.2 *Let $\alpha \subset \Sigma$ be a curve that is disjoint from γ and is not peripheral. If $\beta \subset \Sigma \setminus \gamma$ is a closed curve that is homotopic to α in Σ, then α and β are contained in the same component V of $\Sigma \setminus \gamma$ and are homotopic in V.*

2.6 Goldman bracket

Let Σ be a noncompact surface, and let $\widehat{\pi}(\Sigma) = [\mathbb{S}^1, \Sigma]$, ie $\widehat{\pi}(\Sigma)$ denotes the set of free homotopy classes of curves in Σ. Note that if $p \in \Sigma$, there is a bijection between the set of all conjugacy classes of $\pi_1(\Sigma, p)$ and $\widehat{\pi}(\Sigma)$. For a closed curve α, let $\widehat{\alpha}$ denote the free homotopy class of α.

Let $x, y \in \widehat{\pi}(\Sigma)$ and α, β be smoothly immersed, oriented representatives of x and y, respectively, such that α and β intersect transversally in (finitely many) double points. Given any point $p \in \alpha \cap \beta$, we view α and β as curves in Σ based at p and define $\alpha *_p \beta$ as the product of based curves as in the definition of the fundamental group.

We associate to each intersection point p a *sign* $\varepsilon_p = \pm 1$ as in the definition of algebraic intersection number. Namely, parametrizing α and β so that $\alpha(0) = \beta(0) = p$, if the ordered pair $(\alpha'(0), \beta'(0))$ represents the orientation of Σ at p, then $\varepsilon_p = 1$, otherwise $\varepsilon_p = -1$.

The Goldman bracket [3] $[\cdot, \cdot]\colon \mathbb{Z}[\widehat{\pi}(\Sigma)] \times \mathbb{Z}[\widehat{\pi}(\Sigma)] \to \mathbb{Z}[\widehat{\pi}(\Sigma)]$ is defined as

$$[x, y] = \sum_{p \in \alpha \cap \beta} \varepsilon_p \cdot \widehat{\alpha *_p \beta}. \tag{2}$$

Recall that for a nonempty set S, $\mathbb{Z}[S]$ is the free module generated by S over the ring of integers $\mathbb{Z}$. Also, note that the sum in equation (2) is a finite sum.

Theorem 2.3 (Goldman [3]) *The Goldman bracket satisfies the following properties*:

(1) $[\cdot,\cdot]\colon \mathbb{Z}[\widehat{\pi}(\Sigma)]\times\mathbb{Z}[\widehat{\pi}(\Sigma)]\to\mathbb{Z}[\widehat{\pi}(\Sigma)]$ *is a well-defined bilinear map*; *see* [3, Theorem 5.2.].

(2) *The bracket is skew-symmetric and satisfies the Jacobi identity*, *ie*
   (a) $[x,y]=-[y,x]$ *for all* $x,y\in\widehat{\pi}(\Sigma)$, *and*
   (b) $[x,[y,z]]+[y,[z,x]]+[z,[x,y]]=0$ *for all* $x,y,z\in\widehat{\pi}(\Sigma)$; *see* [3, Theorem 5.3.].

(3) *If* α *is a simple closed curve and* $x=\widehat{\alpha}$ *is the homotopy class of* α, *then*, *for* $y\in\widehat{\pi}(\Sigma)$, $[x,y]=0$ *if and only if* $y=\widehat{\beta}$ *for a closed curve* β *such that* $\beta\cap\alpha=\varnothing$; *see* [3, Theorem 5.17(i)].

We will frequently abuse notation and speak of the Goldman bracket of two curves α and β in a surface Σ to mean the Goldman bracket of their free homotopy classes $\widehat{\alpha}$ and $\widehat{\beta}$ in $\widehat{\pi}(\Sigma)$.

3 Proof of Theorem 1.1

Suppose f is homotopic to an orientation-preserving homeomorphism $g\colon\Sigma'\to\Sigma$. Without loss of generality, we may assume g is an orientation-preserving diffeomorphism since in dimension two, any homeomorphism is isotopic, thus properly homotopic to a diffeomorphism. By the definition of the Goldman bracket, it follows that g commutes with the Goldman bracket. As f and g are homotopic, it follows that f also commutes with the Goldman bracket. This proves one implication of Theorem 1.1.

Conversely, suppose that $f\colon\Sigma'\to\Sigma$ is a homotopy equivalence between noncompact surfaces such that f commutes with the Goldman bracket. Pick base-points $p'\in\Sigma'$ and $p\in\Sigma$ so that $f(p')=p$. As sketched in Section 1.1, we will construct a proper map g that is homotopy equivalent to f. This is an inductive construction, where we construct surfaces K'_m and maps $g_m\colon K'_m\to\Sigma$. To keep track of base-points, we will also construct inductively a family of trees $T'_m\subset K'_m$ starting with $T'_0=p'$ so that $T'_m\cap K'_{m-1}=T'_{m-1}$ for $m>1$ and the terminal vertices of T'_m lie on $\partial K'_m$ with one terminal vertex in each boundary component. We will also modify f by a homotopy in a neighbourhood of the trees T'_m (with f fixed throughout the homotopy on p', and on the set K_{m-1} for $m>1$).

Pick an exhaustion $K_1\subset K_2\subset\cdots\subset K_n\subset\cdots$ of Σ by compact subsurfaces with boundary such that $p\in K_1$ and K_1 (and hence every K_m) is not a disc or an annulus — this is possible as Σ is not a plane or a cylinder. Also choose an exhaustion $L'_1\subset L'_2\subset\cdots\subset L'_n\subset\cdots$ of Σ' by compact bordered subsurfaces such that $p'\in L'_1$

We first construct a compact subsurface $K'_1\subset\Sigma'$, a map $g_1\colon K'_1\to\Sigma$, and a tree $T'_1\subset K'_1$, which satisfy various properties. These properties will allow us to continue the construction inductively and to ensure that the resulting map is proper.

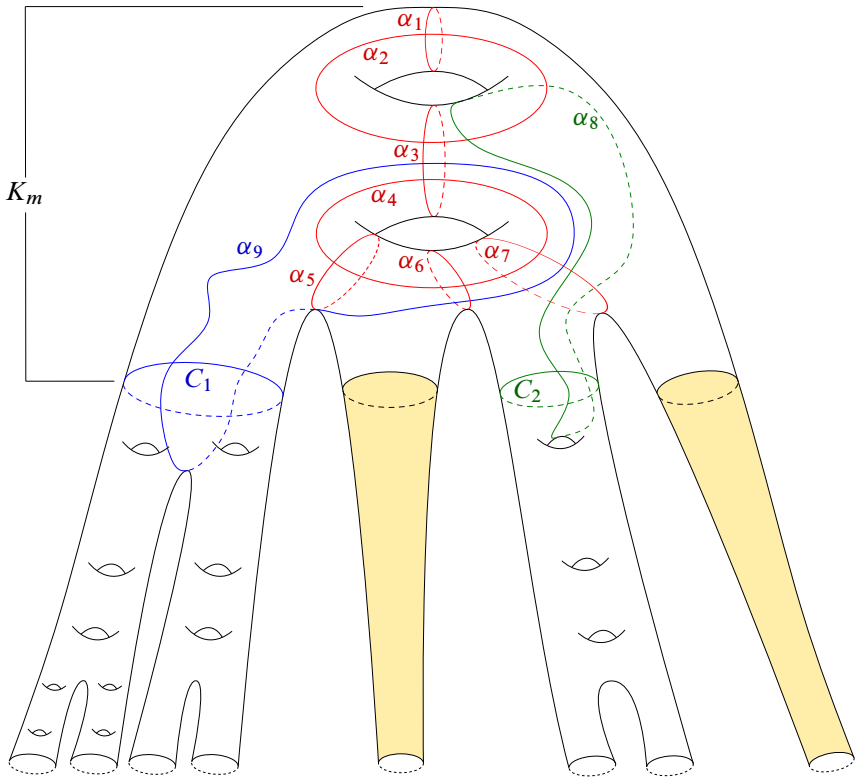


Figure 1: $\{\alpha_1, \dots, \alpha_7\}$ is a filling system of K_m and $\{\alpha_1, \dots, \alpha_9\}$ is an extended filling system for K_m because $I_\Sigma(C_1, \alpha_9) \neq 0 \neq I_\Sigma(C_2, \alpha_8)$.

3.1 Constructing the subsurface

The construction of the subsurface is almost identical in the base case and the general case, so we describe the general case directly. Fix $m \geq 1$. If $m > 1$, assume that a subsurface K'_{m-1} has been constructed. Let $Q'_m := L'_1$ if $m = 1$ and $Q'_m := L'_m \cup K'_{m-1}$ if $m > 1$.

Consider a system of simple closed curves $\alpha_1, \alpha_2, \dots, \alpha_l$ that fill K_m (as defined in Section 2.4). Thus, if γ is a closed curve that can be homotoped to be disjoint from each of the curves α_i, then γ is homotopic to a closed curve in $\Sigma \setminus K_m$. We enlarge this collection to ensure that a peripheral curve must intersect a curve in the system. Namely, if $C \subset \partial K_m$ is a boundary component of K_m which is not the boundary of a component in $\overline{\Sigma \setminus K_m}$ homeomorphic to the half-open cylinder $[0, \infty) \times \mathbb{R}$ or the closed unit disc, then there exists a simple closed curve $\alpha^{(C)}$ so that C and $\alpha^{(C)}$ cannot be homotoped to be disjoint. Add such a curve for each boundary component of K_m that is not the boundary of a cylinder or disc in Σ. Let $\alpha_1, \alpha_2, \dots, \alpha_l, \alpha_{l+1}, \dots, \alpha_k$ be the resulting collection of curves; see Figure 1.

As f is a homotopy equivalence, there exist closed curves $\alpha'_i \subset \Sigma'$, $1 \leq i \leq k$, such that $f(\alpha'_i)$ is homotopic to α_i. Further, we can assume that the curves α'_i are pairwise transversal and intersect only in double points.

Lemma 3.1 *There exists a compact subsurface $K'_m \subset \Sigma'$ that satisfies the following*:

(1) *For all i, $1 \le i \le k$, $\alpha'_i \subset K'_m$.*

(2) *$p' \in K'_m$.*

(3) *$Q'_m \subset K'_m$.*

(4) *The closure of every component of $\Sigma' \setminus K'_m$ is noncompact.*

(5) *Every component of $\overline{\Sigma' \setminus K'_m}$ intersects K'_m in a single component of $\partial K'_m$.*

Proof To ensure the first three conditions, we simply take a compact connected subsurface K'_m that contains $\left(\bigcup_i \alpha'_i\right) \cup Q'_{m-1} \cup \{p'\}$.

Next, since K'_m is compact, the closure $\overline{\Sigma' \setminus K'_m}$ of its complement has finitely many components $V'_1, V'_2, \dots, V'_n$. Suppose the compact complementary components are $V'_{i_1}, V'_{i_2}, \dots V'_{i_l}$. Then we replace K'_m by $K'_m \cup \left(\bigcup_{j=1}^{n} V'_{i_j}\right)$. This is also compact and satisfies condition (4).

Finally, let $V'_1, V'_2, \dots, V'_l$ be the (noncompact) components of $\overline{\Sigma' \setminus K'_m}$. Suppose V'_i intersects K'_m in $n_i > 1$ components of $\partial K'_m$, say $C'_1, C'_2, \dots C'_{n_i}$. For $j = 2, 3, \dots, n_i$, pick disjoint simple arcs β'_j in V'_j connecting C'_1 to C'_j. Replace K'_m by $K'_m \cup \mathcal{N}\left(\bigcup_{j=2}^{n_i} \beta'_j\right)$, where $\mathcal{N}$ denotes a regular neighbourhood in V'_i. Making a similar construction for each V'_i that intersects K'_m in more than one boundary component, we obtain a compact subsurface K'_m that satisfies the condition (5) (and continues to satisfy the other conditions). □

Remark 3.2 The proof of Lemma 3.1 shows that any noncompact surface Σ without boundary has an exhaustion $\{K_i\}$ by compact subsurfaces such that for each $i = 1, 2, \dots$, it holds that

- the closure of every component of $\Sigma \setminus K_i$ is noncompact, and
- every component of $\overline{\Sigma \setminus K_i}$ intersects K_i in a single component of ∂K_i.

3.2 Pushing loop images outside subsurfaces

Choose and fix a compact subsurface K'_m satisfying the conditions of Lemma 3.1. Using the assumption that f commutes with the Goldman bracket gives the following key lemma.

Lemma 3.3 *Let γ' be a closed curve in Σ' that is homotopic to a closed curve in $\Sigma' \setminus K'_m$. Then $f(\gamma')$ is homotopic to a closed curve in $\Sigma \setminus K_m$.*

Proof Since γ' is homotopic to a closed curve in $\Sigma' \setminus K'_m$, γ' is homotopic to a curve that is disjoint from α'_i for all i, $1 \le i \le k$. It follows that the Goldman bracket $[\gamma', \alpha'_i]$ is zero for all i, $1 \le i \le k$.

As f commutes with the Goldman bracket, $[f(\gamma'), \alpha_i] = 0$ for all i. For each i, as α_i is a simple closed curve, it follows that $f(\gamma')$ is homotopic to a curve disjoint from α_i by statement (3) of Theorem 2.3. As a subset of the curves α_i fill K_m, it follows that $f(\gamma')$ is homotopic to a curve in $\Sigma \setminus K_m$. □

3.3 Construction of the map in the base case

We next construct a map $g_1\colon K_1' \to \Sigma$ that is homotopic to $f|_{K_1'}$ and satisfies $g_1(p') = p$ so that $g_1(\partial K_1') \cap K_1 = \varnothing$. As K_1' is a compact surface with a nonempty boundary, there exists a graph $\Gamma' \subset K_1'$ such that K_1' deformation retracts onto Γ'. We can choose Γ' so that $p' \in \Gamma'$ and $\Gamma' \cap \partial K_1' = \varnothing$. A regular neighbourhood $\mathcal{N}(\Gamma')$ of Γ' has a deformation retraction $\rho\colon \mathcal{N}(\Gamma') \to \Gamma'$. We get a map $g_1\colon \mathcal{N}(\Gamma') \to \Sigma$ by setting $g_1|_{\mathcal{N}(\Gamma')} = f \circ \rho$.

Further, $\overline{K_1' \setminus \mathcal{N}(\Gamma')}$ is a disjoint union of annuli $C_i' \times [0,1]$, $1 \le i \le n$, so that the boundary components of K_1' are the curves $C_i' \times \{1\}$, $1 \le i \le n$ and the boundary components of $\mathcal{N}(\Gamma')$ are $C_i' \times \{0\}$, $1 \le i \le n$. By Lemma 3.3, for each i, $1 \le i \le n$, there is a curve $\gamma_i \subset \Sigma \setminus K_1$ so that $f(C_i')$ is homotopic to γ_i. We define g_1 on each annulus $C_i' \times [0,1]$ as follows. We set $g_1(C_i' \times \{1\}) = \gamma_i$. Note that $C_i' \times \{0\}$ is identified with a boundary component of $\mathcal{N}(\Gamma')$, and so g_1 has been defined on $C_i' \times \{0\}$. Finally, by construction $g_1|_{C_i' \times \{0\}}$ and $g_1|_{C_i' \times \{1\}}$ are both homotopic in Σ to $f(C_i')$ and hence homotopic to each other — so we can extend g_1 to $C_i' \times [0,1]$ using a homotopy in Σ.

We next construct the tree T_1' and modify the map f. Let the boundary components of K_1' be $\delta_i' = C_i' \times \{1\}$, $1 \le i \le n$. Pick a point $q_i' \in \delta_i'$ for each i, $1 \le i \le n$. Pick a family of arcs θ_i' with interiors disjoint so that θ_i' is an arc from p' to q_i' in K_1'. Let T_1' be the union of these arcs. As T_1' is contractible and $f(p') = g_1(p')$, we can homotope f fixing p' so that $f|T_1' = g_1|T_1'$.

Note that for $1 \le i \le n$, $f(q_i') = g_1(q_i')$. Further, for $1 \le i \le n$, since

$$(f|_{K_1'})_* = (g_1)_*\colon \pi_1(K_1', p') \to \pi_1(\Sigma, f(p')), \quad \text{and} \quad f|_{\theta_i'} = g_1|_{\theta_i'},$$

we deduce by using the *change of base-point isomorphisms* corresponding to θ_i' and the fact that $f \circ \theta_i' = g_1 \circ \theta_i'$ that $(f|_{K_1'})_* = (g_1)_*\colon \pi_1(K_1', q') \to \pi_1(\Sigma, f(q'))$.

3.4 Inductive properties

We now consider the inductive step, where $m > 1$, and we have already defined a subsurface $K_{m-1}' \subset \Sigma'$, a map $g_{m-1}\colon K_{m-1}' \to \Sigma$, and a tree $T_{m-1}' \subset K_{m-1}'$. Further, we assume that these satisfy the following properties:

(1) The subsurface K_{m-1}' satisfies the conditions of Lemma 3.1.

(2) $g_{m-1}(\partial K_{m-1}') \subset \Sigma \setminus K_{m-1}$.

(3) $f|_{T_{m-1}'} = g_{m-1}|_{T_{m-1}'}$.

(4) The terminal vertices of T_{m-1}' are in $\partial K_{m-1}'$, with one terminal vertex in each component.

(5) For a terminal vertex $q' \in T_{m-1}'$, $(f|_{K_{m-1}'})_* = (g_{m-1})_*\colon \pi_1(K_{m-1}', q') \to \pi_1(\Sigma, q)$.

Observe that we have shown in Section 3.3 that the above properties hold for $m = 2$. We will construct K_m', g_m, and T_m' and show (in Lemma 3.6) that the above properties hold for these.

3.5 Maps on fundamental groups of components

Let K'_m be the subsurface constructed using Lemma 3.1. Let $V'^{(1)}, V'^{(2)}, \dots, V'^{(k)}$ be the components of $\overline{K'_m \setminus K'_{m-1}}$. By Lemma 3.1, each component $V'^{(j)}$ intersects K'_{m-1} in a single boundary component $\delta'^{(j)}$.

We construct extensions $g_m^{(j)}\colon K'_{m-1} \cup V'^{(j)} \to \overline{\Sigma \setminus K_{m-1}}$ of g_{m-1} for each component $V'^{(j)}$, $1 \le j \le k$ in Section 3.6. We first show that we have homomorphisms between fundamental groups of components of $\overline{K'_m \setminus K'_{m-1}}$ and components of $\overline{\Sigma \setminus K_{m-1}}$.

Fix a component $V' = V'^{(j)}$ of $\overline{K'_m \setminus K'_{m-1}}$ and let $\delta' = \delta'^{(j)}$. Let $i'\colon V' \to \Sigma'$ be the inclusion map. By construction $g_{m-1}(\delta') \subset \Sigma \setminus K_{m-1}$. Let V be the component of $\Sigma \setminus K_{m-1}$ containing $g_{m-1}(\delta')$ and $i\colon V \to \Sigma$ be the inclusion map. We construct an extension with image contained in V. Most of the additional work beyond Lemma 3.3 is to control images of curves up to *based* homotopy, not just free homotopy.

Let $q' \in V'$ be the terminal vertex of T'_{m-1} corresponding to V', and let $q = f(q')$. We have

$$(f|_{K'_{m-1}})_* = (g_{m-1})_*\colon \pi_1(K'_{m-1}, q') \to \pi_1(\Sigma, q).$$

Observe that δ' can be regarded as a closed curve based at q'. As the image of δ' is contained in V, $(f|_{K'_{m-1}})_*([\delta']) = (g_{m-1})_*([\delta']) \in i_*(\pi_1(V, q))$. The key idea is to apply Lemma 3.3 to both the curves γ' and $\gamma' * \delta'$ for a curve $\gamma' \subset V'$ based at q'. We first need a lemma about amalgamated free products.

Let A and B be groups with $C \subseteq A$ and $C \subseteq B$ as subgroups of both of them. Let $G := A *_C B$ be the *amalgamated free product* of A and B along C (see [4]). The groups A and B are called the *factors* of G. Following terminology from topology, we say that an element in A or B is *peripheral* if it is conjugate to an element in C.

Lemma 3.4 *Let $a \in A$ be an element such that a is not conjugate to an element in C. Let $g \in G$.*

(1) *Suppose g is conjugate to a nonperipheral element $b \in B$. Then ag is not conjugate to an element in a factor of G.*

(2) *If g is conjugate to a nonperipheral element a' in A and ag is conjugate to an element in a factor, then $g \in A$.*

Proof By [4, Theorem 4.6], if ag is conjugate to an element in a factor, then any cyclically reduced word representing an element conjugate to ag is contained in a factor.

First, suppose g is conjugate to a nonperipheral element $b \in B$, say $g = hbh^{-1}$. We consider a reduced word $h = l_1 \cdot l_2 \cdot l_3 \cdot \ldots \cdot l_k$ representing h (we will separate *letters* in words in the free products by $\cdot$). There are a few different cases corresponding to in which factor the first and last letters of h belong.

We first consider the most nontrivial case, where $h = a_1 \cdot b_1 \cdot \ldots \cdot a_k \cdot b_k$, so

$$g = a_1 \cdot b_1 \cdot \ldots \cdot a_k \cdot b_k \cdot b \cdot b_k^{-1} \cdot a_k^{-1} \cdot \ldots \cdot b_1^{-1} \cdot a_1^{-1}.$$

We claim that a cyclically reduced word conjugate to ag is $(a_1^{-1}aa_1)\cdot b_1\cdot\ldots\cdot a_k\cdot(b_kbb_k^{-1})\cdot a_k^{-1}\cdot\ldots\cdot b_1^{-1}$. This clearly represents the element $a_1^{-1}aga_1$, which is conjugate to ag. As a and b are not peripheral, $a_1^{-1}aa_1\notin C$ and $b_kbb_k^{-1}\notin C$; hence the word is indeed cyclically reduced. As this word is not contained in a factor, ag is not conjugate to an element in a factor.

The other three cases are analogous. In each case we obtain a cyclically reduced word that is not in a factor. For completeness, we list the cases and the cyclically reduced words we obtain:

(1) $h=a_1\cdot b_1\cdot\ldots\cdot a_k$, $k\geq 1$, we obtain $(a_1^{-1}aa_1)\cdot b_1\cdot\ldots\cdot a_k\cdot b\cdot a_k^{-1}\cdot\ldots\cdot b_1^{-1}$.

(2) $h=b_1\cdot a_2\cdot\ldots\cdot a_k\cdot b_k$, we obtain $a\cdot b_1\cdot\ldots\cdot a_k\cdot(b_kbb_k^{-1})\cdot a_k^{-1}\cdot\ldots\cdot b_1^{-1}$.

(3) $h=b_1\cdot a_2\cdot\ldots\cdot a_k$, we obtain $a\cdot b_1\cdot\ldots\cdot a_k\cdot b\cdot a_k^{-1}\cdot\ldots\cdot b_1^{-1}$.

Next, suppose g is conjugate to a nonperipheral element $a'\in A$, say $g=ha'h^{-1}$, and ag is conjugate to an element in a factor. We again consider a reduced word representing h. If this is a single letter in A, ie $h=a_1\in A$, then $g=a_1a'a_1^{-1}\in A$ as claimed. In all other cases, we see above that ag is represented by a cyclically reduced word that is not in a factor, a contradiction. Again, we list the (four) cases and the cyclically reduced words we obtain:

(1) $h=a_1\cdot b_1\cdot\ldots\cdot a_k\cdot b_k$, $k\geq 1$, we obtain $(a_1^{-1}aa_1)\cdot b_1\cdot\ldots\cdot a_k\cdot b_k\cdot a'\cdot b_k^{-1}\cdot\ldots\cdot b_1^{-1}$.

(2) $h=b_1\cdot a_2\cdot\ldots\cdot a_k\cdot b_k$, we obtain $a\cdot b_1\cdot\ldots\cdot a_k\cdot b_k\cdot a'\cdot b_k^{-1}\cdot\ldots\cdot b_1^{-1}$.

(3) $h=b_1\cdot a_2\cdot\ldots\cdot b_{k-1}\cdot a_k$, we obtain $a\cdot b_1\cdot\ldots\cdot a_k\cdot b_{k-1}\cdot(a_ka'a_k^{-1})\cdot b_{k-1}^{-1}\cdot\ldots\cdot b_1^{-1}$.

(4) $h=a_1\cdot b_1\cdot a_2\cdot\ldots\cdot b_{k-1}\cdot a_k$, we obtain $(a_1^{-1}aa_1)\cdot b_1\cdot\ldots\cdot(a_ka'a_k^{-1})\cdot\ldots\cdot b_1^{-1}$, which is not in a factor if $k>1$, ie except for the case $k=1$, $h=a_1$ (mentioned above), where $g=a_1a'a_1^{-1}\in A$. □

We will apply the above with C, the fundamental group of a simple closed curve δ in a surface F that separates F, with A and B the fundamental groups of the closures of the components of $F\setminus\delta$. We use Lemma 2.2 in this context.

Recall that V' is a component of $\overline{K'_m\setminus K'_{m-1}}$ with $\delta'=V'\cap K'_{m-1}$ and V is a component of $\overline{\Sigma\setminus K_{m-1}}$ such that $f(\delta')$ is freely homotopic to a curve in V. Further, q' is a point in δ' and $q=f(q')$. The following lemma says that f sends each loop in V' based at q' to a loop that is homotopic fixing basepoint to a loop in V based at q, up to base-point fixing homotopy.

Lemma 3.5 $$f_*\big(i'_*(\pi_1(V',q'))\big)\subset i_*(\pi_1(V,q)).$$

Proof Let γ' be a curve in V' based at q', so $[\gamma']\in i'_*(\pi_1(V',q'))$. We show that $f_*([\gamma'])\in i_*(\pi_1(V,q))$. If γ' is *peripheral*, ie homotopic to a power of the boundary component δ', then either $\gamma'=\delta'^k$ for some $k\in\mathbb{Z}$ or $\gamma'=g'*\delta'^k*g'^{-1}$ for some $g'\in\pi_1(V',q')$ which is not peripheral. In the first case, $f_*([\gamma'])\in i_*(\pi_1(V,q))$ as $f_*([\delta'])=(g_{m-1})_*([\delta'])\in i_*(\pi_1(V,q))$. In the second case, it suffices to show that the nonperipheral curve g is mapped to a curve in $i_*(\pi_1(V,q))$. Hence, we can assume that γ' is not peripheral.

By Lemma 3.3, there is a curve $\gamma \subset \Sigma \setminus K_{m-1}$ so that $f(\gamma')$ is homotopic to γ. Thus, $\gamma \subset \widehat{V}$ for some component $\widehat{V}$ of $\overline{\Sigma \setminus K_{m-1}}$. Let $\widehat{\delta} = \widehat{V} \cap K_{m-1}$, and let $\widehat{q} \in \widehat{\delta}$. We modify γ by a homotopy so that γ intersects $\widehat{\delta}$ in the single point $\widehat{q}$.

We claim that γ is not peripheral, ie γ is not homotopic to a curve with image in $\partial \widehat{V} = \widehat{\delta}$. Namely, as γ' is a curve in V' that is not peripheral, there is a simple closed curve β' in V' so that γ' and β' are not homotopic to disjoint curves. Hence, by statement (3) of Theorem 2.3, we have $[\beta', \gamma'] \neq 0$. The curve $f(\beta')$ is homotopic to a curve β in $\Sigma \setminus K_{m-1}$. If γ were peripheral, then β and γ would be homotopic to disjoint curves, hence $[\beta, \gamma]$ would become 0. But $[\beta', \gamma'] \neq 0$ implies $[\beta, \gamma] \neq 0$ as f commutes with the Goldman bracket, this is a contradiction. Hence, γ is not peripheral.

We next see that $\widehat{V} = V$. Suppose not, then $\gamma \subset \widehat{V}$ for a component $\widehat{V} \neq V$ of $\overline{\Sigma \setminus K_{m-1}}$. As γ is homotopic to $f(\gamma')$, there exists a path θ in Σ from q to $\widehat{q}$ such that $f(\gamma')$ is homotopic in Σ fixing base-point to $\theta * \gamma * \bar{\theta}$. Observe that $\pi_1(\Sigma, \widehat{q})$ is the amalgamated free product

$$\pi_1(\Sigma, \widehat{q}) = \pi_1(\widehat{V}, \widehat{q}) *_{\pi_1(\widehat{\delta}, \widehat{q})} \pi_1(\overline{\Sigma \setminus \widehat{V}}, \widehat{q}). \tag{3}$$

We identify $\pi_1(\Sigma, q)$ with $\pi_1(\Sigma, \widehat{q})$ using the change of base-point isomorphism determined by θ, ie the isomorphism $[\lambda] \mapsto [\bar{\theta} * \lambda * \theta]$ for a loop λ based at q. Under this identification, $f_*([\gamma']) \equiv [\gamma]$. Further, as $f(\delta')$ is homotopic to a curve η in $V \subset \Sigma \setminus \widehat{V}$, $f_*([\delta'])$ is conjugate to an element of $\pi_1(\Sigma \setminus \widehat{V}, \widehat{q})$, which is nonperipheral in $\Sigma \setminus \widehat{V}$ as K_{m-1} is not an annulus or disc by assumption. Hence, with respect to the amalgamated free product given by equation (3), $f_*([\gamma' * \delta']) = f_*([\gamma']) \cdot f_*([\delta'])$ is of the form ag of Lemma 3.4, statement (1). It follows that $f(\gamma' * \delta')$ is not homotopic to a curve that is disjoint from $\widehat{\delta}$ using Lemma 2.2 (see Figure 2). This contradicts Lemma 3.3 as $\widehat{\delta} \subset \partial K_{m-1}$ and $f(\gamma' * \delta')$ is homotopic to a curve in $\Sigma \setminus K_{m-1}$. Thus, $\widehat{V} = V$.

Next, we can express the fundamental group $\pi_1(\Sigma, q)$ as an amalgamated free product

$$\pi_1(\Sigma, q) = \pi_1(V, q) *_{\pi_1(\delta, q)} \pi_1(\overline{\Sigma \setminus V}, q). \tag{4}$$

Next, as $\gamma \subset V$, modifying by a homotopy, we can assume that γ is a loop based at q. As γ is freely homotopic to $f(\gamma')$, $[\gamma] \in \pi_1(\Sigma, q)$ is conjugate to the element $f_*([\gamma']) \in \pi_1(\Sigma, q)$, which is not peripheral as γ is not peripheral. We claim that $f_*([\delta'])$ is also not peripheral. Namely, as V contains a curve that is not peripheral, namely γ, the boundary component δ of V which is contained in K_m does not bound a cylinder or a disc. Hence, one of the curves α_i in Lemma 3.1 cannot be homotoped to be disjoint from δ. But δ' is disjoint from α_i', so $[\delta', \alpha_i'] \neq 0$, but $[f(\delta'), \alpha_i] = 0$. As in the case of γ', we conclude that $f_*([\delta'])$ is not peripheral (as it is not conjugate to a power of δ).

For the amalgamated free product decomposition (4), we apply statement (2) of Lemma 3.4 with $a = f_*([\delta'])$ and $g = f_*([\gamma'])$. As $f(\delta' * \gamma')$ is conjugate to an element disjoint from K_{m-1}, hence from δ, $f_*([\delta' * \gamma'])$ is conjugate to an element in $\pi_1(V, q)$ or $\pi_1(\overline{\Sigma \setminus V}, q)$, ie an element of a factor. It follows by statement (2) of Lemma 3.4 that $f_*([\gamma']) \in \pi_1(V, q)$, as desired. □

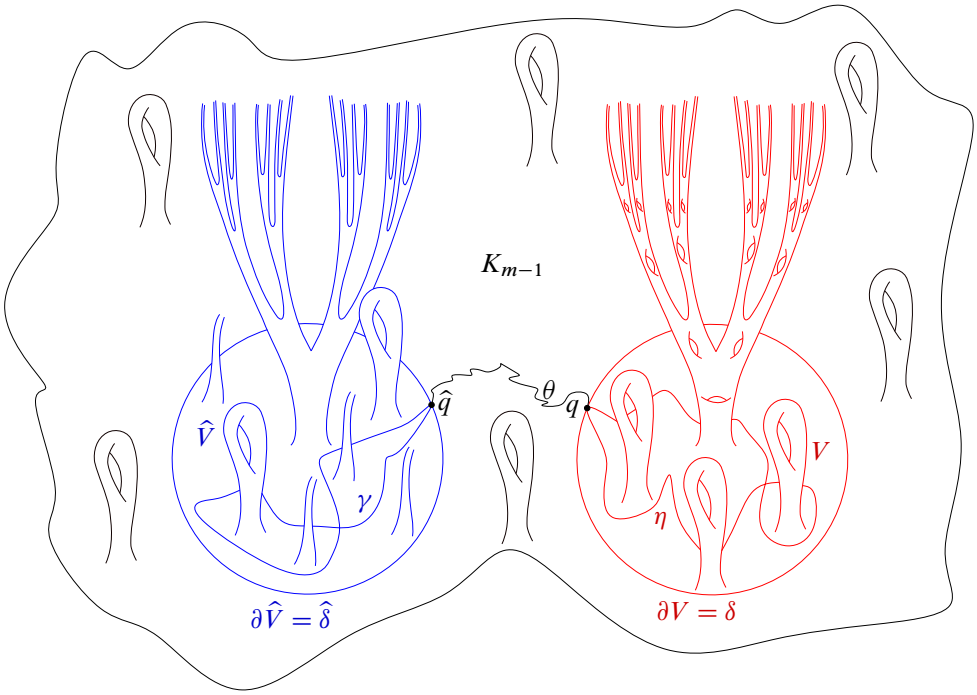


Figure 2: If $\hat{V} \neq V$, $f_*([\gamma']) \cdot f_*([\delta]) = [\gamma * \theta * \eta * \bar{\theta}]$ is not homotopic to a curve disjoint from $\partial\hat{V}$, giving a contradiction.

3.6 Maps on complementary components

The rest of the construction on each component is analogous to the base case with some refinements.

We continue with the notation of Lemma 3.5. On the component $V' = V'^{(j)}$ of $\overline{K_m \setminus K_{m-1}}$, we construct a map $g_m^{(j)}\colon V' \to V$ extending $g_{m-1}|_{\delta'}$ so that $i_* \circ (g_m^{(j)})_* = (f|_{V'})_*\colon \pi_1(V', q') \to \pi_1(\Sigma, q)$ and $g_m^{(j)}(\partial K'_m \cap V') \subset V \subset \Sigma \setminus K_m$.

Namely, we extend the graph in $V' = V'^{(j)}$ consisting of a single vertex q' and the edge δ' to a graph Γ' so that V' deformation retracts to Γ' (this exists as V' has more than one boundary component as a consequence of Lemma 3.1, statement (4)). By Lemma 3.5, we can extend $g_{m-1}|_{\delta'}$ to a map $\varphi\colon \Gamma' \to V$ so that $\varphi_* = (f|_{\Gamma'})_*\colon \pi_1(\Gamma', q') \to \pi_1(\Sigma, q)$ (as any homomorphism from the fundamental group of a graph to that of a topological space is induced by a map). A regular neighbourhood $\mathcal{N}(\Gamma')$ of Γ' has a deformation retraction $\rho\colon \mathcal{N}(\Gamma') \to \Gamma'$. We get a map $g_m^{(j)}\colon \mathcal{N}(\Gamma') \to \Sigma$ by setting $g_m^{(j)}|_{\mathcal{N}(\Gamma')} = \varphi \circ \rho$ (see Figure 3).

Next, $\overline{V' \setminus \mathcal{N}(\Gamma')}$ is a disjoint union of annuli $C'_i \times [0,1]$, $1 \le i \le k$, so that the boundary components of K'_m other than δ' are the curves $C'_i \times \{1\}$, $1 \le i \le k$ and the boundary components of $\mathcal{N}(\Gamma')$ are $C'_i \times \{0\}$, $1 \le i \le k$. By Lemma 3.1, $f(C'_i \times \{1\})$ is homotopic to a curve $\gamma \subset \Sigma \setminus K_m$ for each i, $1 \le i \le k$. We define $g_m^{(j)}$ on $C'_i \times \{1\}$ to map to γ. As $f(C'_i \times \{0\})$ is homotopic to $f(C'_i \times \{1\})$ and $g_m^{(j)}(C'_i \times \{0\})$,

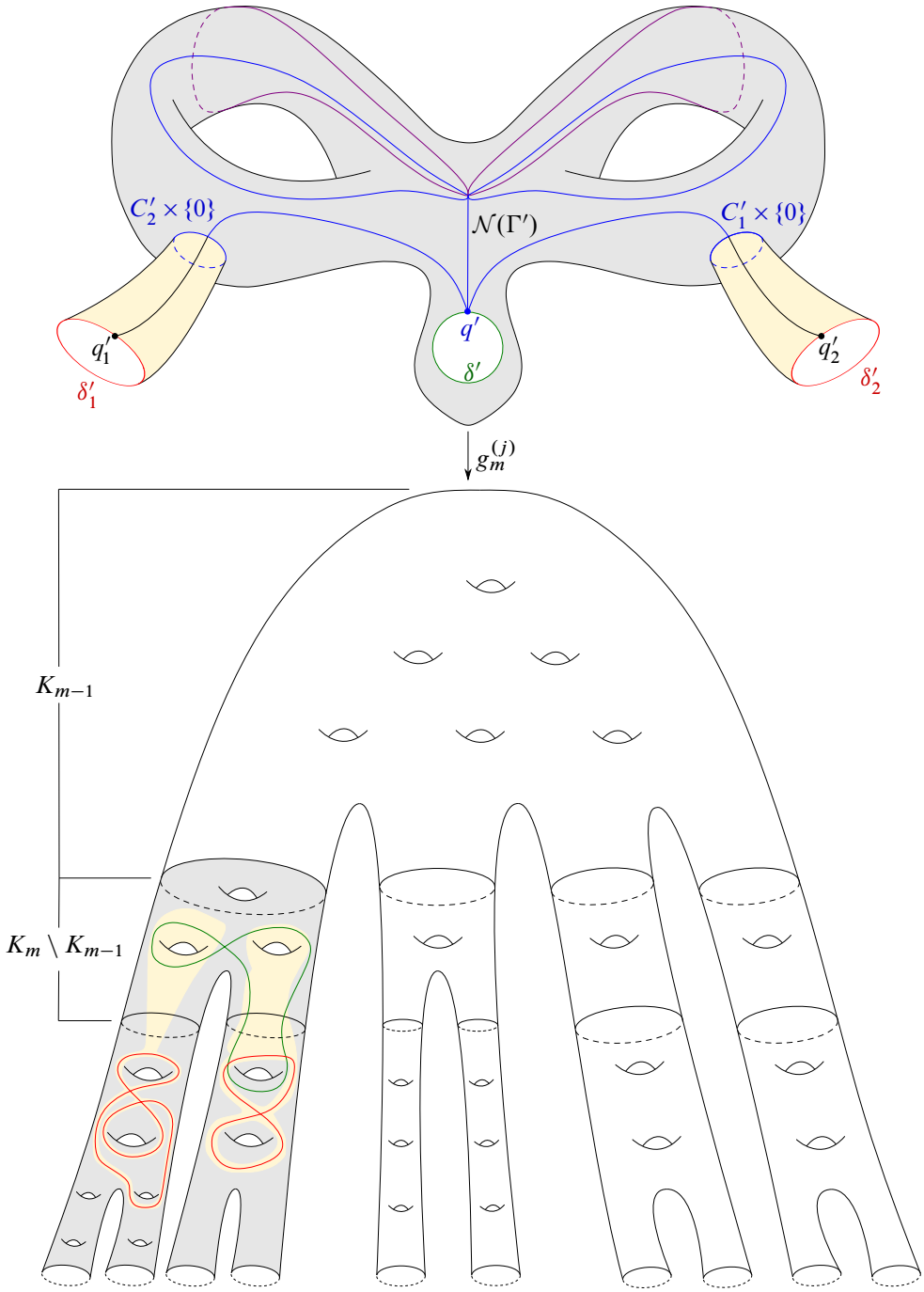


Figure 3: Description of $g_m^{(j)}\colon V' = V_m^{(j)} \to V \subset \Sigma \setminus K_{m-1}$. On the top, ie in V', the purple, blue, and green portions (arcs and circles) form Γ', and blue and black arcs form $T_m'^{(j)}$. At the bottom: the grey and yellow shades indicate $g_m^{(j)}(V')$ and $g_m^{(j)}(C_i' \times [0,1])$, respectively, and red and green loops denote $g_m^{(j)}(C_i' \times \{1\})$ and $g_m^{(j)}(\delta')$, respectively, where $i = 1, 2$.

we deduce that γ is homotopic to $g_m^{(j)}(C' \times \{0\})$ in Σ, and hence in $\Sigma \setminus K_{m-1}$ by Lemma 2.2. We use this homotopy to extend $g_m^{(j)}$ to $C_i' \times [0,1]$.

We also extend the tree T'_{m-1} to a tree $T_m'^{(j)}$ and modify f as in the base case. Let the boundary components of V' other than δ' be $\delta'_1, \dots, \delta'_n$, $\delta'_i = C'_i \times \{1\}$. Pick a point $q'_i \in \delta'_i$ for each i, $1 \le i \le n$. Pick a family of arcs θ'_i with interiors disjoint so that θ'_i is an arc from q' to q'_i in K'_1. Let $T_m'^{(j)}$ be the union of these arcs. As $T_m'^{(j)}$ is contractible we can homotope f fixing $\Sigma' \setminus (\mathrm{int}(V'))$ so that $f|T_m'^{(j)} = g_m^{(j)}|T_m'^{(j)}$.

Note that for $1 \le i \le n$, $f(q'_i) = g_m^{(j)}(q'_i)$. Further, for $1 \le i \le n$, as

$$(f|_{V'})_* = (g_m^{(j)})_* \colon \pi_1(V', q') \to \pi_1(\Sigma, f(q')), \quad \text{and} \quad f|_{\theta'_i} = g_m^{(j)}|_{\theta'_i},$$

we deduce by using the change of base-point isomorphisms corresponding to θ'_i and $f \circ \theta'_i = g_m^{(j)} \circ \theta'_i$ that $(f|_{V'})_* = (g_m^{(j)})_* \colon \pi_1(V', q'_i) \to \pi_1(\Sigma, f(q'_i))$.

3.7 Combining the maps

We define $g_m \colon K'_m \to \Sigma$ as the unique map extending g_{m-1} whose restriction to the component $V'^{(j)}$ of $\overline{K'_m \setminus K'_{m-1}}$ is $g_m^{(j)}$. Observe that $g_m(K'_m \setminus K'_{m-1}) \subset \overline{\Sigma \setminus K_{m-1}}$ and $g_m(\partial K'_m) \subset \Sigma \setminus K_m$.

The tree T'_m is $T'_{m-1} \cup \bigcup_j (T_m'^{(j)})$. As the homotopies of f had disjoint support, we take f to be the result of the composition of the homotopies. We have $f|_{T'_m} = g_m|_{T'_m}$ and the required equalities on the fundamental groups.

To summarize, we have constructed a subsurface, a map, and a tree satisfying the following properties:

Lemma 3.6 *The subsurface K'_m, map g_m and tree T'_m satisfy the following properties*:

(1) *The subsurface K'_m satisfies the conditions of Lemma 3.1.*

(2) $g_m(\partial K'_m) \subset \Sigma \setminus K_m$.

(3) $f|_{T'_{m-1}} = g_{m-1}|_{T'_{m-1}}$.

(4) *The terminal vertices of T'_{m-1} are in $\partial K'_{m-1}$, with one terminal vertex in each component.*

(5) *For a terminal vertex $q' \in T'_{m-1}$, $(f|_{K'_{m-1}})_* = (g_{m-1})_* \colon \pi_1(K'_{m-1}, q') \to \pi_1(\Sigma, q)$.*

(6) $g_m(K'_m \setminus K'_{m-1}) \subset \overline{\Sigma \setminus K_{m-1}}$.

The conditions except (6) are exactly those of Section 3.4, ie the conditions for the inductive step.

3.8 Constructing the proper map

As $K'_m \supset L'_m$ by Lemma 3.1 and the subsurfaces L'_m form an exhaustion, so do the subsurfaces K'_m. We define the map $g\colon \Sigma' \to \Sigma$ as the direct limit of the maps $g_m\colon K'_m \to \Sigma$.

We claim that g is proper. It suffices to show that $g^{-1}(K_n)$ is compact for all n. We claim that $g^{-1}(K_n) \subset K'_{n+1}$, and hence is a closed subset of a compact space, so compact. Namely, suppose $x \in g^{-1}(K_n)$ and $x \notin K'_{n+1}$. Then $x \in K'_m \setminus K'_{m-1}$ for some $m > n+1$. Hence, $g(x) = g_m(x) \subset \overline{\Sigma \setminus K_{m-1}}$ by statement (6) of Lemma 3.6, so $g(x) \notin \operatorname{int}(K_{m-1}) \supset K_{m-2} \supset K_n$ (the last containment by $m > n+1$), a contradiction.

Further, $g_* = f_*\colon \pi_1(\Sigma', p') \to \pi_1(\Sigma, p)$, so g is homotopic to f. By [1, Theorem 1], g is in turn (properly) homotopic to a homeomorphism, and hence so is f.

As g commutes with the Goldman bracket, we deduce that g is orientation preserving. Namely, first, modify g by an isotopy to ensure that g is a diffeomorphism (in dimension 2, homeomorphisms are isotopic, hence properly homotopic to diffeomorphisms). If g is orientation reversing, then as the signs of intersection numbers are reversed, we get the equation $[g_*(x'), g_*(y')] = -g_*([x', y'])$. As $f_* = g_*$ and f_* commutes with the Goldman bracket, we conclude that $[g_*(x'), g_*(y')] = 0$ for all x, y. As g_* is an isomorphism, it follows that the Goldman bracket is trivial on Σ, a contradiction as Σ is assumed not to be a plane or cylinder, so there exist classes x and y with $[x, y] \neq 0$.

This completes the proof of Theorem 1.1.

4 Proof of Theorem 1.2

We sketch the modifications needed to prove Theorem 1.2. A homeomorphism preserves intersection numbers, so clearly (1) implies (2) and (3). Further, clearly (2) implies (3). We show that (3) implies (1).

Assume now that, $I_\Sigma(f_*(x'), f_*(y')) = 0 \iff I_{\Sigma'}(x', y') = 0$ for all $x', y' \in \widehat{\pi}(\Sigma')$. We see that the proof of Theorem 1.1 goes through with this in place of the hypothesis that f preserves the Goldman bracket.

Namely, there are only two places where the hypothesis that the Goldman bracket is preserved are used directly:

(1) In Lemma 3.3 to show that if the closed curve γ' can be homotoped to be disjoint from α'_i, ie $I_{\Sigma'}(\gamma', \alpha'_i) = 0$, then $I_\Sigma(\gamma, \alpha_i) = 0$, where γ is homotopic to $f(\gamma')$ and α_i is homotopic to $f(\alpha'_i)$. Here an additional hypothesis is that α_i is a simple closed curve.

(2) In Lemma 3.5, to show that if the closed curve γ' can be homotoped to be disjoint from β', ie $I_{\Sigma'}(\gamma', \beta') = 0$, then $I_\Sigma(\gamma, \beta) = 0$, where γ is homotopic to $f(\gamma')$ and β is homotopic to $f(\beta)$. Here the additional hypothesis is that β' is a simple closed curve.

Clearly both these follow under the hypothesis that $I_\Sigma(f_*(x'), f_*(y')) = 0 \iff I_{\Sigma'}(x', y') = 0$ for all $x', y' \in \widehat{\pi}(\Sigma')$. Thus, the proof of Theorem 1.1 goes through with this in place of the hypothesis that f preserves the Goldman bracket.

References

[1] **S Das**, *Strong topological rigidity of noncompact orientable surfaces*, Algebr. Geom. Topol. 24 (2024) 4423–4469 MR Zbl

[2] **S Gadgil**, *The Goldman bracket characterizes homeomorphisms*, C. R. Math. Acad. Sci. Paris 349 (2011) 1269–1272 MR Zbl

[3] **W M Goldman**, *Invariant functions on Lie groups and Hamiltonian flows of surface group representations*, Invent. Math. 85 (1986) 263–302 MR Zbl

[4] **W Magnus**, **A Karrass**, **D Solitar**, *Combinatorial group theory*: *presentations of groups in terms of generators and relations*, Interscience, New York (1966) MR Zbl

[5] **J Stillwell**, *Appendix: The Dehn–Nielsen Theorem*, from "Papers on group theory and topology" (M Dehn, editor), Springer (1987) 363–396 MR Zbl

Department of Mathematics, Indian Institute of Science
Bangalore, India

sumantadas@iisc.ac.in, gadgil@iisc.ac.in, ajaynair@iisc.ac.in

Received: 27 July 2023 Revised: 20 March 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3603–3626
DOI: 10.2140/agt.2025.25.3603
Published: 1 October 2025

A geometric computation of cohomotopy groups in codegree one

MICHAEL JUNG
THOMAS O ROT

Using geometric arguments, we compute the group of homotopy classes of maps from a closed $(n+1)$-dimensional manifold to the n-sphere for $n \geq 3$. Our work extends results of Kirby, Melvin and Teichner for closed oriented 4-manifolds, and of Konstantis for closed $(n+1)$-dimensional spin manifolds, considering possibly nonorientable and nonspinnable manifolds. In the process, we introduce two types of manifolds that generalize the notion of odd and even 4-manifolds. Furthermore, for $n \geq 4$, we discuss applications of rank n spin vector bundles and obtain a refinement of the Euler class in the cohomotopy group that fully obstructs the existence of a nonvanishing section.

55Q55; 55N25, 57R15, 57R22

1 Introduction

The cohomotopy sets $\pi^n(X) = [X^{n+k}, S^n]$ of (unpointed) homotopy classes of maps from an $(n+k)$-dimensional cell complex X into an n-sphere play an important role in algebraic and differential topology. If n is large enough, this set can be given a group structure (see Spanier [13]). In the case when $k = 1$ and $n \geq 3$, Steenrod's main theorem [15, Theorem 28.1, page 318] implies that $\pi^n(X)$ fits into the short exact sequence of abelian groups

$$0 \to H^{n+1}(X;\mathbb{Z}_2)\big/\big((\mathrm{Sq}^2 \circ r)(H^{n-1}(X;\mathbb{Z}))\big) \to \pi^n(X) \to H^n(X;\mathbb{Z}) \to 0, \tag{1-1}$$

where $r\colon H^{n-1}(X;\mathbb{Z}) \to H^{n-1}(X;\mathbb{Z}_2)$ is the mod 2 reduction, and the epimorphism pulls back the generator of $H^n(S^n;\mathbb{Z})$. The question about the corresponding group extension has been revisited by Taylor [16] using methods of Larmore and Thomas [10]. While this approach relies on homotopical techniques involving Postnikov towers, it is worth exploring a more geometric approach if X is a manifold. Notably, Kirby, Melvin and Teichner [7] provided a geometric proof for the case of closed, oriented 4-manifolds, contributing to a deeper understanding of the subject. Additionally, Konstantis [9] presented a geometric proof for closed $(n+1)$-dimensional spin manifolds, introducing a new counting invariant in the process. The key tool in both cases is the Pontryagin–Thom construction that provides an isomorphism $\pi^n(X^{n+k}) \cong \mathbb{F}_k(X)$, where $\mathbb{F}_k(X)$ denotes the set of normally framed k-dimensional submanifolds of X up to normally framed bordism in $X \times [0,1]$.

© 2025 The Authors, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

In this article, we aim to complete the geometric picture when $k = 1$ and $n \geq 3$. More precisely, we provide a geometric computation of $\pi^n(X)$ for closed $(n+1)$-dimensional smooth manifolds that are not necessarily spin or orientable. We find that $\mathbb{F}_1(X)$ fits into a dual version of (1-1),

$$0 \to \ker(h) \to \mathbb{F}_1(X) \xrightarrow{h} H_1(X; o_X) \to 0, \tag{1-2}$$

where h maps the bordism class of a normally framed link to the corresponding twisted fundamental class in homology with coefficients in the orientation sheaf o_X. Morally speaking, the map $h\colon \mathbb{F}_1(X) \to H_1(X; o_X)$ forgets the normal framing but remembers its orientation. This allows us to use geometric techniques to determine $\ker(h)$ and compute the extension of the sequence.

As it turns out, the result depends on a certain classification of manifolds that generalizes the notion of odd and even oriented 4-manifolds. More precisely, if $n = 3$, that is if X is a 4-manifold, we say that X is of *type I* if there is a closed embedded surface with an orientable normal bundle but nonvanishing mod 2 self-intersection. For $n \geq 4$, we say that X is of *type I* if there is a closed embedded surface whose normal bundle is orientable but not trivial. The manifold X is of *type II* if it is not of type I. We now state the main theorem of this article.

Theorem 1.1 *If X is of type I, the forgetful map $h\colon \mathbb{F}_1(X) \to H_1(X; o_X)$ is an isomorphism. If X is of type II, then there is a short exact sequence of abelian groups*

$$0 \to \mathbb{Z}_2 \to \mathbb{F}_1(X) \xrightarrow{h} H_1(X; o_X) \to 0 \tag{1-3}$$

whose extension is classified by the unique element in $\operatorname{Ext}(H_1(X; o_X), \mathbb{Z}_2)$ mapped to $w_1^2(X) + w_2(X)$ under δ in the universal coefficient sequence

$$0 \to \operatorname{Ext}(H_1(X; o_X), \mathbb{Z}_2) \xrightarrow{\delta} H^2(X; \mathbb{Z}_2) \to \operatorname{Hom}(H_2(X; o_X), \mathbb{Z}_2) \to 0.$$

Recall that $w_1^2(X) + w_2(X)$ is the primary obstruction of finding a Pin^--structure on X. It follows from the main theorem that the sequence (1-3) splits if and only if X admits a Pin^--structure. In that case, we also obtain a correspondence between splitting maps and Pin^--structures, which is stated in the following theorem:

Theorem 1.2 *Suppose X is Pin^-. Then the sequence (1-3) splits. Moreover, if X is orientable, all equivalent* Spin*-structures of X are in one-to-one correspondence with splitting maps of (1-3). On the other hand, if X is not orientable, all equivalent Pin^--structures of X are in two-to-one correspondence with splitting maps of (1-3). In that case, the difference is given by the action of $w_1(X)$.*

At last, we discuss applications for vector bundles and generalize the invariant introduced in [9]. More precisely, for the case that $n \geq 4$, we construct a refinement of the Euler class in the cohomotopy group $\pi^n(X)$ that gives a full obstruction of finding a nonvanishing section for spin vector bundles over X of rank n. If X is of type I, we obtain an explicit result:

Theorem 1.3 *Suppose X is an $(n+1)$-dimensional connected closed manifold of type I with $n \geq 4$. Let $E \to X$ be an oriented spin vector bundle of rank n. Then the Euler class $e(E)$ is zero if and only if E admits a nonvanishing section.*

If X has a Pin^--structure, we can apply the associated splitting map $\kappa : \mathbb{F}_1(X) \to \mathbb{Z}_2$ to the newly defined obstruction class. The corresponding number $\kappa(E)$ is called the *degree of E* and is well defined if the Euler class $e(E)$ vanishes. This leads to an explicit result for vector bundles over Pin^- manifolds:

Theorem 1.4 *Suppose X is an $(n+1)$-dimensional connected closed manifold that is Pin^- where $n \geq 4$. Let $E \to X$ be an oriented spin vector bundle of rank n. Then E admits a nonvanishing section if and only if both the Euler class $e(E)$ and the degree $\kappa(E)$ vanish.*

The article is organized as follows. We start by revisiting fundamental definitions and classical computations in Section 2. Moving on to Section 3, we present a geometric interpretation of the kernel $\ker(h)$. In Section 4, we prove the first part of Theorem 1.1 concerning manifolds of type I. Shifting focus to type II manifolds, in Section 5 we discuss the proof of the second part of Theorem 1.1. Dedicating Section 6 to Pin^- manifolds, we show the interplay between splitting maps and Pin^--structures, and we derive Theorem 1.2. Finally, in Section 7, we explore the implications for vector bundles and validate Theorem 1.3 as well as Theorem 1.4 both as a corollary of Theorem 7.5.

Acknowledgements

The authors would like to express their gratitude to Panagiotis Konstantis for his valuable time and insightful discussions regarding his results. They also extend their appreciation to Stefan Behrens and Lauran Toussaint for engaging in fruitful discussions that contributed to the project. In closing, our thanks go to the referee for their careful reading and insightful comments, which greatly enhanced the quality of this article.

2 Preliminaries

In what follows we assume that X is a closed, connected $(n+1)$-dimensional smooth manifold, where $n \geq 3$. If $\iota\colon M \looparrowright X$ is an immersion, we denote by ν_ι the associated normal bundle. Provided that $\iota\colon M \hookrightarrow X$ is an embedding, and if the context allows, we may suppress the map ι and denote the normal bundle simply by ν_M. A trivialization of a vector bundle up to a homotopy through trivializations is referred to as a *framing*.

Pin-structures

A Pin-structure generalizes the notion of a Spin-structure to nonorientable vector bundles. While $\mathrm{SO}(r)$ is double covered by $\mathrm{Spin}(r)$, the group $\mathrm{O}(r)$ has two possible central $\mathbb{Z}_2$ extensions, denoted by $\mathrm{Pin}^{\pm}(r)$.

Suppose $E \to X$ is a vector bundle over X of rank r. A $\mathrm{Pin}^{\pm}$-principal bundle over X that lifts the frame bundle of E is called a $\mathrm{Pin}^{\pm}$-*structure over* E. In our article, we focus entirely on Pin^{-}-structures. The obstruction to putting a Pin^{-}-structure on E is given by $w_1^2(E) + w_2(E)$. If E admits a Pin^{-}-structure, the set of all equivalent Pin^{-}-structures is an $H^1(X;\mathbb{Z}_2)$-torsor. In other words, the group $H^1(X;\mathbb{Z}_2)$ acts simply transitive on the set of equivalent Pin^{-}-structures. We denote the (stable) bordism group of r-dimensional manifolds with a Pin^{-}-structure on the *tangent bundle* by $\Omega_r^{\mathrm{Pin}^-}$. An in-depth analysis of both Pin^{-} and Pin^{+}-structures and their low-dimensional bordism groups was done by Kirby and Taylor [8].

Homology with twisted coefficients

Since we are interested in nonorientable manifolds as well, we need to consider *homology with local coefficients*. An exposition of the definition and properties, also for cohomology, can be found in Spanier [14] as well as Davis and Kirk [3]. Recall that the *orientation sheaf* is defined as the local coefficient system

$$o_X = \bigsqcup_{x \in X} H_{n+1}(X, X - \{x\}; \mathbb{Z}),$$

sometimes referred to as *twisted coefficients*. Our object of interest is the associated homology $H_*(X; o_X)$.

We are particularly interested in the following functorial property. If $\iota\colon M \looparrowright X$ is an immersion of a $(k+1)$-dimensional manifold, we get an isomorphism of orientation bundles

$$\iota^*\big(\textstyle\bigwedge^{n+1} TX\big) \cong \bigwedge^{k+1} TM \otimes \bigwedge^{n-k} \nu_\iota.$$

An orientation of ν_ι induces an isomorphism $\bigwedge^{k+1} TM \cong \iota^*\big(\bigwedge^{n+1} TX\big)$, and hence an isomorphism of local coefficients $o_M \cong \iota^* o_X$. This then defines a pushforward $\iota_*\colon H_*(M; o_M) \to H_*(X; o_X)$.

Next, we discuss a universal coefficient theorem for local coefficients.[1] Suppose Γ is a local coefficient system over X, and G is an abelian group. We obtain a new local coefficient system

$$\mathrm{Hom}(\Gamma, G) := \bigsqcup_{x \in X} \mathrm{Hom}(\Gamma(x), G).$$

A possible formulation of a universal coefficient theorem for local coefficients in cohomology is then as follows:

Theorem 2.1 *Let Γ be a local system of X whose fibers are free abelian groups. Then we have a natural short exact sequence*

$$0 \to \mathrm{Ext}(H_{k-1}(X;\Gamma), G) \to H^k(X; \mathrm{Hom}(\Gamma, G)) \to \mathrm{Hom}(H_k(X;\Gamma), G) \to 0.$$

This sequence splits, but not naturally.

[1] Spanier mentions the possibility of deducing such a theorem in [14, page 283], but leaves the statement entirely open as an exercise.

Proof Observe that there is a canonical chain isomorphism

$$C^*(X;\mathrm{Hom}(\Gamma,G)) \xrightarrow{\cong} \mathrm{Hom}(C_*(X;\Gamma),G),$$

where the boundary maps on the right are given by precomposition with ∂. As the chain complex $C_*(X;\Gamma)$ is free abelian when Γ has free abelian fibers, the proof remains the same as for ordinary cohomology. We refer to Hatcher [6, Theorem 3.2] for the details. □

Our special case of interest entails $\Gamma = o_X$ and $G = \mathbb{Z}_2$. Then $\mathrm{Hom}(o_{X,x},\mathbb{Z}_2) \cong \mathbb{Z}_2$ for any $x \in X$, and therefore $\mathrm{Hom}(o_X,\mathbb{Z}_2) \cong X \times \mathbb{Z}_2$. For this special case, we obtain the sequence

$$0 \to \mathrm{Ext}(H_{k-1}(X;o_X),\mathbb{Z}_2) \xrightarrow{\delta} H^k(X;\mathbb{Z}_2) \xrightarrow{r^*} \mathrm{Hom}(H_k(X;o_X),\mathbb{Z}_2) \to 0, \tag{2-1}$$

where $r\colon H_k(X;o_X) \to H_k(X;\mathbb{Z}_2)$ is the mod 2 reduction with $H^k(X;\mathbb{Z}_2)$ and $\mathrm{Hom}(H_k(X;\mathbb{Z}_2),\mathbb{Z}_2)$ identified.

Analogous to the (untwisted) oriented case, we have *twisted Poincaré duality* for any manifold:

$$\mathrm{PD}_{\mathrm{tw}}\colon H^k(X;\mathbb{Z}) \xrightarrow{\cong} H_{n+1-k}(X;o_X).$$

To prove this, replace the ordinary cap product

$$\cap\colon C_p(X;\mathbb{Z}) \times C^q(X;\mathbb{Z}) \to C_{p-q}(X;\mathbb{Z})$$

with the twisted cap product

$$\cap\colon C_p(X;o_X) \times C^q(X;\mathbb{Z}) \to C_{p-q}(X;o_X).$$

Every manifold has a preferred *twisted fundamental class* because

$$H_{n+1}(X, X-\{x\};o_X) \cong H_{n+1}(X, X-\{x\};\mathbb{Z}) \otimes H_{n+1}(X, X-\{x\};\mathbb{Z})$$

has a preferred generator. The isomorphism can be understood through a twisted version of the universal coefficient theorem in homology, but we omit the details. The Poincaré duality isomorphism is then given by the twisted cap product with the twisted fundamental class.

Computations by Thom

It follows from Thom's seminal work [18] that every homology class in degree d with integer coefficients can be represented by a closed d-dimensional submanifold as long as d remains sufficiently small. The twisted case is analogous and is explained by Atiyah [1] in more detail. The idea is to consider the following commutative diagram:

$$\begin{array}{ccc} [X,\mathrm{MSO}(k)] & \xrightarrow{\cong} & \mathbb{L}_{n+1-k}(X;o_X) \\ \downarrow{\scriptstyle u_*} & & \downarrow \\ H^k(X;\mathbb{Z}) & \xrightarrow[\mathrm{PD}_{\mathrm{tw}}]{\cong} & H_{n+1-k}(X;o_X) \end{array}$$

Here, we denote by $\mathbb{L}_d(X; o_X)$ the set of closed d-dimensional submanifolds of X with an orientation of the normal bundle up to bordism in $X \times [0,1]$ with an orientation of the normal bundle. The top horizontal arrow is given by the Pontryagin–Thom construction. The left vertical arrow is induced by the Thom class $u\colon \mathrm{MSO}(k) \to K(\mathbb{Z}, k)$, where $\mathrm{MSO}(k)$ denotes the Thom space of the universal bundle for $\mathrm{SO}(k)$. The right vertical arrow is given by the pushforward of the twisted fundamental class. Thom [18] has shown that $u\colon \mathrm{MSO}(k) \to K(\mathbb{Z}, k)$ is $(k+2)$-connected. That is if we assume $k \geq n-1$, the left vertical arrow is an isomorphism, and we obtain $H_d(X; o_X) \cong \mathbb{L}_d(X; o_X)$ for $d = 1, 2$. Similar results hold for $\mathbb{Z}_2$ coefficients.

Two types of manifolds

Throughout this article, we consider two types of manifolds. Denote by $(w_1^2 + w_2)\colon H_2(X; \mathbb{Z}_2) \to \mathbb{Z}_2$ the functional associated with $w_1^2(X) + w_2(X)$ and by $r\colon H_2(X; o_X) \to H_2(X; \mathbb{Z}_2)$ the mod 2 reduction. We say that

$$X \text{ is of type I } :\Longleftrightarrow\ (w_1^2 + w_2) \circ r \not\equiv 0,$$
$$X \text{ is of type II } :\Longleftrightarrow\ (w_1^2 + w_2) \circ r \equiv 0.$$

It is clear from the definition that Pin^- manifolds are automatically of type II. But not all manifolds of type II are Pin^-. We say that X is of *type IIa* if it is Pin^- and of *type IIb* if it is of type II but not Pin^-.

We now give a geometric interpretation of this condition in terms of normal bundles of surfaces.

Lemma 2.2 *Let $\iota\colon \Sigma \looparrowright X$ be an immersed closed surface. Then*

$$(w_1^2 + w_2)(\iota_*[\Sigma]_2) = \langle w_2(\nu_\iota), [\Sigma]_2 \rangle.$$

Proof A simple computation reveals

$$\begin{aligned}(w_1^2 + w_2)(\iota_*[\Sigma]_2) &= \langle w_1^2(X) + w_2(X), \iota_*[\Sigma]_2 \rangle \\ &= \langle \iota^*(w_1^2(X) + w_2(X)), [\Sigma]_2 \rangle \\ &= \langle w_1^2(\Sigma) + w_2(\Sigma) + w_2(\nu_\iota) + w_1(\Sigma) w_1(\nu_\iota) + w_1^2(\nu_\iota), [\Sigma]_2 \rangle \\ &= \langle w_2(\nu_\iota) + w_1(\Sigma) w_1(\nu_\iota) + w_1^2(\nu_\iota), [\Sigma]_2 \rangle \\ &= \langle w_2(\nu_\iota), [\Sigma]_2 \rangle.\end{aligned}$$

Here we have used the fact that $\iota^*(TX) \cong T\Sigma \oplus \nu_\iota$, and that all surfaces are Pin^-, which means $w_1^2(\Sigma) + w_2(\Sigma) = 0$. The last step follows from Wu's formula which implies $w_1(\Sigma) w_1(\nu_\iota) = \mathrm{Sq}^1(w_1(\nu_\iota))$, and hence $w_1(\Sigma) w_1(\nu_\iota) = w_1^2(\nu_\iota)$. □

Proposition 2.3 *If there exist an immersed closed surface $\iota\colon \Sigma \looparrowright X$ with $w_1(\nu_\iota) = 0$ and $w_2(\nu_\iota) \neq 0$, then X is of type I. Conversely, if X is of type I, then there exists an embedded closed surface $\iota\colon \Sigma \hookrightarrow X$ such that $w_1(\nu_\iota) = 0$ and $w_2(\nu_\iota) \neq 0$.*

Proof Suppose $\iota\colon \Sigma \looparrowright X$ is an immersed closed surface with $w_1(\nu_\iota) = 0$. We have seen in Lemma 2.2 that the functional $w_1^2 + w_2$ measures the triviality of the second Stiefel–Whitney class of the normal

bundle $w_2(\nu_\iota)$. Together with functoriality in twisted coefficients, this shows the first implication. The converse follows from the fact that $H_2(X;o_X)$ is generated by embedded closed surfaces with an orientation of the normal bundle as discussed in the previous paragraph. □

Remark 2.4 It is worth mentioning that Proposition 2.3 implies that if there exists an immersed closed surface $\iota\colon \Sigma \looparrowright X$ with $w_1(\nu_\iota) = 0$ and $w_2(\nu_\iota) \neq 0$, then there automatically exists an embedded one with that property.

Equivalently, we obtain the geometric statement for manifolds of type II.

Corollary 2.5 *If all embedded surfaces $\iota\colon \Sigma \hookrightarrow X$ with $w_1(\nu_\iota) = 0$ have $w_2(\nu_\iota) = 0$, then X is of type II. Conversely, if X is of type II, then all immersed surfaces $\iota\colon \Sigma \looparrowright X$ with $w_1(\nu_\iota) = 0$ must have $w_2(\nu_\iota) = 0$.*

When X is a 4-manifold, the class $w_2(\nu_\Sigma)$ reports mod 2 self-intersection for normally oriented surfaces $\Sigma \subset X$. If X is in addition oriented, being type I is equivalent to X being *odd*, and being type II is equivalent to X being *even*. For higher dimensions, that is if $n \geq 4$, we can give another geometric description. The key ingredient is the following remark.

Remark 2.6 Recall that any vector bundle E of rank $n \geq 3$ over a 2-dimensional cell complex Σ is trivializable if and only if $w_1(E)$ and $w_2(E)$ vanish. This can be seen as follows. Since $n \geq 3$, the second Stiefel Whitney class $w_2(E)$ provides a full obstruction of finding an $(n-1)$-frame over Σ; see Milnor and Stasheff [11, Section 13]. Then the $(n-1)$-frame can be completed to an n-frame over Σ if and only if the bundle is orientable, or equivalently if and only if $w_1(E)$ vanishes.

With that remark, we deduce the following corollary:

Corollary 2.7 *Let $n \geq 4$. Then X is of type I if and only if there is a closed surface $\Sigma \subset X$ such that ν_Σ is orientable but not trivializable. Equivalently, X is of type II if and only if all closed surfaces $\Sigma \subset X$ for which ν_Σ is orientable have that ν_Σ is trivializable.*

The forgetful map

For any k, there is an obvious map $\mathbb{F}_k(X) \to \mathbb{L}_k(X;o_X)$ that forgets the framing but remembers the orientation of the normal bundle. For $k = 1$, we can identify $H_1(X;o_X) \cong \mathbb{L}_1(X;o_X)$, and hence we obtain a group homomorphism

$$h\colon \mathbb{F}_1(X) \to H_1(X;o_X).$$

Since any orientable vector bundle over a 1-dimensional cell complex is trivializable, this map is surjective. Therefore, $\mathbb{F}_1(X)$ always fits into the short exact sequence (1-2). In the following sections, we determine $\ker(h)$ and classify the extension with geometric methods.

3 The kernel of the forgetful map

To gain insights into the kernel of $h\colon \mathbb{F}_1(X) \to H_1(X; o_X)$, we need to familiarize ourselves with certain properties of framings over surfaces with boundary. For what follows, we call a framing *positive* if it is compatible with a given orientation.

Lemma 3.1 *Let Y be a 2-dimensional cell complex and $E \to Y$ be a trivializable vector bundle of rank $n \geq 3$ with a given orientation. Then distinct positive framings of E are in one-to-one correspondence with elements of $H^1(Y; \mathbb{Z}_2)$.*

Proof All distinct positive framings of E are in one-to-one correspondence with homotopy classes in $[Y, \mathrm{SO}(n)]$ via composition with a fixed positive framing. We represent the generator of $H^1(\mathrm{SO}(n), \mathbb{Z}_2) \cong \mathbb{Z}_2$ by a map $u\colon \mathrm{SO}(n) \to K(\mathbb{Z}_2, 1)$. Since $\pi_2(\mathrm{SO}(n)) = 0$, the map u is 3-connected. Hence, the induced map $u_*\colon [Y, \mathrm{SO}(n)] \to H^1(Y; \mathbb{Z}_2)$ is an isomorphism. □

Proposition 3.2 *Let Σ be a compact connected surface with boundary components $\partial\Sigma = C \sqcup C_1 \sqcup \cdots \sqcup C_k$, where $k \geq 0$. Let E be an oriented vector bundle of rank $n \geq 3$ over Σ. Then any choice of positive framing of E over $C_1, \ldots, C_k$ extends to a positive framing over Σ. Moreover, the induced framing over C of any such extension only depends on the choice made over $C_1, \ldots, C_k$.*

Proof Since E is oriented and Σ is homotopy equivalent to a 1-dimensional cell complex, we may fix a positive framing of E. Then by Lemma 3.1, positive framings over any subcomplex $\iota\colon Y \hookrightarrow \Sigma$ are in one-to-one correspondence with elements in $H^1(Y; \mathbb{Z}_2)$. The restriction of a framing over Σ to a subcomplex Y is then given by the induced map $\iota^*\colon H^1(\Sigma; \mathbb{Z}_2) \to H^1(Y; \mathbb{Z}_2)$. If we set $Y = C_1 \sqcup \cdots \sqcup C_k$, then we have the following short exact sequence in relative cohomology:

$$H^1(\Sigma; \mathbb{Z}_2) \xrightarrow{\iota^*} H^1(Y; \mathbb{Z}_2) \to H^2(\Sigma, Y; \mathbb{Z}_2) \cong 0.$$

This sequence shows that all positive framings over Y can be extended across Σ and further restricted to C via $\jmath^*\colon H^1(\Sigma; \mathbb{Z}_2) \to H^1(C; \mathbb{Z}_2)$, where $\jmath\colon C \hookrightarrow \Sigma$ is the inclusion of C. For uniqueness, suppose $\alpha \in H^1(\Sigma; \mathbb{Z}_2) \cong \mathrm{Hom}(H_1(\Sigma; \mathbb{Z}_2), \mathbb{Z}_2)$. Then the restriction $\jmath^*\alpha$ is determined by

$$(\jmath^*\alpha)([C]_2) = \alpha(\jmath_*[C]_2) = \alpha(\iota_*[Y]_2) = \iota^*\alpha([Y]_2),$$

and hence only determined by the restriction $\iota^*\alpha$. The second equality follows from

$$\jmath_*[C]_2 = \iota_*[Y]_2 \in H_1(\Sigma; \mathbb{Z}_2)$$

since Σ is a bordism between C and $Y = C_1 \sqcup \cdots \sqcup C_k$. □

The case $k = 0$ deserves special attention.

Corollary 3.3 *Under the assumptions of Proposition 3.2, if Σ has only one boundary component C, then there is exactly one framing over C that extends over the whole surface Σ.*

We wish to isotope bordisms in $X \times [0,1]$ to cut off disks whose boundaries respect slices. If the projection to $[0,1]$ is Morse, this is immediate, and the subsequent proposition explains how we can achieve that.

Lemma 3.4 *Let Σ be a compact surface and $\iota_0 \colon \Sigma \hookrightarrow X \times [0,1]$ be an embedding mapping boundaries to boundaries. Then there is another embedding $\iota_1 \colon \Sigma \to X \times [0,1]$ mapping boundaries to boundaries such that*

(i) *ι_0 and ι_1 are isotopic relative to $\partial\Sigma$,*

(ii) *$(\mathrm{pr}_2 \circ \iota_1) \colon \Sigma \to [0,1]$ is a Morse function.*

Proof Let $f \colon \Sigma \to [0,1]$ be a Morse function on Σ. If Σ has boundary, we may choose the Morse function f in such a way that $f|_{\partial\Sigma} = (\mathrm{pr}_2 \circ \iota_0)|_{\partial\Sigma}$. Consider the homotopy $\iota_t \colon \Sigma \to X \times [0,1]$ with

$$\iota_t = \big((\mathrm{pr}_1 \circ \iota_0),\ (\mathrm{pr}_2 \circ \iota_0) + t[f - (\mathrm{pr}_2 \circ \iota_0)]\big).$$

We may choose f sufficiently close to $(\mathrm{pr}_2 \circ \iota_0)$ such that ι_t becomes an isotopy. This is possible because embeddings are open in the strong topology. □

For what follows, we fix a contractible circle $U \subset X$ and a disk $D \subset X \times [0,1]$ that bounds U. Then U has two possible normal framings. One framing extends over D, and we denote the corresponding framed circle by U_0. The other framing does not extend over D, and the associated framed circle is denoted by U_1.

Lemma 3.5 *The subgroup $\ker(h) \subset \mathbb{F}_1(X)$ is generated by $[U_1]$.*

Proof For any element $[L,\varphi] \in \ker(h)$, there is a normally oriented bordism $\Sigma \hookrightarrow X \times [0,1]$ from L to the empty set. Notice that Σ may not be connected, ie $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_k$. We isotope each connected component Σ_i according to Lemma 3.4. Then we can cut off an embedded disk D_i from each Σ_i such that ∂D_i sits in one slice of $X \times [0,1]$. Each ∂D_i carries a uniquely determined normal framing via Proposition 3.2. We thus get an induced normal framing of the disjoint union $\partial D_1 \sqcup \cdots \sqcup \partial D_k$ which makes it framed bordant to (L,φ). Now, each framed circle ∂D_i is framed bordant to U_0 or U_1. But since $[U_0] = 0$ in $\mathbb{F}_1(X)$, the element $[L,\varphi]$ is a multiple of $[U_1]$. □

Since two copies of U_1 are framed bordant to U_0 through a pair of pants, the element $[U_1]$ is of order at most two, and we get as a proposition:

Proposition 3.6 *The forgetful map $h \colon \mathbb{F}_1(X) \to H_1(X; o_X)$ is either an isomorphism or a two-to-one epimorphism.*

4 Manifolds of type I

We have seen that $[U_1]$ generates $\ker(h)$ by Lemma 3.5. This allows us to restate the first part of Theorem 1.1 for type I manifolds as follows:

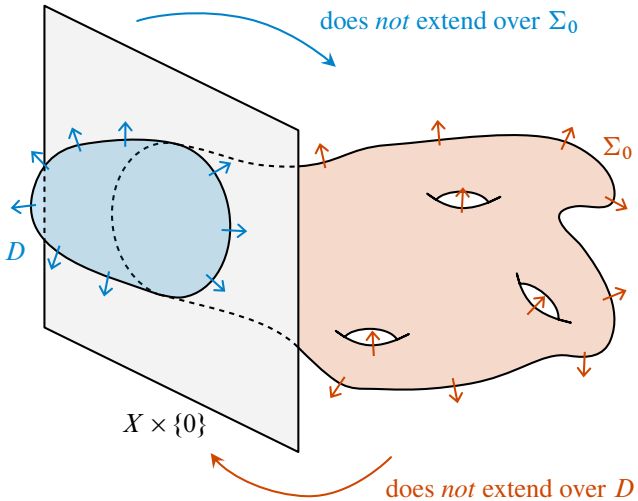


Figure 1: The surface $\Sigma = D \cup \Sigma_0$ in $X \times [0,1]$ where ν_Σ is oriented but not trivializable with many suppressed directions.

Theorem 4.1 *The manifold X is of type I if and only if $[U_1] = 0$ in $\mathbb{F}_1(X)$.*

Proof Suppose X is of type I. According to Proposition 2.3, there is a closed embedded surface $\iota\colon \Sigma \hookrightarrow X$ whose normal bundle is oriented and has nonvanishing second Stiefel–Whitney class. Denote by $\iota_0\colon \Sigma \hookrightarrow X \times [0,1]$ the map $s \mapsto (\iota(s), 0)$. We want to show that $[U_1] = 0$ in $\mathbb{F}_1(X)$. First, we isotope ι_0 to an embedding $\iota_1\colon \Sigma \hookrightarrow X \times [0,1]$ for which $\iota_1 \circ \mathrm{pr}_2$ is Morse using Lemma 3.4. Notice that this gives an isomorphism of normal bundles $\nu_{\iota_1} \cong \nu_\iota \oplus \varepsilon^1$. Then, we can cut off a disk from Σ whose boundary embeds into one slice of $X \times [0,1]$. Without loss of generality, we assume that this disk is given by the fixed disk D. That is, $\partial D = U$ and the framing of U_0 extends over D. Because Σ has a nontrivial normal bundle, the framing of U_0 cannot be extended over $\Sigma_0 = \Sigma - \mathrm{int}(D)$. The situation is depicted in Figure 1. Then, according to Corollary 3.3, the framing of U_1 can be extended over $\Sigma - \mathrm{int}(D)$. This implies $[U_1] = 0$.

For the converse assume that $[U_1] = 0$ in $\mathbb{F}_1(X)$. We want to show that there is an immersed closed surface in X with an oriented normal bundle whose second Stiefel–Whitney class does not vanish. By definition, the framing of U_0 extends over the given disk $D \subset X \times [0,1]$. Since $[U_1] = 0$, the normally framed circle U_1 can be extended over some surface $\Sigma_0 \subset X \times [0,1]$ with boundary U. We obtain a new closed surface $\iota_1\colon \Sigma \hookrightarrow X \times [-1,1]$ by gluing D and Σ_0 together along $U \subset X \times \{0\}$ from opposite sides. Visualization is provided in Figure 1. This surface has an oriented and nontrivial normal bundle of rank $n \geq 3$, and because of Remark 2.6, the second Stiefel Whitney class $w_2(\nu_{\iota_1})$ must not vanish. We project the surface to $X \times \{0\}$ and homotope it to an immersed surface $\iota\colon \Sigma \looparrowright X$. Denote by $\iota_0\colon \Sigma \looparrowright X \times [-1,1]$ the map $s \mapsto (\iota(s), 0)$. Since $n \geq 3$, the maps ι_0 and ι_1 are homotopic through immersions. Again, we obtain $\nu_{\iota_1} \cong \nu_\iota \oplus \varepsilon^1$ and thus $w_2(\nu_\iota) = w_2(\nu_{\iota_1})$. We conclude that ν_ι is oriented and $w_2(\nu_\iota)$ does not vanish, and hence by Proposition 2.3, X is of type I. □

Example 4.2 Consider $X = \mathbb{R}\mathrm{P}^{4k}$ for $k \geq 1$. Let $\mathbb{R}\mathrm{P}^2 \subset \mathbb{R}\mathrm{P}^{4k}$ be the standard embedding. A straightforward computation with Stiefel–Whitney classes shows that $w_1(\nu_{\mathbb{R}\mathrm{P}^2})$ vanishes but $w_2(\nu_{\mathbb{R}\mathrm{P}^2})$ does not.[2] Since $H_1(\mathbb{R}\mathrm{P}^{4k}; o_X) \cong H^{4k-1}(\mathbb{R}\mathrm{P}^{4k}; \mathbb{Z}) = 0$, we deduce $\mathbb{F}_1(\mathbb{R}\mathrm{P}^{4k}) = 0$.

5 Manifolds of type II

Throughout this section, we assume that X is of type II. In that case, the subgroup $\ker(h)$ is isomorphic to $\mathbb{Z}_2$. Hence we obtain the short exact sequence (1-3) stated in the introduction, and it remains to determine the extension.

We discuss some required background in homological algebra. Consider the following short exact sequence of (local) coefficients:

$$0 \to o_X \xrightarrow{\cdot 2} o_X \xrightarrow{\text{mod } 2} X \times \mathbb{Z}_2 \to 0.$$

This gives rise to connecting homomorphisms $\beta_k : H_{k+1}(X; \mathbb{Z}_2) \to H_k(X; o_X)$ which we call *twisted Bockstein homomorphisms*. In cohomology, we have Bockstein homomorphisms

$$\beta^k : H^k(X; \mathbb{Z}_2) \to H^{k+1}(X; \mathbb{Z})$$

associated with the sequence $\mathbb{Z} \to \mathbb{Z} \to \mathbb{Z}_2$. In a similar way to the ordinary Bockstein homomorphisms, the following diagram commutes:

$$\begin{array}{ccccc} H^{n-k}(X; \mathbb{Z}_2) & \xrightarrow{\beta^{n-k}} & H^{n+1-k}(X; \mathbb{Z}) & \xrightarrow{\text{mod } 2} & H^{n+1-k}(X; \mathbb{Z}_2) \\ \downarrow{\scriptstyle \mathrm{PD}_2}\cong & & \downarrow{\scriptstyle \mathrm{PD}_{\mathrm{tw}}}\cong & & \downarrow{\scriptstyle \mathrm{PD}_2}\cong \\ H_{k+1}(X; \mathbb{Z}_2) & \xrightarrow[\beta_k]{} & H_k(X; o_X) & \xrightarrow[\text{mod } 2]{} & H_k(X; \mathbb{Z}_2) \end{array} \tag{5-1}$$

We appeal to Munkres [12, Lemma 69.2] for proof of the commutativity of the left square in the oriented nontwisted case.

Remark 5.1 Note that the top row of (5-1) is precisely given by the first Steenrod square Sq^1. Let $\iota : \Sigma \hookrightarrow X$ be an embedded (or immersed) surface in X. Then β_1 sends the fundamental class $\iota_*[\Sigma]_2$ to a class in $H_1(X; o_X)$ whose mod 2 reduction is the Poincaré dual of the shriek $\iota_! w_1(\nu_\iota)$.[3]

Next, we would like to classify the extension given in (1-3) and give it a geometric interpretation. What follows revisits the discussion in [7, pages 6–7] applied to the twisted case. Consider the free resolution

$$0 \to B_1 \xrightarrow{i} Z_1 \to H_1(X; o_X) \to 0, \tag{5-2}$$

where $i : B_1 \to Z_1$ is the inclusion of 1-boundaries into 1-cycles in the singular chain complex of the local coefficients o_X. Using this free resolution, we can identify $\mathrm{Ext}(H_1(X; o_X), \mathbb{Z}_2)$ with $\mathrm{coker}(i^*)$.

[2] Recall that $H^*(\mathbb{R}\mathrm{P}^{n+1}; \mathbb{Z}_2) \cong \mathbb{Z}_2[a]/\langle a^{n+2}\rangle$ and $w(\mathbb{R}\mathrm{P}^{n+1}) = (1+a)^{n+2}$.

[3] The original result goes back to Thom [17] which is further generalized by Eccles and Grant [4].

That is, an element of $\operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2)$ is represented by a homomorphism $\phi\colon B_1\to\mathbb{Z}_2$. Denote by $\operatorname{Tor}_2(H_1(X;o_X))$ the subgroup of $H_1(X;o_X)$ with all elements of order at most two. Then we get a homomorphism

$$\Psi\colon \operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2)\to \operatorname{Tor}_2(H_1(X;o_X))^*,\quad [\phi]\mapsto([z]\mapsto\phi(2z)), \tag{5-3}$$

where we denoted $\operatorname{Hom}(-,\mathbb{Z}_2)=-^*$. Notice that this map is well defined because

(i) if $[z]\in\operatorname{Tor}_2(H_1(X;o_X))$, we have $2z\in B_1$,

(ii) for any $b\in B_1$, we have $\phi(2(z+b))=\phi(2z)$, and

(iii) if ϕ extends to a map $Z_1\to\mathbb{Z}_2$, then $\phi(2z)=0$.

The next proposition is an application of [5, Theorem 26.5] spelled out for our case.

Proposition 5.2 *The map $\Psi\colon \operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2)\to\operatorname{Tor}_2(H_1(X;o_X))^*$ is an isomorphism.*

Proof Since X is compact, $H_1(X;o_X)$ is a direct sum of cyclic groups. According to Cohen and Gluck [2], there is a stacked basis for the sequence (5-2), meaning there is a basis $\{z_i\}_{i\in I}$ of Z_1 such that $\{n_iz_i\}_{i\in J}$ is a basis of B_1 for $J\subseteq I$ and some $n_i\in\mathbb{Z}$. We split $B_1=B_1^{\text{even}}\oplus B_1^{\text{odd}}$ generated by the n_iz_i with n_i even or odd respectively. For injectivity, assume that $\Psi([\phi])=0$. This means that $\phi|_{B_1^{\text{even}}}\equiv 0$. In that case, ϕ can be extended to a map $Z_1\to\mathbb{Z}_2$, and therefore $[\phi]=0$ in $\operatorname{coker}(i^*)$ as desired. For surjectivity, fix a homomorphism $\varepsilon\colon\operatorname{Tor}_2(H_1(X;o_X))\to\mathbb{Z}_2$. We define $\phi_\varepsilon\colon B_1\to\mathbb{Z}_2$ as follows. On B_1^{even}, we declare $\phi_\varepsilon(n_iz_i)=\varepsilon\big(\big[\frac{1}{2}n_iz_i\big]\big)$, and otherwise we set $\phi_\varepsilon|_{B_1^{\text{odd}}}\equiv 0$. It is now easy to verify that $\Psi([\phi_\varepsilon])=\varepsilon$. □

Lemma 5.3 *The Bockstein β_1 is dual to δ in the twisted universal coefficient sequence (2-1), meaning that the following diagram commutes*:

$$\begin{array}{ccc}
\operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2) & \xrightarrow{\ \delta\ } & H^2(X;\mathbb{Z}_2)\\
\cong\downarrow & & \downarrow\cong\\
\operatorname{Tor}_2(H_1(X;o_X))^* & \xrightarrow{\ \beta_1^*\ } & H_2(X;\mathbb{Z}_2)^*
\end{array}$$

Proof Under the identification $\operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2)\cong\operatorname{coker}(i^*)$, the map

$$\delta\colon\operatorname{Ext}(H_1(X;o_X),\mathbb{Z}_2)\to H^2(X;\mathbb{Z}_2)$$

can be described as follows. Take a representative $\phi\colon B_1\to\mathbb{Z}_2$ and map it to

$$\phi\circ\partial\colon C_2(X;o_X)\to\mathbb{Z}_2.$$

This is a cocycle in the cochain complex $\operatorname{Hom}(C_*(X;o_X),\mathbb{Z}_2)$. But we have a canonical chain isomorphism $\operatorname{Hom}(C_*(X;o_X),\mathbb{Z}_2)\cong C^*(X;\mathbb{Z}_2)$. Under this identification, $\phi\circ\partial$ is mapped to a cocycle

representing the $\mathbb{Z}_2$-cohomology class $\delta([\phi])$. For the other composition, recall the construction of the Bockstein boundary homomorphism. Let $c \in C_2(X;\mathbb{Z}_2)$ be a cycle and $\bar{c} \in C_2(X;o_X)$ a lift of c, ie $\bar{c} \bmod 2 = c$. Then the Bockstein boundary map is given by $\beta_1([c]) = \left[\frac{1}{2}\partial\bar{c}\right]$. Precomposing β_1^* with the isomorphism in (5-3) gives precisely $[\phi \circ \partial]$, and hence the diagram commutes. □

The discussion thus far has primarily focused on homological algebra. The following lemma provides a more geometric perspective.

Lemma 5.4 *Let $\varepsilon_X : \operatorname{Tor}_2(H_1(X;o_X)) \to \mathbb{Z}_2$ be the homomorphism associated with the extension in (1-3). If $[L,\varphi]$ is in the preimage of $\operatorname{Tor}_2(H_1(X;o_X))$ under the map $h : \mathbb{F}_1(X) \to H_1(X;o_X)$, then*

$$(\varepsilon_X \circ h)([L,\varphi]) = \begin{cases} 0 & \text{if and only if } 2[L,\varphi] = [U_0], \\ 1 & \text{if and only if } 2[L,\varphi] = [U_1]. \end{cases} \tag{5-4}$$

Proof Using the free resolution (5-2), the extension of (1-3) is represented by a map $\phi : B_1 \to \mathbb{Z}_2$ that is induced by $\psi : Z_1 \to \mathbb{F}_1(X)$ in the following commutative diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & B_1 & \xrightarrow{i} & Z_1 & \longrightarrow & H_1(X;o_X) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\psi} & & \| & & \\ 0 & \longrightarrow & \mathbb{Z}_2 & \xrightarrow{u} & \mathbb{F}_1(X) & \xrightarrow{h} & H_1(X;o_X) & \longrightarrow & 0 \end{array}$$

For a reference see Spanier [14, Section 5.5.2]. Notice that the equivalence class $[\phi] \in \operatorname{coker}(i^*)$ does not depend on the choice of ψ. Let $[L,\varphi]$ be an element in the preimage of $\operatorname{Tor}_2(H_1(X;o_X))$ under h and $[z] = h([L,\varphi])$. Then $2[L,\varphi] \in \ker(h)$, that is $2[L,\varphi]$ is either $[U_0]$ or $[U_1]$. Because the diagram commutes, we obtain $(u \circ \phi)(2z) = 2[L,\varphi]$. But by definition, we have $(\varepsilon_X \circ h)([L,\varphi]) = \phi(2z)$, and hence (5-4) follows. □

Remark 5.5 The relation established in (5-4) remains invariant under any choice of initial framing φ, as the doubling of this choice has no contribution to the resulting framing, as indicated by Proposition 3.2.

Remark 5.6 By Proposition 3.6 and Lemma 5.4, it follows that any nonzero element $[L,\varphi]$ in the preimage of $\operatorname{Tor}_2(H_1(X;o_X))$ under h has either order two or four in $\mathbb{F}_1(X)$ depending on whether $(\varepsilon_X \circ h)([L,\varphi])$ is 0 or 1.

With our geometric perspective on the group extension, we are almost ready to approach the proof of the main theorem. However, before doing so, we require additional geometric tools.

Proposition 5.7 *Suppose M is a closed $(m+1)$-dimensional manifold with $m \geq 1$. Let $\iota : N \hookrightarrow M$ be an embedded compact manifold of dimension $k+1$ with boundary ∂N. Let τ be a normal framing of $\iota : N \hookrightarrow M$. Suppose V is an outward-pointing vector field in $TN|_{\partial N}$. Then $[\iota(\partial N), (\tau|_{\partial N}, \mathrm{d}\iota(V))] = 0$ in $\mathbb{F}_k(M)$.*

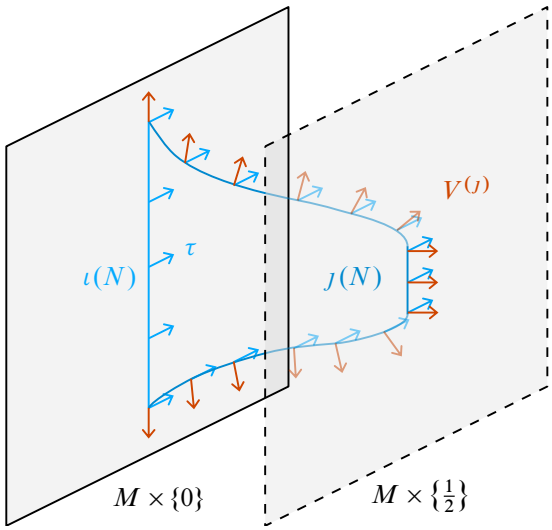


Figure 2: The construction of the framed null-bordism in Proposition 5.7.

Proof We take a collar neighborhood $U \subset N$ together with a diffeomorphism $\psi\colon \partial N \times [0,1) \to U$. Moreover, let $f\colon N \to \left[0, \frac{1}{2}\right]$ be a smooth function such that

(i) $f^{-1}(0) = \partial N$,

(ii) $f(U) = \left[0, \frac{1}{2}\right)$,

(iii) $f|_{M-U} \equiv \frac{1}{2}$, and

(iv) $(f|_U \circ \psi)(p, \cdot)\colon [0,1) \to \left[0, \frac{1}{2}\right)$ is strictly increasing for any $p \in \partial N$.

Consider the embedding $J\colon N \hookrightarrow M \times [0,1]$ defined via $p \mapsto (\iota(p), f(p))$. We extend the vector field V on ∂N to a smooth vector field onto U such that $\mathrm{d}f(V) < 0$ where we denote the extension by the same letter. Now, we define a vector field along $J\colon N \hookrightarrow M \times [0,1]$ given by

$$V^{(J)}(p) := \begin{cases} \left((1-2f(p))\mathrm{d}\iota(V(p)),\ 2f(p)\frac{\partial}{\partial t}\big|_{f(p)}\right) & \text{if } p \in U, \\ \left(0, \frac{\partial}{\partial t}\big|_{f(p)}\right) & \text{if } p \in N - U, \end{cases}$$

where $t \in [0,1]$ is the coordinate in the second component. Because $\mathrm{d}f(V) < 0$ and $f > 0$ in $U - \partial N$, the vector field $V^{(J)}$ represents a section of ν_J. For the normal framing $\tau = (v_1, \dots, v_{n-1})$, we denote $\tau^{(J)} = ((v_1, 0), \dots, (v_{n-1}, 0))$. We hence obtain a normal framing $(\tau^{(J)}, V^{(J)})$ of $J\colon N \hookrightarrow M \times [0,1]$ that restricts to $(\tau|_{\partial N}, \mathrm{d}\iota_0(V))$ on the boundary ∂N. An illustration of this construction is provided in Figure 2. □

As long as we are in the stable range, meaning $2k < m$, the set $\mathbb{F}_k(M)$ is an abelian group and inverses are given by flipping the orientation of the framing. Hence, we obtain the following corollary:

Corollary 5.8 *Suppose M is a closed $(m+1)$-dimensional manifold with $m \geq 1$, and suppose that $2k < m$. Let $\iota\colon N \hookrightarrow M$ be an embedded compact manifold of dimension $k+1$ with boundary*

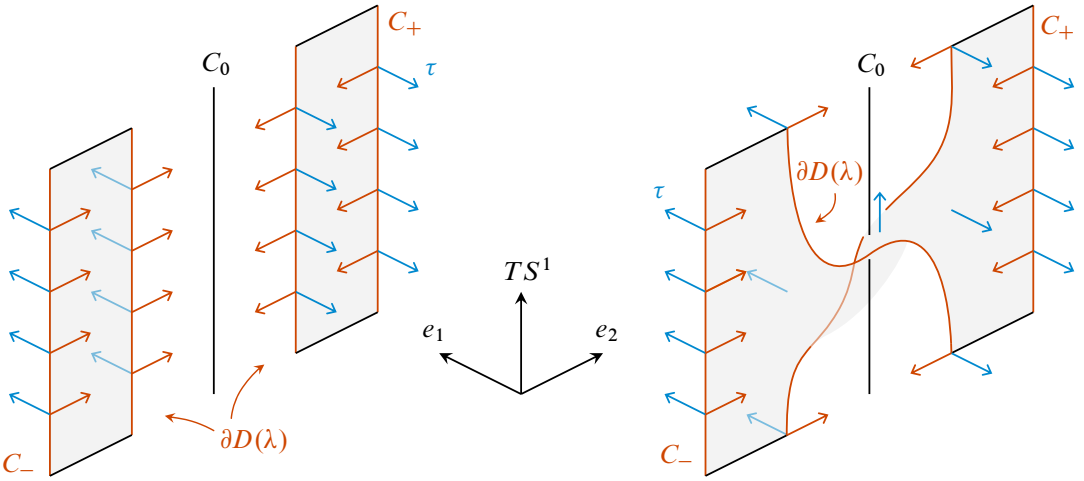


Figure 3: The construction of framed bordisms in Lemma 5.9. Left the case when λ is trivial, and right when λ is the nontrivial bundle.

$\partial N = B_0 \sqcup B_1$. Assume τ is a normal framing of $\iota\colon N \hookrightarrow M$. Suppose V is a vector field in $TN|_{\partial M}$ inward-pointing on B_0 and outward-pointing on B_1. Then in $\mathbb{F}_k(M)$,

$$[\iota(B_0), (\tau|_{B_0}, \mathrm{d}\iota(V|_{B_0}))] = [\iota(B_1), (\tau|_{B_1}, \mathrm{d}\iota(V|_{B_1}))].$$

We apply the corollary to a special situation that will play an important role in the proof of the main theorem.

Lemma 5.9 *Suppose λ is a rank one subbundle of the trivial bundle $\varepsilon^2 = S^1 \times \mathbb{R}^2$. Let $D(\lambda) \subset \varepsilon^2$ be the disk bundle associated with λ, and C_0 be the zero section of ε^2. Then $2[C_0, (e_1, e_2)] = [\partial D(\lambda), (v_1, v_2)]$ in $\mathbb{F}_1(\varepsilon^2)$, where (e_1, e_2) is the standard frame, v_2 is an inward-pointing normal of $\partial D(\lambda) \subset D(\lambda)$, and (v_1, v_2) is completed to a frame of $v_{\partial D(\lambda)}$ that is induced by the orientation of (e_1, e_2).*

Proof Let $C_\pm$ be two copies of C_0 displaced in the positive and negative e_2-direction respectively. We endow $C_\pm$ with the normal frame $(\pm e_1, \pm e_2)$ so that $C_- \sqcup C_+$ represents $2[C_0, (e_1, e_2)]$. We must consider two cases: either λ is trivializable or not. For the first case, assume without loss of generality that $\lambda = S^1 \times \{0\} \times \mathbb{R}$, meaning that e_2 is a frame of λ. Then $\pm e_1$ extends trivially to a normal frame τ of $D(\lambda)$, and the statement follows from Corollary 5.8 as depicted in the left of Figure 3. Suppose that λ is nontrivializable, that is $D(\lambda)$ is homeomorphic to the Möbius band. In that case, the bordism is given by a pair of pants as depicted in the right of Figure 3, and $\pm e_1$ over $C_\pm$ extends to a normal frame τ over the pair of pants. Applying Corollary 5.8, where $M = \varepsilon^2$, we get a normal frame on $\partial D(\lambda)$ that twists around $\partial D(\lambda)$ once. This frame is homotopic (through frames) to (v_1, v_2) defined in the lemma, which proves the claim. □

The following proposition relates Lemma 5.4 to normal bundles of embedded surfaces, and thus provides a deeper geometric insight.

Proposition 5.10 *Assume that X is of type II, and let $\Sigma \subset X$ be a closed connected surface. Moreover, let $L \subset X$ be a link with an oriented normal bundle representing the class $\beta_1([\Sigma]_2) \in H_1(X; o_X)$. Then $w_2(\nu_\Sigma) = 0$ if and only if $2[L, \varphi] = [U_0]$ independent of the choice of normal positive framing φ over L.*

Proof First, observe that L with its orientation on ν_L is homologous in $H_1(X; o_X)$ to a closed curve $C \subset \Sigma \subset X$ with an orientation of ν_C. This follows from naturality of the Bockstein homomorphism

$$\begin{array}{ccc} H_2(\Sigma; \mathbb{Z}_2) & \xrightarrow{\beta_1} & H_1(\Sigma; \iota^* o_X) \\ \downarrow{\scriptstyle \iota_*} & & \downarrow{\scriptstyle \iota_*} \\ H_2(X; \mathbb{Z}_2) & \xrightarrow{\beta_1} & H_1(X; o_X) \end{array}$$

where $\iota\colon \Sigma \hookrightarrow X$ is the inclusion map. Recall that the mod 2 homology class of C is Poincaré dual to $w_1(\nu_\Sigma)$ according to Remark 5.1. Suppose ν_Σ is orientable, meaning that $[C]_2 = 0$ in $H_1(\Sigma; \mathbb{Z}_2)$. It follows that C is bounded by a connected component of $\Sigma - C$. This implies that C with an orientation of ν_C is null-homologous in $H_1(X; o_X)$. Then the statement follows automatically since $[C, \varphi] \in \ker(h)$ for any normal framing φ, and $w_2(\nu_\Sigma) = 0$ by Corollary 2.5 as X is of type II.

Now suppose that ν_Σ is nonorientable. As C represents the Poincaré dual of $w_1(\nu_\Sigma)$, it follows that $\nu_\Sigma|_{\Sigma - C}$ is orientable. Notice that since C is not null-homologous, the complement $\Sigma - C$ remains connected. Denote by ν_C^Σ the normal bundle of C within Σ. Let $N \subset \Sigma$ be the closed tubular neighborhood associated with ν_C^Σ. We cut off an embedded disk D of $\Sigma - N$. Without loss of generality, we may assume that $\partial D = U$. We may also assume that U_0 is framed by (τ, V), where τ extends to a normal framing over D and V is an inward-pointing vector field of $\partial D \subset D$. That is, U_0 extends over a stretched version of D in $X \times [0, 1]$ according to Proposition 5.7. Let φ be any normal framing of C in X. We want to show that $2[C, \varphi] = [U_0]$ if and only if $w_2(\nu_\Sigma) = 0$. By obstruction theory, this is equivalent to showing that $2[C, \varphi] = [U_0]$ if and only if ν_Σ admits $n-2$ linearly independent sections. Notice that we can split ν_C into line bundles as follows:

$$\nu_C \cong \nu_\Sigma|_C \oplus \nu_C^\Sigma \cong \varepsilon^{n-2} \oplus \gamma \oplus \nu_C^\Sigma.$$

Without loss of generality, we may choose the framing $\varphi = (e_1, \dots, e_n)$ in such a way that (e_{n-1}, e_n) is a framing of $\gamma \oplus \nu_C^\Sigma$. Notice that we can extend $(e_1, \dots, e_{n-2})$ over N as an $(n-2)$-framing. By Lemma 5.9, the bordism class $2[C, \varphi]$ is represented by ∂N with normal framing $(e_1, \dots, e_{n-2}, v_1, v_2)$, where v_1 is normal to N, and v_2 is the inward-pointing normal of $\partial N \subset N$. As discussed in Proposition 3.2, the normal framing $\tau = (e_1, \dots, e_{n-2}, v_1)$ can be extended over $\Sigma - (N \cup D)$ to give a framing of $\nu_\Sigma|_U$. If we assume $2[C, \varphi] = [U_0]$, then the framing τ also extends over D. This implies that $(e_1, \dots, e_{n-2})$ is an $(n-2)$-framing of ν_Σ as desired. For the converse direction, assume that $(e_1, \dots, e_{n-2})$ is an $(n-2)$-framing of ν_Σ.[4] But since $\nu_\Sigma|_{\Sigma - N}$ is oriented and $(e_1, \dots, e_{n-2})$ extends over D, the framing $\tau = (e_1, \dots, e_{n-2}, v_1)$ extends over D as well. We thus obtain $2[C, \varphi] = [U_0]$ as required. □

[4] Such a framing predetermines the splitting $\nu_\Sigma|_C \cong \varepsilon^{n-2} \oplus \gamma \oplus \nu_C^\Sigma$ into line bundles.

$(n+1) \bmod 4$	0	1	2	3
$w_1(\nu_{\mathbb{R}\mathrm{P}^2})$	0	1	0	1
$w_2(\nu_{\mathbb{R}\mathrm{P}^2})$	1	1	0	0
type	I	IIb	IIa	IIa
$\pi^n(\mathbb{R}\mathrm{P}^{n+1})$	0	$\mathbb{Z}_4$	$\mathbb{Z}_2$	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$

Table 1: First two Stiefel–Whitney classes of $\nu_{\mathbb{R}\mathrm{P}^2}$ inside $\mathbb{R}\mathrm{P}^{n+1}$ as well as the type of $\mathbb{R}\mathrm{P}^{n+1}$ and the associated cohomotopy group.

Remark 5.11 Let $C \subset X$ be a closed curve with an oriented normal bundle, representing a nonzero element of $\mathrm{Tor}_2(H_1(X; o_X))$. Since $\beta_1 \colon H_2(X; \mathbb{Z}_2) \to \mathrm{Tor}_2(H_1(X; o_X))$ is surjective, there exists an embedded closed surface $\iota \colon \Sigma \hookrightarrow X$ for which C represents $\iota_! w_1(\nu_\Sigma)$. Proposition 5.10 shows that $[C, \varphi] \in \mathbb{F}_1(X)$ with any normal framing φ is of order two if $w_2(\nu_\Sigma) = 0$, and of order four otherwise.

Example 5.12 We consider $X = \mathbb{R}\mathrm{P}^{n+1}$. Let $\mathbb{R}\mathrm{P}^2 \subset \mathbb{R}\mathrm{P}^{n+1}$ be the standard embedding. Notice that the fundamental class $[\mathbb{R}\mathrm{P}^2]_2$ generates $H_2(\mathbb{R}\mathrm{P}^{n+1}; \mathbb{Z}_2)$. The group $H_1(\mathbb{R}\mathrm{P}^{n+1}; o_X)$ is of order at most two. This means that the twisted Bockstein $\beta_1 \colon H_2(\mathbb{R}\mathrm{P}^{n+1}; \mathbb{Z}_2) \to H_1(\mathbb{R}\mathrm{P}^{n+1}; o_X)$ is surjective, and $\beta_1([\mathbb{R}\mathrm{P}^2]_2)$ generates $H_1(\mathbb{R}\mathrm{P}^{n+1}; o_X)$. Notice that $w_1(\nu_{\mathbb{R}\mathrm{P}^2}) = 0$ if and only if $H_1(\mathbb{R}\mathrm{P}^{n+1}; o_X) = 0$. Consequently, if we compute $w_1(\nu_{\mathbb{R}\mathrm{P}^2})$ and $w_2(\nu_{\mathbb{R}\mathrm{P}^2})$, we can determine the type of $\mathbb{R}\mathrm{P}^{n+1}$, and also the associated cohomotopy group $\pi^n(\mathbb{R}\mathrm{P}^{n+1})$ by Remark 5.11. The computation is similar to Example 4.2, and the results are summarized in Table 1.

We are now ready to prove the second part of Theorem 1.1 for type II manifolds. For convenience, we restate the theorem here.

Theorem 5.13 *If X is of type II, then there is an extension of abelian groups*

$$0 \to \mathbb{Z}_2 \to \mathbb{F}_1(X) \xrightarrow{h} H_1(X; o_X) \to 0,$$

classified by the unique element in $\mathrm{Ext}(H_1(X, o_X), \mathbb{Z}_2)$ that maps to $w_1^2(X) + w_2(X)$ via δ in the universal coefficient sequence (2-1).

Proof From the previous discussion, we have seen that the extension is uniquely determined by the functional

$$\varepsilon_X \colon \mathrm{Tor}_2(H_1(X; o_X)) \to \mathbb{Z}_2$$

through (5-3). Together with Lemma 5.3, which relates δ to the dual Bockstein, it is therefore sufficient to show that $\varepsilon_X \circ \beta_1 = w_1^2 + w_2$. Let $\Sigma \subset X$ be a closed connected surface representing a homology class in $H_2(X; \mathbb{Z}_2)$. Suppose the element $\beta_1([\Sigma]_2) \in \mathrm{Tor}_2(H_1(X; o_X))$ is represented by a normally oriented link L. Then it follows from Lemma 5.4 and Proposition 5.10 that both $(\varepsilon_X \circ \beta_1)([\Sigma]_2)$ and $\langle w_2(\nu_\Sigma), [\Sigma]_2 \rangle$ are either 0 or 1 depending on whether $2[L, \varphi]$ is $[U_0]$ or $[U_1]$ for any choice of framing φ. This implies

$$(\varepsilon_X \circ \beta_1)([\Sigma]_2) = \langle w_2(\nu_\Sigma), [\Sigma]_2 \rangle,$$

which is equal to $(w_1^2+w_2)([\Sigma]_2)$ by Lemma 2.2. The statement then follows from the fact that $H_2(X;\mathbb{Z}_2)$ is generated by closed embedded surfaces. □

Corollary 5.14 *The short exact sequence (1-3) splits if and only if X is Pin^-. Therefore, in that case, we have $\mathbb{F}_1(X) \cong H_1(X;o_X) \oplus \mathbb{Z}_2$.*

Corollary 5.14 generalizes the result obtained by Konstantis [9] to nonorientable manifolds. In particular, Konstantis examines the splitting maps and obtains a counting invariant from it. This is the subject of the following sections.

6 Manifolds of type IIa

In the previous section, we have shown that (1-3) splits if and only if X is Pin^-. In this section, we sort out the geometric interpretation of the splitting maps. Throughout this section, we assume that X is Pin^-, and we fix a Pin^--structure on X. Suppose (L_0,φ_0) is a normally framed link in X. As ν_{L_0} is trivialized via φ_0 and $TX|_{L_0}$ carries a Pin^--structure, TL_0 inherits a Pin^--structure σ_0 via stabilization (see [8, Lemma 1.6]). This defines a (stable) bordism class $[L_0,\sigma_0] \in \Omega_1^{\mathrm{Pin}^-}$.

Lemma 6.1 *The bordism class $[L_0,\sigma_0] \in \Omega_1^{\mathrm{Pin}^-}$ as constructed above does not depend on the choice of representative in the bordism class $[L_0,\varphi_0] \in \mathbb{F}_1(X)$.*

Proof We assume that (L_0,φ_0) and (L_1,φ_1) are normally framed bordant through $\Sigma \subset X\times[0,1]$ with normal framing φ. That is, (Σ,φ) restricts to (L_i,φ_i) when intersected with $X\times\{i\}$ for each $i=0,1$. Again, $T\Sigma$ inherits a Pin^--structure σ from $TX|_\Sigma$ because ν_Σ is trivialized through φ. By construction, the restriction of σ to $X\times\{i\}$ agrees with the Pin^--structure σ_i on TL_i induced by φ_i. Therefore, (Σ,σ) is a Pin^- bordism from (L_0,σ_0) to (L_1,σ_1) which proves the claim. □

As discussed in [8, Theorem 2.1], the group $\Omega_1^{\mathrm{Pin}^-}$ is canonically isomorphic to $\mathbb{Z}_2$ generated by the circle S^1_{Lie} with the Lie group framing. This gives rise to a well-defined group homomorphism $\kappa\colon \mathbb{F}_1(X)\to\mathbb{Z}_2$. Next, we show that κ restricted to $\ker(h)$ is an isomorphism.

Proposition 6.2 *The restriction $\kappa|_{\ker(h)}\colon \ker(h)\to\mathbb{Z}_2$ is an isomorphism.*

Proof The statement follows from the proof of $\Omega_1^{\mathrm{Pin}^-}\cong\mathbb{Z}_2$ in [8, Theorem 2.1]. Since the generator of $\Omega_1^{\mathrm{Pin}^-}$ is given by the circle S^1_{Lie} with the Lie group framing, its induced Pin^--structure does not extend over a disk. But $\ker(h)$ is generated by $[U_1]$ as shown in Lemma 3.5, and κ maps $[U_1]$ to the generator $[S^1_{\mathrm{Lie}}]$ in $\Omega_1^{\mathrm{Pin}^-}$. □

This proposition shows that κ is a splitting map for the short exact sequence (1-3). For the rest of this section, we show how exactly those splitting maps and Pin^--structures are tied together. Let $\mathcal{P}\mathrm{in}(X)$ be the set of Pin^--structures up to equivalence. Furthermore, we denote

$$\mathcal{S}\mathrm{plit}(X) := \{\text{splitting maps } \kappa\colon \mathbb{F}_1(X)\to\mathbb{Z}_2 \text{ of (1-3)}\}.$$

The construction in the preceding paragraphs gives rise to a map $\mathcal{P}\mathrm{in}(X) \to \mathcal{S}\mathrm{plit}(X)$. We show that this map satisfies an equivariant-like property with certain group actions.

First note that there is a simply transitive group action by $\mathrm{Hom}(H_1(X;o_X),\mathbb{Z}_2)$ upon $\mathcal{S}\mathrm{plit}(X)$. The group action is described as follows. Take a splitting map $\kappa\colon \mathbb{F}_1(X) \to \mathbb{Z}_2$ and a homomorphism $\gamma\colon H_1(X;o_X) \to \mathbb{Z}_2$. If (L,φ) is a normally framed link in X, we define the action of γ on κ by

$$(\kappa\cdot\gamma)([L,\varphi]) := \kappa([L,\varphi]) + (\gamma\circ h)([L,\varphi]).$$

It is not hard to check that this action is transitive and free.

Lemma 6.3 *The mod 2 reduction $r\colon H_1(X;o_X)\to H_1(X;\mathbb{Z}_2)$ induces an isomorphism*

$$\mathrm{Hom}(H_1(X;o_X),\mathbb{Z}_2) \xrightarrow{\cong} H^1(X;\mathbb{Z}_2)/\langle w_1(X)\rangle.$$

Proof Consider the twisted universal coefficient theorem in cohomology (see Theorem 2.1),

$$0\to \mathrm{Ext}(H_0(X;o_X),\mathbb{Z}_2)\to H^1(X;\mathbb{Z}_2)\xrightarrow{r^*}\mathrm{Hom}(H_1(X;o_X),\mathbb{Z}_2)\to 0.$$

Since X is connected, we have

$$H_0(X;o_X)\cong\begin{cases}\mathbb{Z} & \text{if } X \text{ is orientable,}\\ \mathbb{Z}_2 & \text{if } X \text{ is nonorientable.}\end{cases}$$

This implies that $\mathrm{Ext}(H_0(X;o_X),\mathbb{Z}_2)$ is either 0 or $\mathbb{Z}_2$ depending on whether X is orientable or not. Thus $\ker(r^*)$ is either 0 or $\mathbb{Z}_2$ depending on whether X is orientable or not. Let $w_1\colon H_1(X;\mathbb{Z}_2)\to\mathbb{Z}_2$ denote the homomorphism induced by $w_1(X)$. Recall that elements of $H_1(X;o_X)$ are represented by normally oriented links L. This means that $TX|_L$ must be orientable for all such L. Then, by the definition of $w_1(X)$, we deduce $(w_1\circ r)([L])=0$ for all such L. Therefore $\ker(r^*)=\langle w_1(X)\rangle$, and as r^* is an epimorphism, we conclude the statement. □

The last ingredient allows us to draw a connection to the group action by $H^1(X;\mathbb{Z}_2)$ upon $\mathcal{P}\mathrm{in}(X)$. In particular, Theorem 1.2 is an immediate consequence of the following equivariant property:

Theorem 6.4 *The following diagram of obvious maps commutes*:

$$\begin{array}{ccc}\mathcal{P}\mathrm{in}(X)\times H^1(X;\mathbb{Z}_2) & \xrightarrow{\text{act}} & \mathcal{P}\mathrm{in}(X)\\ \downarrow & & \downarrow\\ \mathcal{S}\mathrm{plit}(X)\times\big(H^1(X;\mathbb{Z}_2)/\langle w_1(X)\rangle\big) & \xrightarrow{\text{act}} & \mathcal{S}\mathrm{plit}(X)\end{array}$$

Proof We fix a Pin^--structure on X. Let $\kappa\colon\mathbb{F}_1(X)\to\mathbb{Z}_2$ be the corresponding induced splitting map. Suppose $\gamma\in H^1(X;\mathbb{Z}_2)/\langle w_1(X)\rangle$, and let $\bar\gamma\in H^1(X;\mathbb{Z}_2)$ be a lift of γ. Moreover assume (C,φ) is a normally framed circle in X. When we identify $H^1(X;\mathbb{Z}_2)\cong\mathrm{Hom}(H_1(X;\mathbb{Z}_2),\mathbb{Z}_2)$, the action of $\bar\gamma$ on the given Pin^--structure can be described as follows:

$$\bar\gamma([C]_2)=\begin{cases}0 & \text{if and only if } \bar\gamma \text{ does } \textit{not} \text{ change the } \mathrm{Pin}^-\text{-structure on } TX|_C,\\ 1 & \text{if and only if } \bar\gamma \text{ changes the } \mathrm{Pin}^-\text{-structure on } TX|_C.\end{cases} \tag{6-1}$$

Because $TX|_C$ is orientable, we have $w_1([C]_2) = 0$. This implies that $\bar{\gamma}([C]_2)$ does not depend on the choice of lift. With the identification in Lemma 6.3, we conclude

$$(\kappa \cdot \gamma)([C,\varphi]) = \kappa([C,\varphi]) + (\gamma \circ h)([C,\varphi]) = \kappa([C,\varphi]) + \bar{\gamma}([C]_2). \tag{6-2}$$

The statement follows from substituting (6-1) into (6-2). □

7 Application to vector bundles

By obstruction theory, a choice of Spin-structure over an oriented vector bundle of rank 3 or higher is equivalent to a choice of framing over the 1-skeleton that extends over the 2-skeleton. This leads to the following lemma:

Lemma 7.1 *Let B be a 1-dimensional cell complex and $E \to B$ be an oriented spin vector bundle of rank $n \geq 3$. Then E is isomorphic to the trivial bundle and a* Spin*-structure uniquely determines a framing of E.*

A geometric proof of the above statement can be found in [9, Lemma 3.1]. For the rest of this section, we assume that $n \geq 4$. Suppose $E \to X$ is an oriented spin vector bundle of rank n, and we fix a Spin-structure over E. Let $s\colon X \to E$ be a global section, and let $0_E\colon X \hookrightarrow E$ be the zero section. We may assume that s intersects 0_E transversally, and we denote $s^{-1}(0_E) := L$. As the intersection is transverse, the differential $\mathrm{d}s\colon TX \to TE$ defines a canonical isomorphism of bundles $\nu_L \cong E|_L$. The bundle $E|_L$ inherits a Spin-structure from E. By Lemma 7.1 this induces a unique framing of $E|_L$ and therefore a normal framing φ of ν_L. This defines a framed bordism class $[L,\varphi] \in \mathbb{F}_1(X)$. The class $h([L,\varphi]) \in H_1(X; o_X)$ is the (twisted) Poincaré dual to the Euler class $e(E) \in H^n(X;\mathbb{Z})$. In other words, the class $[L,\varphi]$ is a refinement of the Euler class $e(E)$. We will show that this class is a *full* obstruction to finding a nonvanishing section on E.

Lemma 7.2 *The class $[L,\varphi]$ does not depend on the choice of section s up to homotopy.*

Proof Suppose that $s'\colon X \to E$ is a section transverse to 0_E that is homotopic (through sections) to s. That is, there is a homotopy $\bar{s}\colon X \times [0,1] \to E$ with $\bar{s}(\cdot,0) = s$ and $\bar{s}(\cdot,1) = s'$. Consider the projection $\mathrm{pr}_1 : X \times [0,1] \to X$ and denote $\overline{E} = \mathrm{pr}_1^* E$. Then $\bar{s}$ is a section of $\overline{E}$ and without loss of generality, we may assume that $\bar{s}$ is transverse to $0_{\overline{E}}$. Denote the zero locus of s' by $L' := (s')^{-1}(0_E)$ and the induced normal framing by φ'. Without loss of generality, we assume that the submanifold $\Sigma := \bar{s}^{-1}(0_{\overline{E}}) \subset X \times [0,1]$ has no connected components without boundary. Then the boundary of Σ is precisely L and L' and $\Sigma \simeq$ 1-dimensional cell complex. The restricted bundle $\overline{E}|_\Sigma$ inherits a Spin-structure from E, and by Lemma 7.1, this gives a unique framing of $\overline{E}|_\Sigma$. This framing induces a normal framing over Σ such that (L,φ) is normally framed to (L',φ'). □

Lemma 7.3 *The class $[L,\varphi]$ does not depend on the choice of the* Spin*-structure over E if $w_n(E) = 0$.*

Proof To begin, let us observe that $\mathrm{PD}_2([L]_2) = w_n(E)$. Consequently, we can deduce that $w_n(E) = 0$ if and only if $[L]_2 = 0$. A change of Spin-structure corresponds to a homomorphism $\pi_1(X) \to \mathbb{Z}_2$, ie an element in $H^1(X;\mathbb{Z}_2)$. That homomorphism takes the homotopy class of a circle C and "twists" the framing over C by an element in $\pi_1(\mathrm{SO}(n)) \cong \mathbb{Z}_2$. However, given that $[L]_2 = 0$, any such change does not impact the framing over L. □

Let W be a connected closed $(n+k)$-dimensional manifold with connected boundary ∂W, where $k \geq 1$. Consider a smooth map $f\colon W \to \mathbb{R}^n$ for which zero is a regular value and $f^{-1}(0) \cap \partial W = \varnothing$. Then $M := f^{-1}(0)$ is a closed submanifold of $\mathrm{int}(W)$. The manifold M comes equipped with an induced normal framing φ. Note that (M,φ) defines an element $[M,\varphi]$ in the *relative* framed bordism group $\mathbb{F}_k(W,\partial W)$.[5] Now, we set

$$g\colon \partial W \to S^{n-1} \quad \text{where } g(x) := \frac{f(x)}{|f(x)|}.$$

Lemma 7.4 *The map g extends over W if and only if $[M,\varphi] = 0$ in $\mathbb{F}_k(W,\partial W)$.*

Proof To begin the proof, we observe that the map $f\colon W \to \mathbb{R}^n$, viewed as a map into the 1-point compactification $\mathbb{R}^n \subset \mathbb{R}^n \cup \{\infty\} = S^n$, can be homotoped to a map $f_{\mathrm{rel}}\colon (W,\partial W) \to (S^n,\infty)$ by radially extending it to infinity in a collar of ∂W. The Pontryagin–Thom construction gives an isomorphism

$$\mathbb{F}_k(W,\partial W) \cong [(W,\partial W),(S^n,\infty)].$$

This isomorphism associates the bordism class $[M,\varphi]$ with the homotopy class $[f_{\mathrm{rel}}]$. When we apply the Puppe sequence to the inclusions

$$\iota\colon \partial W \hookrightarrow W, \quad \jmath\colon (W,\varnothing) \hookrightarrow (W,\partial W),$$

we obtain a (long) exact sequence in relative cohomotopy sets,

$$[\mathrm{S}W,S^n] \xrightarrow{\iota^*} [\mathrm{S}\partial W,S^n] \xrightarrow{\delta} [(W,\partial W),(S^n,\infty)] \xrightarrow{\jmath^*} [W,S^n], \tag{7-1}$$

where S denotes the unreduced suspension functor. Since we assumed $n \geq 4$, we are in the stable range for the suspension and obtain

$$[\mathrm{S}W,S^n] \cong [W,S^{n-1}] \quad \text{and} \quad [\mathrm{S}\partial W,S^n] \cong [\partial W,S^{n-1}].$$

To understand the boundary map δ in (7-1), we consider the suspended map $\mathrm{S}g\colon \mathrm{S}\partial W \to S^n$. Using the inclusion of a collar $\partial W \times [0,1] \hookrightarrow W$, we can extend $\mathrm{S}g$ onto W by assigning it a constant value outside the collar, which produces a map $f'_{\mathrm{rel}}\colon (W,\partial W) \to (S^n,\infty)$. By construction, the map f'_{rel} is homotopic to f_{rel} relative to ∂W. Consequently, we have $\delta[g] = [f_{\mathrm{rel}}]$. The statement now follows from the exactness of (7-1). □

We are ready to prove the main theorem of this section that generalizes the invariant introduced by Konstantis [9].

[5] Here, $\mathbb{F}_k(W,\partial W)$ is the set of normally framed closed k-dimensional submanifolds of $\mathrm{int}(W)$ up to framed bordism.

Theorem 7.5 *Let $E \to X$ be an oriented vector bundle of rank $n \geq 4$ with a* Spin*-structure. Moreover, denote by (L, φ) the zero locus of a section transverse to the zero section together with the normal framing induced by the* Spin*-structure. Then E admits a nonvanishing section if and only if $[L, \varphi] = 0$ in $\mathbb{F}_1(X)$.*

Proof Suppose E admits a nonvanishing section s'. Then the zero locus of s' is the empty set, and because of Lemma 7.2, we get $[L, \varphi] = 0$.

For the converse, assume that $[L, \varphi] = 0$ in $\mathbb{F}_1(X)$. Let $\Sigma \subset X \times [0, 1]$ be the associated framed null-bordism. Notice that Σ may not be connected, ie $\Sigma = \Sigma_1 \sqcup \cdots \sqcup \Sigma_m$. Without loss of generality, we assume that each connected component Σ_k has a nonempty boundary. We isotope each Σ_k relative to $\partial\Sigma_k$ to a submanifold $\iota_k \colon \Sigma_k \hookrightarrow X$. There, we attach a closed collar, ie we extend ι_k to an embedding

$$\Sigma_k \cup_{\partial\Sigma_k} (\partial\Sigma_k \times [0, 1]) \hookrightarrow X,$$

and consider the tubular neighborhood of this embedding. The corresponding disk bundle gives us a compact manifold with corners. We can smoothen the corners to produce a compact manifold W_k with connected and smooth boundary ∂W_k. By construction, we have $[\partial\Sigma_k, \varphi_k] = 0$ in $\mathbb{F}_1(W_k, \partial W_k)$, where φ_k is the restriction of φ to $\partial\Sigma_k$. As W_k is homotopy equivalent to Σ_k, we have $E|_{W_k} \cong W_k \times \mathbb{R}^n$. If we restrict s to W_k, we get an induced map $f_k \colon W_k \to \mathbb{R}^n$ whose zero locus is precisely $\partial\Sigma_k$. We can now apply Lemma 7.4 to deform s into a section that has no zeroes in each W_k. □

Remark 7.6 The proof above can be interpreted as a version of the Whitney trick. To see this, we apply the same trick if the rank of E is equal to the dimension of X, namely $n+1$. In that case, the null-bordism is given by a bunch of arcs that isotope to arcs in X. We concatenate these arcs with their corresponding lifts along the section s and obtain closed curves in the total space E that bound contractible disks. These disks can now be used to "push" s from 0_E in a neighborhood of each arc.

Now, we would like to relate Theorem 7.5 to type I and type II manifolds. If X is a manifold of type I, the map $h \colon \mathbb{F}_1(X) \to H_1(X; o_X)$ is an isomorphism. Therefore, Theorem 1.3 is an immediate consequence of Theorem 7.5. To obtain a well-defined invariant in the case of type IIa manifolds, we require the next lemma.

Lemma 7.7 *Suppose that X is of type IIa, and choose a* Pin^-*-structure over X. We denote by $\kappa \colon \mathbb{F}_1(X) \to \mathbb{Z}_2$ the induced splitting map. Moreover, let E and (L, φ) be as in Theorem 7.5. If $w_n(E) = 0$, then $\kappa([L, \varphi])$ does not depend on the chosen* Pin^-*-structure.*

Proof The proof is similar to the proof of Lemma 7.3 and follows from Theorem 6.4. If $w_n(E) = 0$, we obtain from (6-2) that $\kappa([L, \varphi])$ remains unchanged for any choice of Pin^--structure. □

Together with Lemma 7.3, this allows us to define *the degree of E*,

$$\kappa(E) := \kappa([L, \varphi]),$$

provided that $w_n(E) = 0$. If X is oriented, this quantity is precisely the counting invariant introduced by Konstantis [9]. In the same spirit, we obtain Theorem 1.4, which is now a corollary of Theorem 7.5, bearing in mind that $w_n(E)$ is the mod 2 reduction of $e(E)$.

References

[1] **M F Atiyah**, *Bordism and cobordism*, Proc. Cambridge Philos. Soc. 57 (1961) 200–208 MR Zbl

[2] **J M Cohen**, **H Gluck**, *Stacked bases for modules over principal ideal domains*, J. Algebra 14 (1970) 493–505 MR Zbl

[3] **J F Davis**, **P Kirk**, *Lecture notes in algebraic topology*, Graduate Studies in Mathematics 35, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[4] **P J Eccles**, **M Grant**, *Self-intersections of immersions and Steenrod operations*, Acta Math. Hungar. 137 (2012) 272–281 MR Zbl

[5] **S Eilenberg**, **S MacLane**, *Group extensions and homology*, Ann. of Math. 43 (1942) 757–831 MR Zbl

[6] **A Hatcher**, *Algebraic topology*, Cambridge Univ. Press (2002) MR Zbl

[7] **R Kirby**, **P Melvin**, **P Teichner**, *Cohomotopy sets of 4-manifolds*, from "Proceedings of the Freedman Fest" (V Krushkal, Z Wang, editors), Geom. Topol. Monogr. 18, Geom. Topol. Publ., Coventry (2012) 161–190 MR Zbl

[8] **R C Kirby**, **L R Taylor**, Pin *structures on low-dimensional manifolds*, from "Geometry of low-dimensional manifolds, II" (S K Donaldson, C B Thomas, editors), London Math. Soc. Lecture Note Ser. 151, Cambridge Univ. Press (1990) 177–242 MR Zbl

[9] **P Konstantis**, *A counting invariant for maps into spheres and for zero loci of sections of vector bundles*, Abh. Math. Semin. Univ. Hambg. 90 (2020) 183–199 MR Zbl

[10] **L L Larmore**, **E Thomas**, *Group extensions and principal fibrations*, Math. Scand. 30 (1972) 227–248 MR Zbl

[11] **J W Milnor**, **J D Stasheff**, *Characteristic classes*, Annals of Mathematics Studies 76, Princeton Univ. Press (1974) MR Zbl

[12] **J R Munkres**, *Elements of algebraic topology*, Addison-Wesley, Menlo Park, CA (1984) MR Zbl

[13] **E Spanier**, *Borsuk's cohomotopy groups*, Ann. of Math. 50 (1949) 203–245 MR

[14] **E H Spanier**, *Algebraic topology*, Springer (1995) MR Zbl

[15] **N E Steenrod**, *Products of cocycles and extensions of mappings*, Ann. of Math. 48 (1947) 290–320 MR

[16] **L R Taylor**, *The principal fibration sequence and the second cohomotopy set*, from "Proceedings of the Freedman Fest" (V Krushkal, Z Wang, editors), Geom. Topol. Monogr. 18, Geom. Topol. Publ., Coventry (2012) 235–251 MR

[17] **R Thom**, *Variétés plongées et i-carrés*, C. R. Acad. Sci. Paris 230 (1950) 507–508 MR Zbl

[18] **R Thom**, *Quelques propriétés globales des variétés différentiables*, Comment. Math. Helv. 28 (1954) 17–86 MR Zbl

Department of Mathematics, Vrije Universiteit Amsterdam
Amsterdam, Netherlands

m.jung@vu.nl, t.o.rot@vu.nl

Received: 9 August 2023 Revised: 25 March 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3627–3677
DOI: 10.2140/agt.2025.25.3627
Published: 1 October 2025

Calabi–Yau structure on the Chekanov–Eliashberg algebra of a Legendrian sphere

NOÉMIE LEGOUT

We prove that the Chekanov–Eliashberg algebra of a horizontally displaceable n-dimensional Legendrian sphere in the contactization of a Liouville manifold is an $(n+1)$-Calabi–Yau differential graded algebra. In particular it means that there is a quasi-isomorphism of DG bimodules between the diagonal bimodule and the inverse dualizing bimodule associated to the Chekanov–Eliashberg algebra. On some cyclic version of these bimodules, computing the Hochschild homology and cohomology of the Chekanov–Eliashberg algebra, we construct A_∞-operations and show that the Calabi–Yau isomorphism extends to a family of maps satisfying the A_∞-functor equations.

53D42, 57R58

1. Introduction 3627
2. Background on DG bimodules 3630
3. Moduli spaces 3632
4. The Chekanov–Eliashberg DGA 3635
5. The Rabinowitz DG bimodule 3636
6. The Calabi–Yau structure 3638
7. Products and higher-order structure maps 3652
8. Example: the unknot 3674
References 3676

1 Introduction

We consider n-dimensional Legendrian submanifolds in a contact manifold (Y,α) which is the contactization of a Liouville manifold (P,β). Among the numerous invariants of Legendrians up to Legendrian isotopy, many of them are derived from the famous Chekanov–Eliashberg algebra (C-E algebra) [8; 22], originally defined by Chekanov for Legendrian links in $\mathbb{R}^3$ and generalized to higher-dimensional Legendrians in the contactization of a Liouville domain by Ekholm, Etnyre and Sullivan [19; 20]. This algebra is a unital differential graded algebra (DGA) associated to a Legendrian Λ and generated by Reeb chords of Λ; see Section 4 for a brief recall of the definition. In this paper we take the coefficient ring to be the field $\mathbb{Z}_2$.

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

It is well known that linearized versions of the C-E algebra satisfy a particular type of duality, which was first proved by Sabloff [33] for Legendrian knots, and generalized to higher dimensions by Ekholm, Etnyre and Sabloff [18] and to the bilinearized case by Bourgeois and Chantraine [1]. We prove in this paper that under some assumptions the full C-E algebra, ie not linearized, also satisfies a similar type of duality. This duality is expressed in terms of differential graded bimodules (DG bimodules). Namely, if $\mathcal{A}$ denotes the C-E algebra of a n-dimensional Legendrian sphere, then $\mathcal{A}$ is a DG $\mathcal{A}$-bimodule called the *diagonal bimodule*, and the *inverse dualizing bimodule* of $\mathcal{A}$ is the DG $\mathcal{A}$-bimodule $\mathcal{A}^! := \mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A}\otimes\mathcal{A})$. We prove that there is a quasi-isomorphism

$$\mathcal{CY}\colon \mathcal{A} \xrightarrow{\simeq} \mathcal{A}^![-n-1] \tag{1}$$

satisfying $\mathcal{CY} \simeq \mathcal{CY}^![-n-1]$. Together with the fact that $\mathcal{A}$ is a *homologically smooth* differential graded algebra (see Section 2), this leads to the following result:

Theorem 1.1 *Let $\Lambda \subset Y$ be an horizontally displaceable Legendrian sphere. Then its C-E-algebra is a $(n+1)$-Calabi–Yau differential graded algebra.*

The definition of Calabi–Yau structure we use here is the one of Ginzburg [26, Definition 3.2.3] (see also Kontsevich and Soibelman [28]), although we have some opposite sign convention for degrees.

For the following reasons, it was reasonable to expect that the C-E algebra of a displaceable Legendrian sphere admits a Calabi–Yau structure. In his thesis, Ganatra [24] showed that the wrapped Fukaya category $\mathcal{W}$ of a Weinstein manifold is a noncompact Calabi–Yau category, where the Calabi–Yau structure is induced similarly as for (1) by an equivalence of A_∞-bimodules between the diagonal bimodule $\mathcal{W}$ and an analogue of the inverse dualizing bimodule in the A_∞-setting, $\mathcal{W}^!$. Now, observe that the wrapped Fukaya category of a Weinstein manifold X (of finite type) is generated by the Lagrangian cocores obtained by attaching critical handles to Legendrian spheres in the ideal contact boundary of X, as proved by Chantraine, Dimitroglou Rizell, Ghiggini and Golovko [7], and Ganatra, Pardon and Shende [25]. A Calabi–Yau structure on the C-E algebra of these Legendrian spheres should then be possible to construct by using the surgery isomorphism of Bourgeois, Ekholm and Eliashberg [3] (see also Ekholm and Lekili [21] and Ekholm [17]). Recall that the surgery isomorphism gives an A_∞-quasi-isomorphism between the wrapped Floer cohomology of the cocores, ie the endomorphism groups of the generators of the wrapped Fukaya category, and the C-E algebra of the Legendrian attaching spheres. Note that X is taken here to be a subcritical Weinstein manifold. When X is the standard ball, the C-E algebra of a Legendrian in its boundary can be computed inside a contact Darboux ball where it is displaceable; see, for example, the work of Dimitroglou Rizell, Ekholm and Tonkonog [14].

This suggests that a Calabi–Yau structure on the C-E algebra of a Legendrian sphere could potentially be obtained from the one defined by Ganatra via the surgery isomorphism. However we will not adopt this method in this paper. Instead, we restrict ourselves to Legendrian spheres in the contactization of a Liouville manifold and introduce a version of the Rabinowitz Floer homology for Legendrians with coefficients in the free $\mathcal{A}$-bimodule $\mathcal{A}\otimes\mathcal{A}$ of rank 1, where $\mathcal{A}$ denotes the C-E algebra.

Rabinowitz Floer homology was originally defined as an homology theory for contact-type hypersurfaces by Cieliebak and Frauenfelder in [9]. A relative theory, Lagrangian Rabinowitz Floer homology, has then been introduced in [31] for exact Lagrangians in a Liouville manifold, and more recently both the nonrelative and relative theories have been generalized to the case of Liouville cobordisms admitting a filling by Cieliebak and Oancea in [10]. These previous theories were defined in the Hamiltonian setting and in [29] we introduced an SFT-type version of the Lagrangian Rabinowitz Floer homology for Lagrangians in a (trivial) Liouville cobordism. There, we used augmentations of the C-E algebras of Legendrians in the negative ends of Lagrangian cobordisms in order to define a complex over $\mathbb{Z}_2$. In this paper, we consider only the Rabinowitz complex of cylinders over a Legendrian submanifold and we don't use augmentations, so that we get a DG bimodule with coefficients in the C-E algebras of the Legendrians.

More precisely, the Rabinowitz DG bimodule considered here is generated by mixed chords of a 2-copy $\Lambda_0 \cup \Lambda_1$ of a Legendrian Λ, where Λ_1 is a small negative push-off of $\Lambda_0 := \Lambda$, and its differential is defined by a count of pseudoholomorphic discs with boundary on $\mathbb{R} \times (\Lambda_0 \cup \Lambda_1)$. The differential is lower triangular and so the Rabinowitz bimodule is the cone of a DG bimodule map $C_+(\Lambda_0, \Lambda_1) \to C_-(\Lambda_0, \Lambda_1)$, where C_+ is generated by chords from Λ_0 to Λ_1 (in bijective correspondence with chords of Λ) and C_- is generated by chords from Λ_1 to Λ_0 (in bijective correspondence with chords of Λ and critical points of a Morse function on Λ). For a Legendrian sphere Λ and its C-E algebra $\mathcal{A}$, the Rabinowitz bimodule can be described as the cone of a slightly different DG bimodule map

$$\mathrm{CY}\colon \widehat{C}_+(\Lambda_0, \Lambda_1) \to \check{C}_-(\Lambda_0, \Lambda_1),$$

which corresponds to a slight modification of the action filtration. Here $\widehat{C}_+(\Lambda_0, \Lambda_1)$ is generated by Reeb chords from Λ_0 to Λ_1 as well as the maximum of a given Morse function on Λ, and $\check{C}_-(\Lambda_0, \Lambda_1)$ is generated by chords from Λ_1 to Λ_0 and the minimum of the Morse function. Then, we show that $\widehat{C}_+(\Lambda_0, \Lambda_1)[-n-1]$ is quasi-isomorphic to $\mathcal{A}$ and $\check{C}_-(\Lambda_0, \Lambda_1)$ is quasi-isomorphic to $\mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A}\otimes\mathcal{A})$. The invariance of the Rabinowitz homology up to Legendrian isotopies implies that if Λ is horizontally displaceable, then the complex is acyclic. The shifted map $\mathrm{CY}[-n-1]$ provides thus a Calabi–Yau quasi-isomorphism $\mathcal{A} \to \mathcal{A}^![-n-1]$.

In general, we expect that the Rabinowitz complex is acyclic for any Legendrian in the contact boundary of a subcritical Weinstein manifold. For example the periodic orbit version of Rabinowitz Floer homology and Rabinowitz Floer homology for fillable Legendrians both vanish there; see [10]. We expect also that the Calabi–Yau structure we construct here coincides with that constructed by Ganatra in [24], and plan to show this in future work with Asplund. One advantage with the perspective taken here compared to that of Ganatra is from the point of view of computability. The pseudoholomorphic discs that define the operations we consider can be computed using Ekholm's theory of gradient flow trees [15], while the operations defined by Hamiltonian perturbations seem to lack general techniques for computation.

By taking bimodule tensor products of both bimodules $\widehat{C}_+(\Lambda_0, \Lambda_1)$ and $\check{C}_-(\Lambda_0, \Lambda_1)$ with the diagonal bimodule $\mathcal{A}$ one gets complexes which we denote by $\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ and $\check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ and which compute

the Hochschild homology and cohomology of $\mathcal{A}$, respectively. On these complexes we construct higher-order maps, ie for any $(d+1)$-copy $\Lambda_0 \cup \cdots \cup \Lambda_d$ we construct maps

$$\widehat{\mathfrak{m}}_d \colon \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_d),$$
$$\check{\mathfrak{m}}_d \colon \check{C}_-^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_d)$$

satisfying the A_∞-equations; see Section 7. The maps $\check{\mathfrak{m}}_d$ are defined by a count of pseudoholomorphic discs with boundary on $\mathbb{R} \times (\Lambda_0 \cup \cdots \cup \Lambda_d)$, with d negative asymptotics which are inputs and one positive asymptotic which is the output. These maps are well known: they compute the A_∞-structure of the augmentation category $\mathcal{A}\mathrm{ug}_-(\Lambda)$ of Bourgeois and Chantraine [1] with a formal unit added (corresponding to the minimum of the Morse function here). One the other side, the maps $\widehat{\mathfrak{m}}_d$ are defined by a count of certain 2-level pseudoholomorphic buildings appearing in the boundary of the compactification of 1-dimensional moduli spaces. In this sense it can be seen as a *secondary-type* product. Finally, we show that the map CY induced on the cyclic complexes extends to a family of maps

$$\mathrm{CY}_d \colon \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_d)$$

satisfying the A_∞-functor equations.

Observe that the map CY gives an isomorphism between the Hochschild homology and cohomology of $\mathcal{A}$, which after generalizing our definition of the Rabinowitz complex to Legendrians in more general contact manifolds, would recover the quasi-isomorphism between Hochschild homology and cohomology for the wrapped Fukaya category in [24]. Moreover we presumably recover the relation between the different product structures as constructed by Bourgeois, Ekholm and Eliashberg in [2].

Acknowledgments I am very grateful to Georgios Dimitroglou Rizell for all the discussions we had together and his generosity in sharing ideas. I also thank Paolo Ghiggini, who gave me valuable comments on a first version of the paper. In particular he pointed out a remark which I was using in some proof but which was wrong. Finally I thank the referee, whose comments helped me to correct more mistakes. This work is supported by the Knut and Alice Wallenberg Foundation under the grants KAW 2021.0191 and KAW 2021.0300, and by the Swedish Research Council under the grant number 2020-04426.

2 Background on DG bimodules

Let $(\mathcal{A}_0, \partial_{\mathcal{A}_0})$, $(\mathcal{A}_1, \partial_{\mathcal{A}_1})$ be unital differential graded algebras (DGAs) over $\mathbb{Z}_2$ (we restrict to $\mathbb{Z}_2$ for simplicity here but it is not strictly necessary). A *DG $\mathcal{A}_1$-$\mathcal{A}_0$-bimodule* is a graded $\mathcal{A}_1$-$\mathcal{A}_0$-bimodule $\mathcal{B}$ endowed with a degree-1 differential $\partial_{\mathcal{B}}$ such that

$$\partial_{\mathcal{B}}(\boldsymbol{\alpha}_1 b \boldsymbol{\alpha}_0) = \partial_{\mathcal{A}_1}(\boldsymbol{\alpha}_1) b \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 \partial_{\mathcal{B}}(b) \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 b \partial_{\mathcal{A}_0}(\boldsymbol{\alpha}_0).$$

If $(\mathcal{A}_0, \partial_{\mathcal{A}_0}) = (\mathcal{A}_1, \partial_{\mathcal{A}_1}) = (\mathcal{A}, \partial_{\mathcal{A}})$, we write simply $\mathcal{A}$-bimodule instead of $\mathcal{A}$-$\mathcal{A}$-bimodule.

Example 2.1 (a) A DGA $(\mathcal{A}, \partial_{\mathcal{A}})$ is itself a DG $\mathcal{A}$-bimodule, called the *diagonal bimodule*.

(b) The tensor product of a DGA $\mathcal{A}$ with itself over $\mathbb{Z}_2$ is also a DG $\mathcal{A}$-bimodule, with differential given by $\partial_{\mathcal{A}\otimes\mathcal{A}}(a_1 \otimes a_0) = \partial_{\mathcal{A}}(a_1) \otimes a_0 + a_1 \otimes \partial_{\mathcal{A}}(a_0)$. This bimodule carries two different bimodule structures, the *outer* and the *inner* bimodule structure:

$$\mu^{\text{out}}_{\mathcal{A}\otimes\mathcal{A}} \colon \mathcal{A} \otimes (\mathcal{A} \otimes \mathcal{A}) \otimes \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}, \quad \boldsymbol{\alpha} \otimes (a \otimes a') \otimes \boldsymbol{\alpha}' \mapsto \boldsymbol{\alpha} a \otimes a' \boldsymbol{\alpha}',$$
$$\mu^{\text{in}}_{\mathcal{A}\otimes\mathcal{A}} \colon \mathcal{A} \otimes (\mathcal{A} \otimes \mathcal{A}) \otimes \mathcal{A} \to \mathcal{A} \otimes \mathcal{A}, \quad \boldsymbol{\alpha} \otimes (a \otimes a') \otimes \boldsymbol{\alpha}' \mapsto a \boldsymbol{\alpha} \otimes \boldsymbol{\alpha}' a'.$$

(c) Given two DG $\mathcal{A}_1$-$\mathcal{A}_0$-bimodules $\mathcal{B}$ and $\mathcal{C}$, the set $\operatorname{Hom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\mathcal{B}, \mathcal{C})$ of bimodule maps is a chain complex whose differential is given by $D(\phi) = \phi \circ \partial_{\mathcal{B}} + \partial_{\mathcal{C}} \circ \phi$.

In this paper, we write $-\otimes-$ instead of $-\otimes_{\mathbb{Z}_2}-$ for the tensor product over $\mathbb{Z}_2$. A *DG morphism* of $\mathcal{A}_1$-$\mathcal{A}_0$-bimodules from $\mathcal{B}$ to $\mathcal{C}$ is a degree-0 element ϕ in $\operatorname{Hom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\mathcal{B}, \mathcal{C})$ which commutes with the differentials of $\mathcal{B}$ and $\mathcal{C}$; in other words, it is a degree-0 cycle in $(\operatorname{Hom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\mathcal{B}, \mathcal{C}), D)$. A *quasi-isomorphism* of DG bimodules is a DG morphism which induces an isomorphism in homology.

Following [23], we recall that a DG $\mathcal{A}_1$-$\mathcal{A}_0$-bimodule $(\mathcal{B}, \partial_{\mathcal{B}})$ is *free* if it is isomorphic to $\mathcal{A}_1 \otimes V \otimes \mathcal{A}_0$ where V is a $\mathbb{Z}_2$-vector space generated by cycles. In this case a *free generating set* for $\mathcal{B}$ is a basis of V. Given a DGA $\mathcal{A}$ the diagonal bimodule is not a free bimodule while $\mathcal{A} \otimes \mathcal{A}$ endowed with either the outer or the inner structure is free, generated by $1 \otimes 1$. A DG bimodule $\mathcal{B}$ is called *semifree* if there is a filtration

$$\{0\} \subset F_0\mathcal{B} \subset F_1\mathcal{B} \subset \cdots \subset \mathcal{B}$$

such that $F_i\mathcal{B}$ is a DG subbimodule for all $i \geq 0$, $\bigcup F_i\mathcal{B} = \mathcal{B}$, and $F_0\mathcal{B}$ and $F_{i+1}\mathcal{B}/F_i\mathcal{B}$ are free bimodules. We say that a semifree DG bimodule $\mathcal{B}$ as above is of *finite rank* if there is a $k \in \mathbb{N}$ such that $F_k\mathcal{B} = \mathcal{B}$. A *semifree resolution* of a DG bimodule $\mathcal{B}$ is a semifree DG bimodule $R_{\mathcal{B}}$ together with a quasi-isomorphism of DG bimodules $R_{\mathcal{B}} \to \mathcal{B}$.

In the category of DG $\mathcal{A}_1$-$\mathcal{A}_0$-bimodules, $\operatorname{RHom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(-, \mathcal{C})$ denotes the right derived functor of the functor $\operatorname{Hom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(-, \mathcal{C})$. Let $R_{\mathcal{B}}$ be any semifree resolution of the $\mathcal{A}_1$-$\mathcal{A}_0$-bimodule $\mathcal{B}$; we have by definition that $\operatorname{RHom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\mathcal{B}, \mathcal{C}) = \operatorname{Hom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(R_{\mathcal{B}}, \mathcal{C})$, which is well defined up to quasi-isomorphism. Given a DGA $\mathcal{A}$, the *inverse dualizing bimodule*, denoted by $\mathcal{A}^!$, is $\operatorname{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A} \otimes \mathcal{A})$. The elements of $\operatorname{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A} \otimes \mathcal{A})$ are bimodule morphisms from the diagonal bimodule $\mathcal{A}$ to $\mathcal{A} \otimes \mathcal{A}$ where $\mathcal{A} \otimes \mathcal{A}$ is endowed with the inner bimodule structure. Then, $\operatorname{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A} \otimes \mathcal{A})$ is a DG $\mathcal{A}$-bimodule with a bimodule structure induced by the outer bimodule structure on $\mathcal{A} \otimes \mathcal{A}$, ie for $\phi \in \operatorname{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A} \otimes \mathcal{A})$ which can be written as $\phi = \phi_l \otimes \phi_r$ and for any $\boldsymbol{v}, \boldsymbol{a}, \boldsymbol{w} \in \mathcal{A}$ we have

$$(\boldsymbol{v} \cdot \phi \cdot \boldsymbol{w})(\boldsymbol{a}) = \boldsymbol{v}\phi_l(\boldsymbol{a}) \otimes \phi_r(\boldsymbol{a})\boldsymbol{w}.$$

If $\phi \colon \mathcal{B} \to \mathcal{C}$ is a DG morphism of $\mathcal{A}$-bimodules in the derived sense, then there is an induced morphism $\phi^! \colon \mathcal{C}^! \to \mathcal{B}^!$ given by $\phi^!(c^!) = c^! \circ \phi$.

Remark 2.2 The semifree resolution mentioned above in order to compute $\mathrm{RHom}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\mathcal{B},\mathcal{C})$ can be taken to be the bar resolution of $\mathcal{B}$; however this one is not of finite rank. In this paper we will be interested in the inverse dualizing bimodule $\mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A},\mathcal{A}\otimes\mathcal{A})$ of the Chekanov–Eliashberg DGA (see Section 4) of a closed Legendrian submanifold. In this case, the diagonal bimodule $\mathcal{A}$ admits a finite-rank semifree resolution; see Section 6.4.

The purpose of this paper is to prove that under some hypothesis the Chekanov–Eliashberg algebra of a Legendrian sphere is a Calabi–Yau DGA. Let us recall the necessary definition.

Definition 2.3 [28] Let $(\mathcal{A},\partial_{\mathcal{A}})$ be a DGA.

(a) A DG $\mathcal{A}$-bimodule is *perfect* if it is quasi-isomorphic to a direct summand of a finite-dimensional semifree DG bimodule.

(b) The DGA $\mathcal{A}$ is *homologically smooth* if it is perfect as a DG $\mathcal{A}$-bimodule.

Definition 2.4 [26] A homologically smooth DGA $\mathcal{A}$ is called *d-Calabi–Yau* if there is a quasi-isomorphism of DG $\mathcal{A}$-bimodules

$$\phi\colon \mathcal{A}\to\mathcal{A}^!{}[-d]$$

such that $\phi\simeq\phi^![-d]$.

Notation 2.5 The convention we use for shifts in this paper is that if $|a|$ is the degree of an element of a DG bimodule $\mathcal{B}$, then the same element viewed in $\mathcal{B}[d]$ has degree $|a|+d$. Then, for a DG bimodule morphism $f\colon\mathcal{B}\to\mathcal{C}$ we denote by $f[d]\colon\mathcal{B}[d]\to\mathcal{C}[d]$ the shifted map.

3 Moduli spaces

We will be working in the same setting as in [29] and refer to Sections 2.2–2.6 of the mentioned paper for more details about the moduli spaces of pseudoholomorphic discs we consider here. Throughout this paper, when we consider a Legendrian submanifold of Y we always assume that it is closed and nondegenerate in the sense that it admits a finite number of Reeb chords which are isolated and correspond to transverse intersection points of the Lagrangian projection on P.

Let $\Lambda\subset Y$ be a Legendrian and denote by $\mathcal{R}(\Lambda)$ the set of Reeb chords of Λ. For any chord $\gamma\in\mathcal{R}(\Lambda)$ we denote by $\mathrm{CZ}(\gamma)$ its Conley–Zehnder index; see [19].

Remark 3.1 When Λ is not connected, there are additional choices (as, for example, choices of paths between the various connected components) to make in order to define the Conley–Zehnder index of a Reeb chord connecting two distinct connected components; see [13]. The index of a chord depends on these additional choices but for any two chords from a connected component to another, the difference in index is independent of the choices. Moreover, we assume in this paper that the Legendrians we consider always have Maslov number 0 and that the first Chern class of P vanishes. In this case we get a well-defined $\mathbb{Z}$-valued Conley–Zehnder index (after the potentially additional choices discussed above).

Let J_P be an almost complex structure on (P, β) which is compatible with $d\beta$ and cylindrical outside of a compact set in the cylindrical end of P. We call such a structure *admissible*. Then, we denote by J the almost complex structure on $(\mathbb{R} \times Y, d(e^t \alpha))$ which is the cylindrical lift of J_P, ie the unique cylindrical almost complex structure J on $\mathbb{R} \times Y$ such that the projection $\pi_P : \mathbb{R} \times P \times \mathbb{R} \to P$ is (J, J_P)-holomorphic.

Given Reeb chords $\gamma, \gamma_1, \dots, \gamma_d$ of Λ we denote by

$$\widehat{\mathcal{M}}_\Lambda(\gamma; \gamma_1, \dots, \gamma_d)$$

the moduli space of J-holomorphic discs with boundary on $\mathbb{R} \times \Lambda$ having a positive asymptotic at γ and negative asymptotics at $\gamma_1, \dots, \gamma_d$. As the boundary condition for these discs is cylindrical as well as the almost complex structure, there is an action of $\mathbb{R}$ by translation on these moduli spaces. We denote by

$$\mathcal{M}_\Lambda(\gamma; \gamma_1, \dots, \gamma_d) =: \widehat{\mathcal{M}}_\Lambda(\gamma; \gamma_1, \dots, \gamma_d)/\mathbb{R}$$

the quotient by this action.

Let $\Lambda_0, \dots, \Lambda_d \subset Y$ be $d+1$ Legendrian submanifolds such that the link $\Lambda_0 \cup \cdots \cup \Lambda_d$ is nondegenerate. For any $0 \leq i \neq j \leq d$, we denote by $\mathcal{R}(\Lambda_i, \Lambda_j)$ the set of Reeb chords from Λ_j to Λ_i. Such chords are called *mixed* while chords in $\mathcal{R}(\Lambda_i)$ are called *pure*. Let $\gamma_{0d} \in \mathcal{R}(\Lambda_0, \Lambda_d)$, $(\gamma_1, \dots, \gamma_d)$ be a d-tuple of Reeb chords such that $\gamma_i \in \mathcal{R}(\Lambda_{i-1}, \Lambda_i) \cup \mathcal{R}(\Lambda_i, \Lambda_{i-1})$, and $\boldsymbol{\delta}_i$ for $0 \leq i \leq d$ words of Reeb chords of Λ_i. We denote by

$$\mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{0d}; \boldsymbol{\delta}_0, \gamma_1, \boldsymbol{\delta}_1, \dots, \gamma_d, \boldsymbol{\delta}_d)$$

the quotient by the action of $\mathbb{R}$ of the moduli space of pseudoholomorphic discs satisfying the following:

- The boundary of the discs lie on the ordered $(d+1)$-tuple of Lagrangians $\mathbb{R} \times \Lambda_0, \dots, \mathbb{R} \times \Lambda_d$ when following the boundary counterclockwise.
- The discs in this moduli space are positively asymptotic to the Reeb chord γ_{0d}, positively or negatively asymptotic to the Reeb chords γ_i, and negatively asymptotic to the words of Reeb chords $\boldsymbol{\delta}_i$.

Similarly, for a chord $\gamma_{d0} \in \mathcal{R}(\Lambda_d, \Lambda_0)$, and other chords as above, we denote by

$$\mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{d0}; \boldsymbol{\delta}_0, \gamma_1, \boldsymbol{\delta}_1, \dots, \gamma_d, \boldsymbol{\delta}_d)$$

the quotient by the action of $\mathbb{R}$ of the moduli space of pseudoholomorphic discs with the same boundary conditions as above, negatively asymptotic to the Reeb chord γ_{d0}, and with the same asymptotic conditions as above for the other punctures.

Remark 3.2 In the notation we employ, the first asymptotic γ, γ_{0d} or γ_{d0} will always be the output of a map defined by a count of rigid discs in the corresponding moduli spaces. When the Lagrangian boundary conditions for the pseudoholomorphic discs is the ordered $(d+1)$-tuple $(\mathbb{R} \times \Lambda_0, \dots, \mathbb{R} \times \Lambda_d)$, knowing if the first asymptotic is a positive or negative asymptotic is enough to determine if the other mixed asymptotics are positive or negative, according to their direction. More precisely, if a chord γ_i is in $\mathcal{R}(\Lambda_i, \Lambda_{i-1})$ then it will always be a positive asymptotic while if it is in $\mathcal{R}(\Lambda_{i-1}, \Lambda_i)$ it will always be a negative asymptotic.

By [13], if the almost complex structure J is the cylindrical lift of an admissible almost complex structure on P which is regular (that is, such that the pseudoholomorphic discs $\pi_P \circ u$ are transversely cut out, for any pseudoholomorphic disc u in any of the moduli spaces described above), then J is regular meaning that the moduli spaces we described above are transversely cut out. The necessary transversality results for moduli spaces of pseudoholomorphic discs in P with boundary on $\pi_P(\Lambda)$ are carried out in [19]. When transversality holds, these moduli spaces are thus smooth manifolds which can moreover be compactified in the sense of Gromov.

The dimension of a moduli space can be expressed in terms of the Conley–Zehnder indices of its asymptotics. For moduli spaces with only pure asymptotics, we have

$$\dim \mathcal{M}_\Lambda(\gamma;\gamma_1,\dots,\gamma_d) = (\mathrm{CZ}(\gamma)-1) - \sum_{i=1}^{d} (\mathrm{CZ}(\gamma_i)-1) - 1.$$

Then, for a word of pure Reeb chords $\boldsymbol{\delta} = \delta_1 \dots \delta_k$ let us denote by $\ell(\boldsymbol{\delta}) = k$ its length and define $\mathrm{CZ}(\boldsymbol{\delta}) := \sum_{i=1}^{k} \mathrm{CZ}(\delta_i)$. Moreover, let us denote by j^+ the number of positive mixed Reeb chord asymptotics among $\{\gamma_1,\dots,\gamma_d\}$. We have

$$\begin{aligned}\dim \mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{0d};\boldsymbol{\delta}_0,\gamma_1,\boldsymbol{\delta}_1,\dots,\gamma_d,\boldsymbol{\delta}_d) &= (\mathrm{CZ}(\gamma_{0d})-1) + \sum_{\gamma_i\in\mathcal{R}(\Lambda_i,\Lambda_{i-1})} (\mathrm{CZ}(\gamma_i)-1) \\ &\quad - \sum_{\gamma_i\in\mathcal{R}(\Lambda_{i-1},\Lambda_i)} (\mathrm{CZ}(\gamma_i)-1) - \sum (\mathrm{CZ}(\boldsymbol{\delta}_i)-\ell(\boldsymbol{\delta}_i)) + (2-n)j^+ - 1\end{aligned}$$

and

$$\begin{aligned}\dim \mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{d0};\boldsymbol{\delta}_0,\gamma_1,\boldsymbol{\delta}_1,\dots,\gamma_d,\boldsymbol{\delta}_d) &= -(\mathrm{CZ}(\gamma_{d0})-1) + \sum_{\gamma_i\in\mathcal{R}(\Lambda_i,\Lambda_{i-1})} (\mathrm{CZ}(\gamma_i)-1) \\ &\quad - \sum_{\gamma_i\in\mathcal{R}(\Lambda_{i-1},\Lambda_i)} (\mathrm{CZ}(\gamma_i)-1) - \sum (\mathrm{CZ}(\boldsymbol{\delta}_i)-\ell(\boldsymbol{\delta}_i)) + (2-n)(j^+-1) - 1.\end{aligned}$$

We refer to [5, Section 4.3] for the computation of these dimensions. In the following we add an exponent to indicate the dimension of the moduli space, ie $\mathcal{M}^i_\Lambda(\gamma;\gamma_1,\dots,\gamma_d)$ denotes an i-dimensional moduli space. We call *rigid* the pseudoholomorphic discs in a 0-dimensional moduli space.

We define the *action* $\mathfrak{a}(\gamma_{ij})$ of a Reeb chord γ_{ij} by $\mathfrak{a}(\gamma_{ij}) = \int_{\gamma_{ij}} \alpha$ if $i > j$, and $\mathfrak{a}(\gamma_{ij}) = -\int_{\gamma_{ij}} \alpha$ if $i \le j$. By positivity of energy for pseudoholomorphic discs in the moduli spaces defined above, we have the following. If the moduli space $\mathcal{M}_\Lambda(\gamma;\gamma_1,\dots,\gamma_d)$ is not empty then the action of the asymptotics satisfies

$$-\mathfrak{a}(\gamma) + \sum_{i=1}^{d} \mathfrak{a}(\gamma_i) \ge 0.$$

If $\mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{0d};\boldsymbol{\delta}_0,\gamma_1,\boldsymbol{\delta}_1,\dots,\gamma_d,\boldsymbol{\delta}_d)$ is not empty then we have

$$-\mathfrak{a}(\gamma_{0d}) + \sum_{\gamma_i\in\mathcal{R}(\Lambda_{i-1},\Lambda_i)} \mathfrak{a}(\gamma_i) + \sum_{\gamma_i\in\mathcal{R}(\Lambda_i,\Lambda_{i-1})} \mathfrak{a}(\gamma_i) + \sum_{j=0}^{d} \mathfrak{a}(\boldsymbol{\delta}_j) \ge 0.$$

Finally if $\mathcal{M}_{\Lambda_{0\dots d}}(\gamma_{d0};\delta_0,\gamma_1,\delta_1,\dots,\gamma_d,\delta_d)$ is nonempty then we have

$$-\mathfrak{a}(\gamma_{d0})+\sum_{\gamma_i\in\mathcal{R}(\Lambda_{i-1},\Lambda_i)}\mathfrak{a}(\gamma_i)+\sum_{\gamma_i\in\mathcal{R}(\Lambda_i,\Lambda_{i-1})}\mathfrak{a}(\gamma_i)+\sum_{j=0}^{d}\mathfrak{a}(\delta_j)\geq 0.$$

In particular, the definition of action we choose here implies that the maps which will be defined later by a count of pseudoholomorphic discs in the moduli spaces above will be action decreasing.

4 The Chekanov–Eliashberg DGA

In this section we briefly recall the definition of the Chekanov–Eliashberg algebra (C-E algebra) of a Legendrian originally defined in [8; 22] and refer to [8; 19; 20] for details.

Let $\Lambda\subset Y$ be a Legendrian and denote by $C(\Lambda)$ the $\mathbb{Z}_2$-module generated by Reeb chords of Λ. The Chekanov–Eliashberg DGA of Λ, denoted by $\mathcal{A}(\Lambda)$, is the tensor algebra of $C(\Lambda)$ over $\mathbb{Z}_2$:

$$\mathcal{A}(\Lambda)=\bigoplus_{i\geq 0}C(\Lambda)^{\otimes i}$$

with $C(\Lambda)^{\otimes 0}=\mathbb{Z}_2$. The grading of a Reeb chord $\gamma\in\mathcal{R}(\Lambda)$ is given by $|\gamma|_{\mathcal{A}}=1-\mathrm{CZ}(\gamma)$, and we extend it to the whole algebra $\mathcal{A}(\Lambda)$ by

$$|\gamma_1\dots\gamma_d|_{\mathcal{A}}=|\gamma_1|_{\mathcal{A}}+\dots+|\gamma_d|_{\mathcal{A}}.$$

Remark 4.1 In this paper we use a slightly unconventional grading for the C-E algebra. This way the differential will be a map of degree 1 instead of a map of degree -1. We make this choice in order for the Rabinowitz DG bimodule defined in the next section to have a differential of degree 1, as it generalizes (in the sense that we don't use augmentations of the C-E-algebra) the Rabinowitz complex defined in [29] whose differential has degree 1.

The differential $\partial_{\mathcal{A}}$ on $\mathcal{A}$ is given on Reeb chords by

$$\partial_{\mathcal{A}}(\gamma)=\sum_{d\geq 0}\sum_{\gamma_1,\dots,\gamma_d}\#\mathcal{M}^0_{\Lambda}(\gamma;\gamma_1,\dots,\gamma_d)\cdot\gamma_1\dots\gamma_d$$

and extends to the whole algebra $\mathcal{A}$ by Leibniz rule, that is,

$$\partial_{\mathcal{A}}(\gamma_1\gamma_2)=\partial_{\mathcal{A}}(\gamma_1)\gamma_2+\gamma_1\partial_{\mathcal{A}}(\gamma_2).$$

Remark 4.2 In the original definition of the C-E algebra of a Legendrian, the differential is defined by a count of pseudoholomorphic discs with boundary on the projection $\pi_P(\Lambda)$, for $\pi_P\colon\mathbb{R}\times P\to P$. Dimitroglou Rizell proved in [13] that the differential can equivalently be defined by a count of pseudoholomorphic discs with boundary on $\mathbb{R}\times\Lambda\subset\mathbb{R}\times Y$. We use this latter perspective in this paper.

Theorem 4.3 [8; 19; 20] *The map $\partial_{\mathcal{A}}$ is a degree-1 map which satisfies $\partial_{\mathcal{A}}^2=0$.*

The homology of the complex $(\mathcal{A}(\Lambda),\partial_{\mathcal{A}})$ is called the *Legendrian contact homology* of Λ.

5 The Rabinowitz DG bimodule

Let $\Lambda_0, \Lambda_1 \subset Y$ be Legendrian submanifolds and $\mathcal{A}(\Lambda_0)$, $\mathcal{A}(\Lambda_1)$ denote their C-E DGAs. We denote by $C(\Lambda_0, \Lambda_1)$ the graded $\mathbb{Z}_2$-module generated by chords in $\mathcal{R}(\Lambda_0, \Lambda_1)$, and graded with the Conley–Zehnder index. We further denote by $C_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$ the $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodule generated by chords in $\mathcal{R}(\Lambda_0, \Lambda_1)$, ie elements of $C_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$ are linear combinations of words $\boldsymbol{v}_1\gamma_{01}\boldsymbol{v}_0$ where $\boldsymbol{v}_i$ are words of Reeb chords of Λ_i for $i = 0, 1$ and $\gamma_{01} \in \mathcal{R}(\Lambda_0, \Lambda_1)$. The degree of $\boldsymbol{v}_1\gamma_{01}\boldsymbol{v}_0$ is given by $\mathrm{CZ}(\gamma_{01}) + |\boldsymbol{v}_0|_{\mathcal{A}_0} + |\boldsymbol{v}_1|_{\mathcal{A}_1}$. Analogously denote by $C_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_1, \Lambda_0)$ the $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodule generated by chords in $\mathcal{R}(\Lambda_1, \Lambda_0)$.

The Rabinowitz DG bimodule $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$ is a DG $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodule which is defined as follows. The underlying graded bimodule has two different types of generators:

$$\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1) = C_+(\Lambda_0, \Lambda_1) \oplus C_-(\Lambda_0, \Lambda_1),$$

where

- $C_+(\Lambda_0, \Lambda_1)$ is the $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodule whose elements are the same as in $C_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_1, \Lambda_0)$ but the grading of a mixed chord γ_{10} is modified by taking the negative and adding n, that is, $|\gamma_{10}|_{\mathrm{RFC}(\Lambda_0,\Lambda_1)} = n - \mathrm{CZ}(\gamma_{10})$,
- $C_-(\Lambda_0, \Lambda_1) = C_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$.

The differential on $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$ is given by a lower triangular 2×2-matrix

$$\mathfrak{m}_1 = \begin{pmatrix} \boldsymbol{\Delta}_1^{++} & 0 \\ \boldsymbol{b}_1^{-+} & \boldsymbol{b}_1^{--} \end{pmatrix}$$

for which we describe the components on generators of $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0, \Lambda_1)$. The map $\boldsymbol{\Delta}_1^{++}\colon C_+(\Lambda_0, \Lambda_1) \to C_+(\Lambda_0, \Lambda_1)$ on generators is given by the Legendrian contact homology differential of $\Lambda_0 \cup \Lambda_1$ restricted to mixed chords from Λ_0 to Λ_1, namely for a chord $\gamma_{10} \in \mathcal{R}(\Lambda_1, \Lambda_0)$ we have

$$\boldsymbol{\Delta}_1^{++}(\gamma_{10}) = \sum_{\beta_{10}, \boldsymbol{\delta}_0, \boldsymbol{\delta}_1} \#\mathcal{M}^0_{\Lambda_{01}}(\beta_{10}; \boldsymbol{\delta}_0, \gamma_{10}, \boldsymbol{\delta}_1) \cdot \boldsymbol{\delta}_1\, \beta_{10}\, \boldsymbol{\delta}_0,$$

where the sum is over all mixed chords $\beta_{01} \in \mathcal{R}(\Lambda_0, \Lambda_1)$ and all words of pure Reeb chords $\boldsymbol{\delta}_i$ of Λ_i for $i = 0, 1$. Then, we extend it to bimodule elements $\boldsymbol{v}_1\gamma_{10}\boldsymbol{v}_0$ by

$$\boldsymbol{\Delta}_1^{++}(\boldsymbol{v}_1\gamma_{10}\boldsymbol{v}_0) = \partial_{\mathcal{A}(\Lambda_1)}(\boldsymbol{v}_1)\gamma_{10}\boldsymbol{v}_0 + \boldsymbol{v}_1\boldsymbol{\Delta}_1^{++}(\gamma_{10})\boldsymbol{v}_0 + \boldsymbol{v}_1\gamma_{10}\partial_{\mathcal{A}(\Lambda_0)}(\boldsymbol{v}_0)$$

and then by linearity. The component $\boldsymbol{b}_1^{--}\colon C_-(\Lambda_0, \Lambda_1) \to C_-(\Lambda_0, \Lambda_1)$ is given by the restriction to mixed chords from Λ_1 to Λ_0 of the Legendrian contact cohomology differential:

$$\boldsymbol{b}_1^{--}(\gamma_{01}) = \sum_{\beta_{01}, \boldsymbol{\delta}_0, \boldsymbol{\delta}_1} \#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01}; \boldsymbol{\delta}_0, \gamma_{01}, \boldsymbol{\delta}_1) \cdot \boldsymbol{\delta}_1\, \beta_{01}\, \boldsymbol{\delta}_0\,.$$

We extend it to bimodule elements $\boldsymbol{v}_1\gamma_{01}\boldsymbol{v}_0$ by

$$\boldsymbol{b}_1^{--}(\boldsymbol{v}_1\gamma_{01}\boldsymbol{v}_0) = \partial_{\mathcal{A}(\Lambda_1)}(\boldsymbol{v}_1)\gamma_{01}\boldsymbol{v}_0 + \boldsymbol{v}_1\,\boldsymbol{b}_1^{--}(\gamma_{01})\boldsymbol{v}_0 + \boldsymbol{v}_1\gamma_{01}\partial_{\mathcal{A}(\Lambda_0)}(\boldsymbol{v}_0).$$

Finally, the component $\boldsymbol{b}_1^{-+}\colon C_+(\Lambda_0,\Lambda_1)\to C_-(\Lambda_0,\Lambda_1)$ is the *banana map* defined on generators by

$$\boldsymbol{b}_1^{-+}(\gamma_{10})=\sum_{\beta_{01},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\cdot\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0$$

and we extend it to bimodule elements $\boldsymbol{v}_1\gamma_{10}\boldsymbol{v}_0$ by

$$\boldsymbol{b}_1^{-+}(\boldsymbol{v}_1\gamma_{10}\boldsymbol{v}_0)=\boldsymbol{v}_1\,\boldsymbol{b}_1^{-+}(\gamma_{10})\boldsymbol{v}_0.$$

Proposition 5.1 *We have*:

(a) $\mathfrak{m}_1$ *is a degree*-1 *map.*

(b) $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0,\Lambda_1)$ *is a DG* $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$*-bimodule*; *ie* $\mathfrak{m}_1^2=0$.

Proof (a) This is obtained by the dimension formula of the moduli spaces; see Section 3.

(b) This is obtained by considering the algebraic contributions of pseudoholomorphic buildings arising in the boundary of 1-dimensional moduli spaces of the following type:

- $\mathcal{M}^1_{\Lambda_{01}}(\beta_{10};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)$: strips with a positive asymptotic at γ_{10}, a negative asymptotic at β_{10} and negative pure asymptotics at the words $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_1$.
- $\mathcal{M}^1_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1)$: strips with a positive asymptotic at β_{10} and negative asymptotic at γ_{10}.
- $\mathcal{M}^1_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)$: bananas with positive asymptotics at β_{10} and γ_{10}. □

Equivalently we have that $(C_+(\Lambda_0,\Lambda_1),\boldsymbol{\Delta}_1^{++})$ and $(C_-(\Lambda_0,\Lambda_1),\boldsymbol{b}_1^{--})$ are DG $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodules, and $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0,\Lambda_1)$ is the cone of the DG bimodule map $\boldsymbol{b}_1^{-+}\colon C_+(\Lambda_0,\Lambda_1)[1]\to C_-(\Lambda_0,\Lambda_1)$.

Notation 5.2 We suppress the exponents "$++$", "$-+$" and "$--$" indicating the output and input of each component of the differential, and define $\boldsymbol{\Delta}_1:=\boldsymbol{\Delta}_1^{++}$ and $\boldsymbol{b}_1:=\boldsymbol{b}_1^{-+}+\boldsymbol{b}_1^{--}$. We will write explicitly, when needed, if we consider some restriction of the map $\boldsymbol{b}_1$.

Remark 5.3 If the C-E DGAs $\mathcal{A}(\Lambda_0)$ and $\mathcal{A}(\Lambda_1)$ admit augmentations (see [8]) ε_0 and ε_1, respectively, over $\mathbb{Z}_2$, we can use them to turn the bimodule coefficients into elements of $\mathbb{Z}_2$ and thus get a $\mathbb{Z}_2$-module $\mathrm{RFC}(\Lambda_0,\Lambda_1)$. This latter module corresponds to the one defined in [29] in the case where the pair of Lagrangian cobordisms is a pair of trivial cylinders. Moreover, if Λ_1 is a copy of Λ_0, this complex is isomorphic to the complex of the 2-copy described in [18].

We recall now a sufficient condition for the Rabinowitz complex to be acyclic.

Definition 5.4 A pair of Legendrians (Λ_0,Λ_1) in $Y=P\times\mathbb{R}$ is said to be *horizontally displaceable* if there is a Hamiltonian isotopy φ_t of P such that $\Pi_P(\Lambda_0)\cap\varphi_1(\Pi_P(\Lambda_1))=\varnothing$, where $\Pi_P\colon Y\to P$ is the projection.

Theorem 5.5 [29] *If* (Λ_0,Λ_1) *is horizontally displaceable*, *then* $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0,\Lambda_1)$ *is acyclic.*

Remark 5.6 In [29], Theorem 5.5 is proved for the $\mathbb{Z}_2$-module $\mathrm{RFC}(\Lambda_0,\Lambda_1)$ but the proof extends directly to the bimodule case.

Assume that Λ_1 is a perturbation of Λ_0 by a small negative Morse function; see Section 6.1 below. If the perturbation is sufficiently small then by invariance of the C-E algebra [20] (and the fact that the almost complex structure on $\mathbb{R}\times Y$ is the cylindrical lift of an admissible almost complex structure on P; see [13]), the DGAs $\mathcal{A}(\Lambda_0)$ and $\mathcal{A}(\Lambda_1)$ are canonically isomorphic in the sense that there is a canonical identification of Reeb chords of Λ_0 with Reeb chords of Λ_1 such that the differentials coincide under this identification. In this case, we denote by $(\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1),\mathfrak{m}_1)$ the Rabinowitz bimodule, where $\mathcal{A}:=\mathcal{A}(\Lambda_0)=\mathcal{A}(\Lambda_1)$. It is a DG $\mathcal{A}$-bimodule.

6 The Calabi–Yau structure

The goal of this section is to prove the following theorem:

Theorem 6.1 *The C-E algebra $\mathcal{A}(\Lambda)$ of a horizontally displaceable Legendrian sphere $\Lambda\subset Y$ is an $(n+1)$-Calabi–Yau differential graded algebra.*

In order to prove the theorem, we want to find a quasi-isomorphism $\mathcal{A}\to \mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A},\mathcal{A}\otimes\mathcal{A})$ of DG $\mathcal{A}$-bimodules. We will start by describing DG bimodules $\widehat{C}_+(\Lambda_0,\Lambda_1)$ and $\check{C}_-(\Lambda_0,\Lambda_1)$ which are, respectively, a quotient bimodule and a subbimodule of $\mathrm{RFC}_{\mathcal{A}_1\text{-}\mathcal{A}_0}(\Lambda_0,\Lambda_1)$ for a 2-copy $\Lambda_0\cup\Lambda_1$ of Λ. We show then that these bimodules are quasi-isomorphic to $\mathcal{A}$ and $\mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A},\mathcal{A}\otimes\mathcal{A})$, respectively (with some degree shifts). Finally, we define a DG bimodule morphism $\mathrm{CY}\colon \widehat{C}_+(\Lambda_0,\Lambda_1)\to\check{C}_-(\Lambda_0,\Lambda_1)$ and show that the cone of this morphism is quasi-isomorphic to the DG bimodule $\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$. By acyclicity of $\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$, CY is a quasi-isomorphism and thus we get the sought quasi-isomorphism.

6.1 The 2-copy $\Lambda_0\cup\Lambda_1$

In all of Section 6 we will assume that $\Lambda\subset Y$ is an n-dimensional Legendrian sphere. Let $f\colon\Lambda\to\mathbb{R}$ be a $\mathcal{C}^1$-small negative Morse function with exactly one maximum and one minimum, such that the norm of f is much smaller than the length of the shortest Reeb chord of Λ. Let Λ_1 denote the 1-jet of f in a standard neighborhood of Λ (identified with a neighborhood of the 0-section in $J^1(\Lambda)$) and define $\Lambda_0:=\Lambda$. Then we say that $\Lambda_0\cup\Lambda_1$ is a 2-copy of Λ.

Each Reeb chord γ of Λ gives rise to two mixed Reeb chords of $\Lambda_0\cup\Lambda_1$: $\gamma_{01}\in\mathcal{R}(\Lambda_0,\Lambda_1)$ and $\gamma_{10}\in\mathcal{R}(\Lambda_1,\Lambda_0)$. The choice of capping paths in order to define the Conley–Zehnder index of mixed chords is made in such a way that $\mathrm{CZ}(\gamma_0)=\mathrm{CZ}(\gamma_{01})=\mathrm{CZ}(\gamma_{10})$. The Legendrian $\Lambda_0\cup\Lambda_1$ admits two additional mixed Reeb chords from Λ_1 to Λ_0 corresponding to the critical points of f and called *Morse chords*. We denote by y_{01} the Reeb chord corresponding to the maximum of f and by x_{01} the one corresponding to the minimum. Note that $\ell(y_{01})<\ell(x_{01})$ because the function f is negative, and so we will call y_{01} the *minimum Morse Reeb chord* and x_{01} the *maximum Morse Reeb chord*. Note that $\mathrm{CZ}(y_{01})=0$ while $\mathrm{CZ}(x_{01})=n$. Finally, we denote by $C^l_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$ the $\mathcal{A}(\Lambda_1)$-$\mathcal{A}(\Lambda_0)$-bimodule (or $\mathcal{A}$-bimodule for short, as the algebras $\mathcal{A}(\Lambda_0)$ and $\mathcal{A}(\Lambda_1)$ are canonically identified) generated by *long chords* from Λ_1 to Λ_0, ie chords from Λ_1 to Λ_0 which are not Morse chords.

6.2 The bimodules $\widehat{C}_+(\Lambda_0,\Lambda_1)$ and $\check{C}_-(\Lambda_0,\Lambda_1)$

The DG $\mathcal{A}$-bimodule $(\widehat{C}_+(\Lambda_0,\Lambda_1),\widehat{\mathfrak{m}}_1)$ is defined as follows. The underlying graded bimodule is generated by the positive action generators of the Rabinowitz bimodule and the maximum Morse chord:

$$\widehat{C}_+(\Lambda_0,\Lambda_1) = C_+(\Lambda_0,\Lambda_1)[1] \oplus \langle x_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}[1],$$

where $\langle x_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}$ denotes the $\mathcal{A}$-subbimodule of $C_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$ generated by x_{01}. So the chord x_{01} in $\langle x_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}[1]$ has degree $n+1 = \mathrm{CZ}(x_{01})+1$. The differential $\widehat{\mathfrak{m}}_1$ is given on generators by

$$\widehat{\mathfrak{m}}_1(\gamma_{10}) = \mathbf{\Delta}_1(\gamma_{10}) + \gamma_1 x_{01} + x_{01}\gamma_0, \quad \widehat{\mathfrak{m}}_1(x_{01}) = 0,$$

where γ_i denotes the pure chord of Λ_i corresponding to γ_{10}.

Proposition 6.2 *The bimodule $(\widehat{C}_+(\Lambda_0,\Lambda_1),\widehat{\mathfrak{m}}_1)$ is a semifree DG bimodule.*

We state here a lemma that we will use repeatedly in several proofs.

Lemma 6.3 [18] *For a 2-copy $\Lambda_0\cup\Lambda_1$ of Λ_0 we have*:

(a) *For every Reeb chord $\gamma_{01}\in C^l(\Lambda_0,\Lambda_1)$, there are exactly two rigid pseudoholomorphic strips with positive asymptotic at γ_{01} and negative asymptotic at the minimum Reeb chord y_{01}. Moreover, each of these strips has exactly one pure negative chord asymptotic which is the chord γ_0 for one strip and γ_1 for the other (where γ_i denotes the pure Reeb chord of Λ_i corresponding to γ_{01}).*

(b) *For every Reeb chord $\gamma_{10}\in\mathcal{R}(\Lambda_1,\Lambda_0)$, there are exactly two rigid pseudoholomorphic discs with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$ which are bananas with positive asymptotics at γ_{10} and at the maximum Reeb chord x_{01}. Moreover, each of these bananas has exactly one pure negative chord asymptotic which is the chord γ_0 for one banana and γ_1 for the other.*

(c) *The count of rigid pseudoholomorphic strips with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$ admitting a positive puncture at the maximum x_{01} and a negative puncture at a chord β_{01}, vanishes.*

Proof By [18, Theorem 3.6], rigid pseudoholomorphic strips with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$, a positive asymptotic at γ_{01} and a negative asymptotic at the minimum Reeb chord y_{01} (which corresponds to the maximum of the function f) correspond to rigid *generalized discs* which consist of a pseudoholomorphic disc with boundary on $\mathbb{R}\times\Lambda_0$ with a positive asymptotic at γ_0 and a negative gradient flow line of f from the maximum critical point to a point on the boundary of the disc. By rigidity, the pseudoholomorphic disc must be a constant strip $\mathbb{R}\times\gamma_0$ and then there are two ways the flow line can be attached to it (either on $\mathbb{R}\times\{\text{starting point of }\gamma_0\}$, or on $\mathbb{R}\times\{\text{ending point of }\gamma_0\}$). These give the two strips in (a).

For (b), any rigid banana with positive asymptotics at γ_{10} and at the maximum Reeb chord x_{01} corresponds to a rigid generalized discs consisting of a pseudoholomorphic disc with boundary on $\mathbb{R}\times\Lambda_0$ with a positive asymptotic at γ_0 and a negative gradient flow line of f flowing from a point on the boundary of the disc to the minimum critical point (remember x_{01} corresponds to the minimum of f). By rigidity

again the disc must be constant and the two possible ways the flow line can be attached to it provides the two bananas in (b).

For (c), by action reasons the chord β_{01} must be a Morse chord and the strip has no negative pure Reeb chord asymptotics. Thus, [18, Theorem 3.6] tells us that such discs are in bijective correspondence with negative gradient flow lines of f from the critical point corresponding to β_{01} to the minimum. But in our case f has only two critical points so there are either no flow line (for degree reasons in the case $n \geq 2$), or exactly two flow lines (from the maximum to the minimum). □

Proof of Proposition 6.2 We compute $(\widehat{\mathfrak{m}}_1)^2(\gamma_{10}) = \widehat{\mathfrak{m}}_1(\mathbf{\Delta}_1(\gamma_{10})) + \partial_{\mathcal{A}}(\gamma_1)x_{01} + x_{01}\partial_{\mathcal{A}}(\gamma_0)$. For any term $\boldsymbol{\delta}_1\,\beta_{10}\,\boldsymbol{\delta}_0$ appearing in $\mathbf{\Delta}_1(\gamma_{10})$, we have

$$\begin{aligned}\widehat{\mathfrak{m}}_1(\boldsymbol{\delta}_1\,\beta_{10}\,\boldsymbol{\delta}_0) &= \partial_{\mathcal{A}}(\boldsymbol{\delta}_1)\beta_{10}\,\boldsymbol{\delta}_0 + \boldsymbol{\delta}_1(\mathbf{\Delta}_1(\beta_{10}) + \beta_1 x_{01} + x_{01}\beta_0)\,\boldsymbol{\delta}_0 + \boldsymbol{\delta}_1\,\beta_{10}\partial_{\mathcal{A}}(\boldsymbol{\delta}_0)\\ &= \partial_{\mathcal{A}}(\boldsymbol{\delta}_1)\beta_{10}\,\boldsymbol{\delta}_0 + \boldsymbol{\delta}_1\,\mathbf{\Delta}_1(\beta_{10})\,\boldsymbol{\delta}_0 + \boldsymbol{\delta}_1\,\beta_{10}\partial_{\mathcal{A}}(\boldsymbol{\delta}_0) + \boldsymbol{\delta}_1(\beta_1 x_{01} + x_{01}\beta_0)\,\boldsymbol{\delta}_0\\ &= \mathbf{\Delta}_1(\boldsymbol{\delta}_1\,\beta_{10}\,\boldsymbol{\delta}_0) + \boldsymbol{\delta}_1(\beta_1 x_{01} + x_{01}\beta_0)\,\boldsymbol{\delta}_0\,.\end{aligned}$$

So we have

$$(\widehat{\mathfrak{m}}_1)^2(\gamma_{10}) = (\mathbf{\Delta}_1)^2(\gamma_{10}) + \partial_{\mathcal{A}}(\gamma_1)x_{01} + x_{01}\partial_{\mathcal{A}}(\gamma_0) + \sum_{\boldsymbol{\delta}_1\,\beta_{10}\,\boldsymbol{\delta}_0 \in \mathbf{\Delta}_1(\gamma_{10})} \boldsymbol{\delta}_1(\beta_1 x_{01} + x_{01}\beta_0)\,\boldsymbol{\delta}_0\,.$$

We know that $(\mathbf{\Delta}_1)^2(\gamma_{10}) = 0$ and it remains to understand why the last three terms vanish.

The boundary of the compactification of 1-dimensional (after dividing by translation) moduli spaces of bananas with positive asymptotics at γ_{10} and x_{01}, and pure negative Reeb chord asymptotics consists of broken discs which are 2-level buildings connected by a Reeb chord. This Reeb chord is either a mixed chord or a pure chord. In the first case, using Lemma 6.3(c), the only (nonvanishing) possibility is that the upper level of the building is a disc contributing to $\mathbf{\Delta}_1(\gamma_{10})$ with a positive asymptotic at γ_{10}, a negative asymptotic at a chord β_{10} and negative pure Reeb chord asymptotics; and the lower level is a rigid banana with positive asymptotic at β_{10} and x_{01}. By Lemma 6.3(b) there are two such bananas and so we get two buildings contributing to $\boldsymbol{\delta}_1(\beta_1 x_{01} + x_{01}\beta_0)\,\boldsymbol{\delta}_0$, for any $\boldsymbol{\delta}_1\,\beta_{10}\,\boldsymbol{\delta}_0 \in \mathbf{\Delta}_1(\gamma_{10})$. In the second case, the top level of the building is a rigid banana with positive asymptotics at γ_{10} and x_{01}. By Lemma 6.3(b) the pure connecting Reeb chord is either γ_0 or γ_1. The lower level of the building is thus a disk contributing to $\partial_{\mathcal{A}}(\gamma_i)$, for $i = 0, 1$ depending on which one is the connecting pure Reeb chord. This implies that $\partial_{\mathcal{A}}(\gamma_1)x_{01} + x_{01}\partial_{\mathcal{A}}(\gamma_0) = 0$.

The semifree property of $\widehat{C}_+(\Lambda_0, \Lambda_1)$ is deduced directly from the action filtration on chords. Indeed, denote by γ_{10}^j for $1 \leq j \leq k$ the generators of $\widehat{C}_+(\Lambda_0, \Lambda_1)$ which are long chords. Up to relabeling, there are real numbers $\ell_0, \ell_1, \ldots, \ell_k$ such that

$$\mathfrak{a}(x_{01}) < \ell_0 < 0 < \mathfrak{a}(\gamma_{10}^1) < \ell_1 < \cdots < \ell_{k-1} < \mathfrak{a}(\gamma_{10}^k) < \ell_k.$$

The $\mathbb{Z}_2$-vector spaces $F_j\widehat{C}_+ = \{\gamma \in C_+(\Lambda_0,\Lambda_1)[1] \oplus \langle x_{01}\rangle_{\mathbb{Z}_2}[1] : \mathfrak{a}(\gamma) < \ell_j\}$, for $0 \le j \le k$, produce a filtration

$$\mathcal{A}\otimes F_0\widehat{C}_+\otimes\mathcal{A} \subset \mathcal{A}\otimes F_1\widehat{C}_+\otimes\mathcal{A} \subset \cdots \subset \mathcal{A}\otimes F_k\widehat{C}_+\otimes\mathcal{A} = \widehat{C}_+(\Lambda_0,\Lambda_1)$$

giving that $\widehat{C}_+(\Lambda_0,\Lambda_1)$ is semifree and of finite rank. □

Let us now describe the $\mathcal{A}$-bimodule $(\check{C}_-(\Lambda_0,\Lambda_1),\check{\mathfrak{m}}_1)$. The underlying graded bimodule is

$$\check{C}_-(\Lambda_0,\Lambda_1) = C^l_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)\oplus\langle y_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}$$

and as a $\mathcal{A}$-subbimodule of $C_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$ the elements are graded with the Conley–Zehnder index. The differential $\check{\mathfrak{m}}_1$ on generators is given by the restriction to $\check{C}_-(\Lambda_0,\Lambda_1)$ of the map $\boldsymbol{b}_1^{--}$ which was defined in Section 5. By Lemma 6.3(a) we have

$$\check{\mathfrak{m}}_1(y_{01}) = \boldsymbol{b}_1^{--}(y_{01}) = \sum_{\gamma\in\mathcal{R}(\Lambda)} \gamma_1\gamma_{01} + \gamma_{01}\gamma_0.$$

Proposition 6.4 *The bimodule $(\check{C}_-(\Lambda_0,\Lambda_1),\check{\mathfrak{m}}_1)$ is a semifree DG bimodule.*

Proof Consider the boundary of the compactification of moduli spaces of strips with a positive and a negative mixed asymptotics at a generator of $\check{C}_-(\Lambda_0,\Lambda_1)$. This boundary consists of 2-level buildings connected by either a mixed chord or a pure chord. If the connecting mixed chord is x_{01}, then by Lemma 6.3(c) the building will either not exist or arise in pair so its contribution vanishes algebraically. The other cases are exactly those contributing to $(\check{\mathfrak{m}}_1)^2$.

The argument showing that $\check{C}_-(\Lambda_0,\Lambda_1)$ is semifree is analogous to what was done in the proof of Proposition 6.2. □

6.3 Alternative cone description of the Rabinowitz complex of a sphere

We will define a DG bimodule map

$$\mathrm{CY}\colon \widehat{C}_+(\Lambda_0,\Lambda_1) \to \check{C}_-(\Lambda_0,\Lambda_1) \tag{2}$$

and prove that the cone of CY is quasi-isomorphic to the DG bimodule $\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$. For γ_{10}, x_{01} generators of $\widehat{C}_+(\Lambda_0,\Lambda_1)$, we set

$$\mathrm{CY}(\gamma_{10}) = \sum_{\substack{\beta_{01}\in\mathcal{R}^l(\Lambda_0,\Lambda_1)\cup\{y_{01}\}\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1}} \#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\cdot\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0,$$

$$\mathrm{CY}(x_{01}) = \sum_{\substack{\beta_{01}\in\mathcal{R}^l(\Lambda_0,\Lambda_1)\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1}} \#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)\cdot\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0\,.$$

See Figure 1. In the last equation β_{01} can never be the minimum Morse chord y_{01} for energy reasons. Observe also that as ungraded maps the map CY is given on generators of $\widehat{C}_+(\Lambda_0,\Lambda_1)$ by the component of the map $\boldsymbol{b}_1$ taking values in $\check{C}_-(\Lambda_0,\Lambda_1)$.

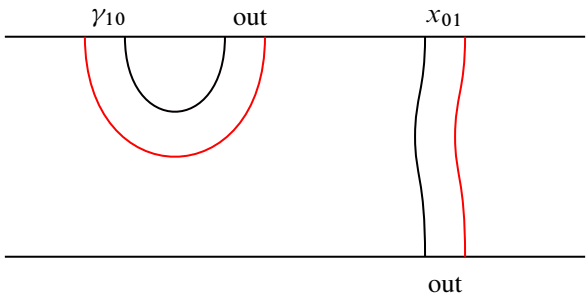


Figure 1: Pseudoholomorphic discs contributing to the map CY.

Lemma 6.5 *The map* CY *is a DG bimodule map.*

Proof We compute first

$$\mathrm{CY}\circ\widehat{\mathfrak{m}}_1(x_{01})+\check{\mathfrak{m}}_1\circ\mathrm{CY}(x_{01})=\check{\mathfrak{m}}_1\circ\mathrm{CY}(x_{01})=(\boldsymbol{b}_1^{--})^2(x_{01})=0,$$

where the second and third equality hold simply by definition. Then, we have

$$\begin{aligned}
\mathrm{CY}\circ\widehat{\mathfrak{m}}_1(\gamma_{10})+\check{\mathfrak{m}}_1\circ\mathrm{CY}(\gamma_{10}) &= \mathrm{CY}(\boldsymbol{\Delta}_1(\gamma_{10})+\gamma_1 x_{01}+x_{01}\gamma_0)+\check{\mathfrak{m}}_1(\mathrm{CY}(\gamma_{10}))\\
&= \mathrm{CY}(\boldsymbol{\Delta}_1(\gamma_{10}))+\gamma_1\mathrm{CY}(x_{01})+\mathrm{CY}(x_{01})\gamma_0+\check{\mathfrak{m}}_1(\mathrm{CY}(\gamma_{10}))\\
&= \mathrm{CY}\Big(\sum \#\mathcal{M}^0(\beta_{10};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\boldsymbol{\delta}_1\beta_{10}\boldsymbol{\delta}_0\Big)+\gamma_1\mathrm{CY}(x_{01})+\mathrm{CY}(x_{01})\gamma_0\\
&\qquad +\widehat{\mathfrak{m}}_1\Big(\sum \#\mathcal{M}^0(\xi_{01};\boldsymbol{\sigma}_0,\gamma_{10},\boldsymbol{\sigma}_1)\boldsymbol{\sigma}_1\xi_{01}\boldsymbol{\sigma}_0\Big)\\
&= \sum \#\mathcal{M}^0(\nu_{01},\boldsymbol{\delta}'_0,\beta_{10},\boldsymbol{\delta}'_1)\#\mathcal{M}^0(\beta_{10};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\boldsymbol{\delta}_1\boldsymbol{\delta}'_1\nu_{01}\boldsymbol{\delta}'_0\boldsymbol{\delta}_0 \qquad (3)\\
&\quad +\gamma_1\mathrm{CY}(x_{01})+\mathrm{CY}(x_{01})\gamma_0 \qquad (4)\\
&\quad +\sum \#\mathcal{M}^0(\xi_{01};\boldsymbol{\sigma}_0,\gamma_{10},\boldsymbol{\sigma}_1)(\partial_{\mathcal{A}}(\boldsymbol{\sigma}_1)\xi_{01}\boldsymbol{\sigma}_0+\boldsymbol{\sigma}_1\xi_{01}\partial_{\mathcal{A}}(\boldsymbol{\sigma}_0)) \qquad (5)\\
&\quad +\sum \#\mathcal{M}^0(\nu_{01};\boldsymbol{\sigma}'_0,\xi_{01},\boldsymbol{\sigma}'_1)\#\mathcal{M}^0(\xi_{01};\boldsymbol{\sigma}_0,\gamma_{10},\boldsymbol{\sigma}_1)\boldsymbol{\sigma}_1\boldsymbol{\sigma}'_1\nu_{01}\boldsymbol{\sigma}'_0\boldsymbol{\sigma}_0. \qquad (6)
\end{aligned}$$

The terms of this last equation correspond actually to the algebraic contributions of the pseudoholomorphic buildings appearing in the boundary of the compactification of moduli spaces of bananas with two positive asymptotics, one at γ_{10} and another at a generator of $\check{C}_-(\Lambda_0,\Lambda_1)$, and some negative pure Reeb chord asymptotics; see Figure 2. □

The cone of CY, $\mathrm{Cone}(\mathrm{CY})=\widehat{C}_+(\Lambda_0,\Lambda_1)[-1]\oplus\check{C}_-(\Lambda_0,\Lambda_1)$, is a DG $\mathcal{A}$-bimodule, and we prove that it is quasi-isomorphic to the Rabinowitz bimodule $\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$. This is almost trivial, indeed consider the bimodule map

$$\nu\colon \mathrm{Cone}(\mathrm{CY})\to\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$$

sending each mixed chord of the 2-copy $\Lambda_0\cup\Lambda_1$ to itself.

Proposition 6.6 *ν is a quasi-isomorphism of DG bimodules.*

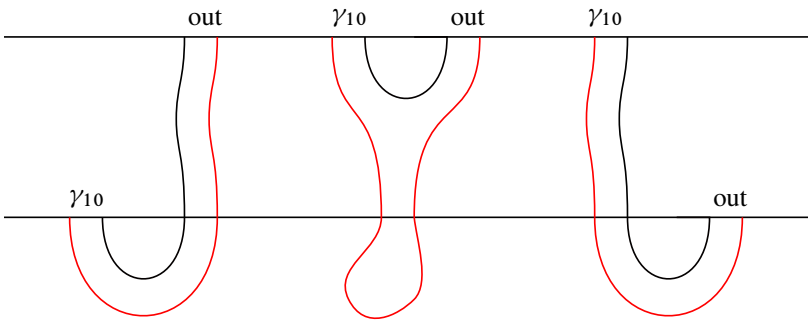


Figure 2: Types of pseudoholomorphic buildings in the boundary of moduli spaces of bananas with a positive asymptotic at γ_{10}. When the connecting Reeb chords between the two components of the leftmost building is a long chord or y_{01}, then it contributes algebraically to (6); and if it is the chord x_{01}, then it contributes algebraically to (4). The middle building schematizes the contributions of type (5) and the rightmost building contributes to (3).

Proof We have

$$\begin{aligned}\nu\circ\mathfrak{m}_1^{\mathrm{Cone(CY)}}(\gamma_{10})+\mathfrak{m}_1\circ\nu(\gamma_{10}) &= \nu(\widehat{\mathfrak{m}}_1(\gamma_{10})+\mathrm{CY}(\gamma_{10}))+\mathfrak{m}_1(\gamma_{10})\\ &= \nu(\boldsymbol{\Delta}_1(\gamma_{10})+\gamma_1x_{01}+x_{01}\gamma_0+\mathrm{CY}(\gamma_{10}))+\boldsymbol{\Delta}_1(\gamma_{10})+\boldsymbol{b}_1(\gamma_{10})\\ &= \boldsymbol{\Delta}_1(\gamma_{10})+\gamma_1x_{01}+x_{01}\gamma_0+\mathrm{CY}(\gamma_{10})+\boldsymbol{\Delta}_1(\gamma_{10})+\boldsymbol{b}_1(\gamma_{10}).\end{aligned}$$

Observe that $\gamma_1x_{01}+x_{01}\gamma_0+\mathrm{CY}(\gamma_{10})=\boldsymbol{b}_1(\gamma_{10})$ because $\mathrm{CY}(\gamma_{10})$ is defined by a count of bananas with positive asymptotics at γ_{10} and at a chord in $\mathcal{R}^l(\Lambda_0,\Lambda_1)$, and the two other terms correspond to bananas with positive asymptotics at γ_{10} and x_{10}. So the sum above vanishes. Then the relation $\nu\circ\mathfrak{m}_1^{\mathrm{Cone(CY)}}+\mathfrak{m}_1\circ\nu=0$ when the input is x_{01} or any generator of $\check{C}_-(\Lambda_0,\Lambda_1)$ follows directly from the definition of the maps. □

6.4 The Calabi–Yau isomorphism

In this subsection, we assume that Λ_0 is horizontally displaceable, which implies that the complex $\mathrm{RFC}_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_0,\Lambda_1)$ is acyclic. Hence it implies that the complex Cone(CY) defined in the previous section is also acyclic and thus that the degree-0 map $\mathrm{CY}\colon\widehat{C}_+(\Lambda_0,\Lambda_1)\to\check{C}_-(\Lambda_0,\Lambda_1)$ is a quasi-isomorphism of DG bimodules. We will show that this quasi-isomorphism is a Calabi–Yau isomorphism, by showing that there are quasi-isomorphisms of DG bimodules $\mathcal{A}\simeq\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]$ and $\mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A},\mathcal{A}\otimes\mathcal{A})\simeq\check{C}_-(\Lambda_0,\Lambda_1)$.

We define a bimodule map $F\colon\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]\to\mathcal{A}$ by

$$F(\boldsymbol{v}_1\gamma_{10}\boldsymbol{v}_0)=0,\qquad F(\boldsymbol{v}_1x_{01}\boldsymbol{v}_0)=\boldsymbol{v}_1\boldsymbol{v}_0$$

for $\boldsymbol{v}_i$ words of Reeb chords of Λ_i which on the right-hand side are canonically identified with words of Reeb chords of Λ_0.

Proposition 6.7 *The map F is a quasi-isomorphism of DG bimodules.*

Proof The fact that F is a degree-0 chain map follows directly from the definition. It is thus a DG bimodule morphism and we check that it is a quasi-isomorphism by proving that its cone is acyclic. For a word of Reeb chords $\boldsymbol{w}_1 = w^1 \dots w^k$ of Λ_1, write

$$\overline{\boldsymbol{w}}_1 = \sum_{j=1}^{k} w^1 \dots w^{j-1} w_{10}^j w^{j+1} \dots w^k \in C_{\mathcal{A}\text{-}\mathcal{A}}(\Lambda_1, \Lambda_0).$$

Consider the DG bimodule $\mathrm{Cone}(F)$, whose differential we denote by $\partial^{\mathrm{Cone}(F)}$. We define a degree -1 $\mathbb{Z}_2$-linear map $h\colon \mathrm{Cone}(F) \to \mathrm{Cone}(F)$ by

$$h(\boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0) = 0, \quad h(\boldsymbol{w}_1 x_{01} \boldsymbol{v}_0) = \overline{\boldsymbol{w}}_1 \boldsymbol{v}_0, \quad h(a^1 \dots a^k) = x_{01} a^1 \dots a^k$$

and show that h defines a homotopy between the identity map and the zero map on $\mathrm{Cone}(F)$, ie that

$$\partial^{\mathrm{Cone}(F)} \circ h + h \circ \partial^{\mathrm{Cone}(F)} = \mathrm{id}_{\mathrm{Cone}(F)}.$$

We check this relation for the three different types of elements in $\mathrm{Cone}(F)$. First

$$\begin{aligned}
(\partial^{\mathrm{Cone}(F)} \circ h + h \circ \partial^{\mathrm{Cone}(F)})&(\boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0) \\
&= 0 + h \circ \widehat{\mathfrak{m}}_1(\boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0) + h \circ F(\boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0) \\
&= h(\partial_{\mathcal{A}}(\boldsymbol{w}_1)\gamma_{10}\boldsymbol{v}_0 + \boldsymbol{w}_1 \gamma_{10} \partial_{\mathcal{A}}(\boldsymbol{v}_0) + \boldsymbol{w}_1 \boldsymbol{\Delta}_1(\gamma_{10})\boldsymbol{v}_0 + \boldsymbol{w}_1(\gamma x_{01} + x_{01}\gamma)\boldsymbol{v}_0) + 0 \\
&= h(\boldsymbol{w}_1 \gamma x_{01} \boldsymbol{v}_0 + \boldsymbol{w}_1 x_{01} \gamma \boldsymbol{v}_0) \\
&= \overline{\boldsymbol{w}}_1 \gamma \boldsymbol{v}_0 + \boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0 + \overline{\boldsymbol{w}}_1 \gamma \boldsymbol{v}_0 \\
&= \boldsymbol{w}_1 \gamma_{10} \boldsymbol{v}_0.
\end{aligned}$$

Then,

$$\begin{aligned}
(\partial^{\mathrm{Cone}(F)} \circ h + h \circ \partial^{\mathrm{Cone}(F)})(\boldsymbol{w}_1 x_{01} \boldsymbol{v}_0) &= \partial^{\mathrm{Cone}(F)}(\overline{\boldsymbol{w}}_1 \boldsymbol{v}_0) + h(\partial_{\mathcal{A}}(\boldsymbol{w}_1) x_{01} \boldsymbol{v}_0 + \boldsymbol{w}_1 x_{01} \partial_{\mathcal{A}}(\boldsymbol{v}_0) + \boldsymbol{w}_1 \boldsymbol{v}_0) \\
&= \widehat{\mathfrak{m}}_1(\overline{\boldsymbol{w}}_1)\boldsymbol{v}_0 + \overline{\boldsymbol{w}}_1 \partial_{\mathcal{A}}(\boldsymbol{v}_0) + \overline{\partial_{\mathcal{A}}(\boldsymbol{w}_1)}\boldsymbol{v}_0 + \overline{\boldsymbol{w}}_1 \partial_{\mathcal{A}}(\boldsymbol{v}_0) + x_{01}\boldsymbol{w}_1 \boldsymbol{v}_0 \\
&= \widehat{\mathfrak{m}}_1(\overline{\boldsymbol{w}}_1)\boldsymbol{v}_0 + \overline{\partial_{\mathcal{A}}(\boldsymbol{w}_1)}\boldsymbol{v}_0 + x_{01}\boldsymbol{w}_1 \boldsymbol{v}_0.
\end{aligned}$$

Observe that $\overline{\partial_{\mathcal{A}}(\boldsymbol{w}_1)} = \boldsymbol{\Delta}_1(\overline{\boldsymbol{w}}_1)$ which is one component of $\widehat{\mathfrak{m}}_1(\overline{\boldsymbol{w}}_1)$. Assuming $\boldsymbol{w}_1 = w^1 \dots w^k$ we thus get

$$\begin{aligned}
(\partial^{\mathrm{Cone}(F)} \circ h + h \circ \partial^{\mathrm{Cone}(F)})(\boldsymbol{w}_1 x_{01} \boldsymbol{v}_0) &= \sum_{j=1}^{k} (w^1 \dots w^{j-1}(w^j x_{01} + x_{01} w^j) w^{j+1} \dots w^k)\boldsymbol{v}_0 + x_{01}\boldsymbol{w}_1 \boldsymbol{v}_0 \\
&= \boldsymbol{w}_1 x_{01} \boldsymbol{v}_0.
\end{aligned}$$

Finally, for a word $a^1 \dots a^k$ in $\mathcal{A}$, we compute

$$\begin{aligned}
(\partial^{\mathrm{Cone}(F)} \circ h + h \circ \partial^{\mathrm{Cone}(F)})(a^1 \dots a^k) &= \partial^{\mathrm{Cone}(F)}(x_{01} a^1 \dots a^k) + h \circ \partial_{\mathcal{A}}(a^1 \dots a^k) \\
&= x_{01} \partial_{\mathcal{A}}(a^1 \dots a^k) + F(x_{01} a^1 \dots a^k) + x_{01} \partial_{\mathcal{A}}(a^1 \dots a^k) \\
&= a^1 \dots a^k.
\end{aligned}$$

□

Observe that Proposition 6.7 implies that $\widehat{C}_+(\Lambda_0, \Lambda_1)[-n-1]$ is a semifree resolution of the diagonal DG bimodule $\mathcal{A}$. This means in particular that there is a quasi-isomorphism of DG bimodules $\mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A}\otimes\mathcal{A}) \simeq \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0, \Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})$. This semifree resolution is also of finite rank so $\mathcal{A}$ is homologically smooth.

Remark 6.8 Such a finite-rank semifree resolution is explicitly given in [27, Section 3.6] for a DG category associated to a quiver (with some specific properties).

We consider now a bimodule map $G\colon \check{C}_-(\Lambda_0,\Lambda_1)\to \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1],\mathcal{A}\otimes\mathcal{A})$ defined on generators by $G(\gamma_{01})=\phi_\gamma$, where $\phi_\gamma\colon \widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]\to\mathcal{A}\otimes\mathcal{A}$ is the bimodule map defined on generators by

$$\phi_\gamma(\beta_{10})=\begin{cases}1\otimes 1 & \text{if } \beta=\gamma,\\ 0 & \text{otherwise,}\end{cases}\qquad \phi_\gamma(x_{01})=0,$$

and similarly $G(y_{01})=\phi_y$ which is the bimodule map defined by $\phi_y(\beta_{10})=0$ and $\phi_y(x_{01})=1\otimes 1$.

Proposition 6.9 *The map G is a quasi-isomorphism of DG bimodules.*

Proof The differential on $\mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1],\mathcal{A}\otimes\mathcal{A})$ is given by $D(\phi)=\phi\circ\widehat{\mathfrak{m}}_1+\partial_{\mathcal{A}\otimes\mathcal{A}}\circ\phi$. We want to prove that

$$G\circ\check{\mathfrak{m}}_1+D\circ G=0 \tag{7}$$

and in order to do so we check that (7) is satisfied when applied to each generator of $\check{C}_-(\Lambda_0,\Lambda_1)$. Let first γ_{01} be a generator of $\check{C}_-(\Lambda_0,\Lambda_1)$ which is a long chord. We want to prove

$$G\circ\check{\mathfrak{m}}_1(\gamma_{01})+G(\gamma_{01})\circ\widehat{\mathfrak{m}}_1+\partial_{\mathcal{A}\otimes\mathcal{A}}\circ G(\gamma_{01})=0. \tag{8}$$

The left-hand side being a map $\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]\to\mathcal{A}\otimes\mathcal{A}$, we prove that it vanishes for each generator of $\widehat{C}_+(\Lambda_0,\Lambda_1)$. We divide it into three cases:

(a) the generator γ_{10} which corresponds to γ_{01} in the 2-copy link,

(b) a long chord generator β_{10} which is not γ_{10}, and finally

(c) the Morse generator x_{01}.

(a) We have

$$\partial_{\mathcal{A}\otimes\mathcal{A}}\circ G(\gamma_{01})(\gamma_{10})=\partial_{\mathcal{A}\otimes\mathcal{A}}\circ\phi_\gamma(\gamma_{10})=\partial_{\mathcal{A}\otimes\mathcal{A}}(1\otimes 1)=0.$$

Moreover $G(\gamma_{01})\circ\widehat{\mathfrak{m}}_1(\gamma_{10})=\phi_\gamma\circ\widehat{\mathfrak{m}}_1(\gamma_{10})=0$ because γ_{10} cannot be the output of $\widehat{\mathfrak{m}}_1(\gamma_{10})$ for action reasons, and so we get 0 by definition of ϕ_γ. Similarly, $G\circ\check{\mathfrak{m}}_1(\gamma_{01})(\gamma_{10})=0$ because γ_{01} is not an output of $\check{\mathfrak{m}}_1(\gamma_{01})$ for action reasons.

(b) In this case observe first that $\partial_{\mathcal{A}\otimes\mathcal{A}}\circ G(\gamma_{01})(\beta_{10})=\partial_{\mathcal{A}\otimes\mathcal{A}}\circ\phi_\gamma(\beta_{10})=0$. Then, the (rigid) pseudoholomorphic discs contributing to $\phi_\gamma\circ\widehat{\mathfrak{m}}_1(\beta_{10})$ are discs with a positive asymptotic at β_{10}, a negative asymptotic at γ_{10} and potentially negative asymptotics to pure Reeb chords; see Figure 3. Consider such a disc which contributes $v_1\dots v_r\gamma_{10}w_1\dots w_s$ to $\widehat{\mathfrak{m}}_1(\beta_{10})$. Then, note that

$$\phi_\gamma\in\mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1],\mathcal{A}\otimes\mathcal{A}),$$

where $\mathcal{A}\otimes\mathcal{A}$ is endowed with the *inner* bimodule structure, which means that

$$\phi_\gamma(v_1\dots v_r\gamma_{10}w_1\dots w_s)=v_1\dots v_r\cdot\phi_\gamma(\gamma_{10})\cdot w_1\dots w_s=v_1\dots v_r\cdot 1\otimes 1\cdot w_1\dots w_s=w_1\dots w_s\otimes v_1\dots v_r.$$

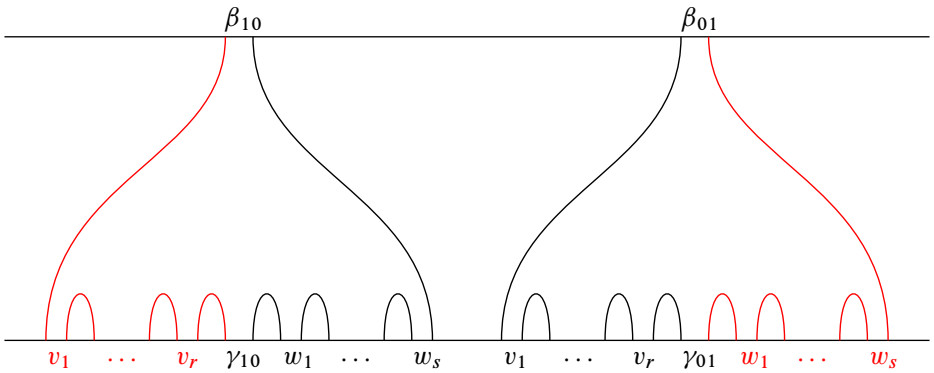


Figure 3: Left: pseudoholomorphic disc contributing to $\phi_\gamma \circ \widehat{\mathfrak{m}}_1(\beta_{10})$; right: pseudoholomorphic discs contributing to $G \circ \check{\mathfrak{m}}_1(\gamma_{01})(\beta_{10})$.

Now let us look at the last term in (8) applied to β_{10}, which is $G \circ \check{\mathfrak{m}}_1(\gamma_{01})(\beta_{10})$. The pseudoholomorphic discs contributing $w_1 \dots w_s \otimes v_1 \dots v_r$ to this term are discs schematized on the right in Figure 3. Indeed, such a disc contributes $w_1 \dots w_s \beta_{01} v_1 \dots v_r$ to $\check{\mathfrak{m}}_1(\gamma_{01})$ and then

$$\begin{aligned} G(w_1 \dots w_s \beta_{01} v_1 \dots v_r)(\beta_{10}) &= (w_1 \dots w_s \cdot G(\beta_{01}) \cdot v_1 \dots v_r)(\beta_{10}) \\ &= (w_1 \dots w_s \cdot \phi_\beta \cdot v_1 \dots v_r)(\beta_{10}) \\ &= w_1 \dots w_s \otimes v_1 \dots v_r, \end{aligned}$$

where for the last equality we use the outer bimodule structure (and the definition of ϕ_β). Finally, observe that the discs schematized in Figure 3 correspond to lifted discs in the terminology of [18] and that there is a bijective correspondence between discs on the left and right of the figure, they both correspond to a rigid disc with boundary on Λ with a pure positive asymptotic at β and pure negative asymptotics at $v_1, \dots, v_r, \gamma, w_1, \dots, w_s$ in this order; see, eg [1, Theorem 3.2]. Thus we get (8) as for any disc contributing to $\phi_\gamma \circ \widehat{\mathfrak{m}}_1(\beta_{10})$ corresponds (bijectively) a disc contributing the same thing to $G \circ \check{\mathfrak{m}}_1(\gamma_{01})(\beta_{10})$.

(c) In this case we have directly that

$$G \circ \check{\mathfrak{m}}_1(\gamma_{01})(x_{01}) + G(\gamma_{01}) \circ \widehat{\mathfrak{m}}_1(x_{01}) + \partial_{\mathcal{A}\otimes\mathcal{A}} \circ G(\gamma_{01})(x_{01}) = 0$$

because y_{01} is never an output of $\check{\mathfrak{m}}_1(\gamma_{01})$ for action reasons (and so $G \circ \check{\mathfrak{m}}_1(\gamma_{01})(x_{01}) = 0$), $\widehat{\mathfrak{m}}_1(x_{01}) = 0$, and $G(\gamma_{01})(x_{01}) = \phi_\gamma(x_{01}) = 0$.

Finally we check that (7) is satisfied when applied to the Morse generator $y_{01} \in \check{C}_-(\Lambda_0, \Lambda_1)$. Recall that $\check{\mathfrak{m}}_1(y_{01}) = \sum_{\gamma \in \mathcal{R}(\Lambda)} \gamma \gamma_{01} + \gamma_{01} \gamma$, so we have

$$G \circ \check{\mathfrak{m}}_1(y_{01})(\gamma_{10}) = \gamma \otimes 1 + 1 \otimes \gamma, \quad G \circ \check{\mathfrak{m}}_1(y_{01})(x_{01}) = 0.$$

On the other side, for a chord γ_{10}, we have

$$\begin{aligned} D \circ G(y_{01})(\gamma_{10}) &= G(y_{01}) \circ \widehat{\mathfrak{m}}_1(\gamma_{10}) + \partial_{\mathcal{A}\otimes\mathcal{A}} \circ G(y_{01})(\gamma_{10}) \\ &= G(y_{01}) \circ \widehat{\mathfrak{m}}_1(\gamma_{10}) + 0 \\ &= G(y_{01})(\boldsymbol{\Delta}_1(\gamma_{10}) + \gamma x_{01} + x_{01}\gamma) \\ &= G(y_{01})(\gamma x_{01} + x_{01}\gamma) = \gamma \otimes 1 + 1 \otimes \gamma. \end{aligned}$$

And finally,

$$\begin{aligned} D \circ G(y_{01})(x_{01}) &= G(y_{01}) \circ \widehat{\mathfrak{m}}_1(\gamma_{10}) + \partial_{\mathcal{A}\otimes\mathcal{A}} \circ G(y_{01})(x_{01}) \\ &= 0 + \partial_{\mathcal{A}\otimes\mathcal{A}}(1 \otimes 1) = 0. \end{aligned}$$

Observe that G admits an inverse $G^{-1}\colon \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A}) \to \check{C}_-(\Lambda_0,\Lambda_1)$ which maps a bimodule morphism $\phi \in \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})$ given by

$$\phi(\gamma_{10}) = \sum_{j=1}^{k_\gamma} \boldsymbol{w}_\gamma^j \otimes \boldsymbol{v}_\gamma^j, \quad \phi(x_{01}) = \sum_{j=1}^{k_x} \boldsymbol{w}_x^j \otimes \boldsymbol{v}_x^j$$

to $G^{-1}(\phi) = \sum_{\gamma\in\mathcal{R}(\Lambda)} \sum_{j=1}^{k_\gamma} \boldsymbol{w}_\gamma^j \gamma_{01} \boldsymbol{v}_\gamma^j + \sum_{j=1}^{k_x} \boldsymbol{w}_x^j y_{01} \boldsymbol{v}_x^j$. □

Proof of Theorem 6.1 From the proof of Proposition 6.7 we obtained that $\mathcal{A}$ is homologically smooth with $\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]$ a finite-dimensional semifree resolution. Consider the shifted Calabi–Yau map

$$\mathcal{CY} := \mathrm{CY}[-n-1]\colon \widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1] \to \check{C}_-(\Lambda_0,\Lambda_1)[-n-1].$$

If Λ_0 is horizontally displaceable then CY is a quasi-isomorphism. Moreover, $\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1]$ is quasi-isomorphic to $\mathcal{A}$ by the map F and the target $\check{C}_-(\Lambda_0,\Lambda_1)[-n-1]$ is quasi-isomorphic to $\mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})[-n-1]$ by the map $G[-n-1]$. Remember also that we have $\mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})[-n-1] \cong \mathrm{RHom}_{\mathcal{A}\text{-}\mathcal{A}}(\mathcal{A}, \mathcal{A}\otimes\mathcal{A})[-n-1]$, so we get a quasi-isomorphism $\mathcal{A} \to \mathcal{A}^!\,[-n-1]$. Finally, we need to check that $\mathcal{CY} = \mathcal{CY}^!\,[-n-1]$. Note that

$$\mathcal{CY}^!\colon \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\check{C}_-(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A}) \to \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})$$

is by definition given by $\mathcal{CY}^!(\phi) = \phi \circ \mathcal{CY}$. Observe that the target of $\mathcal{CY}^!$ is quasi-isomorphic to $\check{C}_-(\Lambda_0,\Lambda_1)$ via the map G^{-1}, while its domain is quasi-isomorphic to $\widehat{C}_+(\Lambda_0,\Lambda_1)$ via the DG bimodule map

$$H\colon \widehat{C}_+(\Lambda_0,\Lambda_1) \to \mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\check{C}_-(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})$$

defined on generators by $H(\gamma_{10})(\gamma_{01}) = 1\otimes 1$ and $H(\gamma_{10})$ vanishes on other generators of $\check{C}_-(\Lambda_0,\Lambda_1)$; and $H(x_{01})(y_{01}) = 1\otimes 1$ and vanishes otherwise. Indeed, by a similar argument as for Proposition 6.9, we prove that H is a quasi-isomorphism. First, we check that $H\circ\widehat{\mathfrak{m}}_1 + D\circ H = 0$ by proving that this equation holds when applied to any generator of $\widehat{C}_+(\Lambda_0,\Lambda_1)$. Let γ_{10} be a long chord generator, we prove that

$$H \circ \widehat{\mathfrak{m}}_1(\gamma_{10}) + H(\gamma_{10}) \circ \check{\mathfrak{m}}_1 + \partial_{\mathcal{A}\otimes\mathcal{A}} \circ H(\gamma_{10}) = 0. \tag{9}$$

Notice that the left-hand side is a bimodule map in $\mathrm{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\check{C}_-(\Lambda_0,\Lambda_1)[-n-1], \mathcal{A}\otimes\mathcal{A})$ so we check that it vanishes when applied to any generator of $\check{C}_-(\Lambda_0,\Lambda_1)$.

(a) For the generator γ_{01} corresponding to γ_{10} we have $H\circ\widehat{\mathfrak{m}}_1(\gamma_{10})(\gamma_{01}) = H(\gamma_{10})\circ\check{\mathfrak{m}}_1(\gamma_{01}) = 0$ because γ_{10} never appears as an output of $\widehat{\mathfrak{m}}_1(\gamma_{10})$, and γ_{01} never appears as an output of $\check{\mathfrak{m}}_1(\gamma_{01})$. Moreover, $\partial_{\mathcal{A}\otimes\mathcal{A}} \circ H(\gamma_{10})(\gamma_{01}) = \partial_{\mathcal{A}\otimes\mathcal{A}}(1\otimes 1) = 0$ and thus we get (9).

(b) For a generator β_{01} which is a long chord not equal to γ_{01} we have $\partial_{\mathcal{A}\otimes\mathcal{A}}\circ H(\gamma_{10})(\beta_{01})=0$. Then we can prove $H\circ\widehat{\mathfrak{m}}_1(\gamma_{10})(\beta_{01})+H(\gamma_{10})\circ\check{\mathfrak{m}}_1(\beta_{01})=0$ by following the same argument as in the proof of Proposition 6.9 considering pseudoholomorphic discs as in Figure 3 but exchanging β_{10} with γ_{10} and β_{01} with γ_{01}.

(c) Finally for the generator y_{01} we have first $\partial_{\mathcal{A}\otimes\mathcal{A}}(\gamma_{10})(y_{01})=0$. Then by definition $\widehat{\mathfrak{m}}_1(\gamma_{10})=\boldsymbol{\Delta}_1(\gamma_{10})+\gamma x_{01}+x_{01}\gamma$ so

$$H\circ\widehat{\mathfrak{m}}_1(\gamma_{10})(y_{01})=H(\boldsymbol{\Delta}_1(\gamma_{10}))(y_{01})+H(\gamma x_{01}+x_{01}\gamma)(y_{01})=0+\gamma\otimes 1+1\otimes\gamma$$

For the other term, we have $\check{\mathfrak{m}}_1(y_{01})=\sum_{\gamma\in\mathcal{R}(\Lambda)}\gamma\gamma_{01}+\gamma_{01}\gamma$ and thus $H(\gamma_{10})\circ\check{\mathfrak{m}}_1(y_{01})=1\otimes\gamma+\gamma\otimes 1$, which gives (9).

It remains to prove that $H\circ\widehat{\mathfrak{m}}_1(x_{01})+D\circ H(x_{01})=0$. First, for a long chord β_{01} generator of $\check{C}_-(\Lambda_0,\Lambda_1)$ we have $H\circ\widehat{\mathfrak{m}}_1(x_{01})(\beta_{01})=\partial_{\mathcal{A}\otimes\mathcal{A}}\circ H(x_{01})(\beta_{01})=0$ (recall $\widehat{\mathfrak{m}}_1(x_{01})=0$), and $H(x_{01})\circ\check{\mathfrak{m}}_1(\beta_{01})=0$ because y_{01} is never an output of $\check{\mathfrak{m}}_1(\beta_{01})$ for action reasons. Finally we have $H\circ\widehat{\mathfrak{m}}_1(x_{01})(y_{01})=0$, $H(x_{01})\circ\check{\mathfrak{m}}_1(y_{01})=0$ as y_{01} is never an output of $\check{\mathfrak{m}}_1(y_{01})$ and $\partial_{\mathcal{A}\otimes\mathcal{A}}\circ H(x_{01})(y_{01})=\partial_{\mathcal{A}\otimes\mathcal{A}}(1\otimes 1)=0$.

The map H admits an inverse that is defined similarly as the inverse of G so we skip writing the formula.

Thus we get a DG bimodule map

$$G^{-1}\circ\mathcal{CY}^!\circ H\colon\widehat{C}_+(\Lambda_0,\Lambda_1)\to\check{C}_-(\Lambda_0,\Lambda_1).$$

We compute this map on generators:

$$G^{-1}\circ\mathcal{CY}^!\circ H(\gamma_{10})=G^{-1}\circ H(\gamma_{10})\circ\mathcal{CY}\,.$$

And $H(\gamma_{10})\circ\mathcal{CY}\in\operatorname{Hom}_{\mathcal{A}\text{-}\mathcal{A}}(\widehat{C}_+(\Lambda_0,\Lambda_1)[-n-1],\mathcal{A}\otimes\mathcal{A})$ is given by a count of bananas or strips with a positive asymptotic at γ_{01}, ie

$$\begin{aligned}H(\gamma_{10})\circ\mathcal{CY}(\beta_{10})&=H(\gamma_{10})\Bigg(\sum_{\xi_{01}\in\mathcal{R}^I(\Lambda_0,\Lambda_1)\cup\{y_{01}\}}\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\xi_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,\xi_{01}\,\boldsymbol{\delta}_0\Bigg)\\&=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\gamma_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\otimes\boldsymbol{\delta}_0\end{aligned}$$

and

$$H(\gamma_{10})\circ\mathcal{CY}(x_{01})=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\gamma_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\otimes\boldsymbol{\delta}_0\,.$$

So we get

$$G^{-1}\circ\mathcal{CY}^!\circ H(\gamma_{10})=\sum_{\beta_{10},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\gamma_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0+\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\gamma_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,y_{01}\,\boldsymbol{\delta}_0$$

while

$$\mathcal{CY}(\gamma_{10})=\sum_{\substack{\beta_{01}\in\mathcal{R}^I(\Lambda_0,\Lambda_1)\\\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}}\#\mathcal{M}^0(\beta_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0+\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(y_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,y_{01}\,\boldsymbol{\delta}_0\,.$$

Similarly we can compute

$$G^{-1}\circ\mathcal{CY}^!\circ H(x_{01})=\sum_{\beta_{10},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(y_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0$$

while

$$\mathcal{CY}(x_{01})=\sum\#\mathcal{M}^0(\beta_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)\,\boldsymbol{\delta}_1\,\beta_{01}\,\boldsymbol{\delta}_0\,.$$

So in order to get $\mathcal{CY}=(G^{-1}\circ\mathcal{CY}^!\circ H)[-n-1]:=G^{-1}[-n-1]\circ\mathcal{CY}[-n-1]\circ H[-n-1]$ we need to check that:

(a) For any $\gamma_{10}\in\mathcal{R}(\Lambda_1,\Lambda_0)$, $\beta_{01}\in\mathcal{R}^l(\Lambda_0,\Lambda_1)$, and their corresponding $\gamma_{01}\in\mathcal{R}^l(\Lambda_0,\Lambda_1)$ and $\beta_{10}\in\mathcal{R}(\Lambda_1,\Lambda_0)$, we have

$$\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\beta_{01};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\gamma_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1).$$

(b) For any $\beta_{01}\in\mathcal{R}^l(\Lambda_0,\Lambda_1)$ and its corresponding $\beta_{10}\in\mathcal{R}(\Lambda_1,\Lambda_0)$, we have

$$\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(\beta_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0(y_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1).$$

The point (a) follows from [18, Theorem 3.6]. Indeed the count of rigid bananas with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$ and positive asymptotics at β_{10} and γ_{01} is in bijective correspondence with the count of rigid strips with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1^A)$, where Λ_1^A is a translation of Λ_1 in the positive Reeb direction by $A\gg 0$ such that the only mixed Reeb chords are from Λ_0 to Λ_1^A, with a positive asymptotic at β_{10}^A and a negative asymptotic at γ_{10}^A, where β_{10}^A and γ_{10}^A are the mixed chords of $\Lambda_0\cup\Lambda_1^A$ corresponding to β_{10} and γ_{01}, respectively. By [18, Theorem 3.6], this count of rigid strips with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1^A)$ corresponds to the count of strips with boundary on $\mathbb{R}\times\Lambda$ with a positive asymptotic at β (the pure chord corresponding to β_{10}) and a negative asymptotic at γ. Similarly, the count of rigid bananas with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$ and positive asymptotics at β_{01} and γ_{10} is in bijective correspondence with the count of rigid strips with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1^{-A})$ with a positive asymptotic at β_{01}^{-A} and negative asymptotic at γ_{01}^{-A}. This last count is also in bijective correspondence with the count of rigid strips with boundary on $\mathbb{R}\times\Lambda$, positively asymptotic to β and negatively asymptotic to γ.

For (b), by [18, Theorem 3.6] a rigid strip with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$, a positive asymptotic at β_{01} and a negative asymptotic at x_{01} corresponds to a generalized disc consisting of a disc with boundary on $\mathbb{R}\times\Lambda$ having a positive asymptotic at β together with a negative gradient flow line of f from the minimum (remember that the maximum Reeb chord $x_{01}\in\mathcal{R}(\Lambda_0,\Lambda_1)$ corresponds to the minimum critical point x of f) to a point on the boundary of the disc. But there is no nonconstant negative gradient flow line flowing from the minimum so the boundary of the disc has to pass by x (more precisely the boundary crosses $\mathbb{R}\times\{x\}$ as we are in the cylindrical setting). On the other side, a rigid banana with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1)$ and positively asymptotic to β_{10} and y_{01} corresponds to a rigid generalized disc consisting of a disc with boundary on $\mathbb{R}\times\Lambda$ positively asymptotic to β together with a negative

gradient flow line of f from a point on the boundary of the disc to the maximum critical point y. Again, this flow line must be constant and the boundary of the disc must pass through $\mathbb{R}\times\{y\}$. Now we check that the count of these two types of discs is the same.

If $\dim\Lambda = 1$, assume $\mathcal{M}^0(\beta_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)$ is not empty, ie assume that there is a pseudoholomorphic disc with boundary on $\mathbb{R}\times\Lambda$ passing through $\mathbb{R}\times\{x\}$, positively asymptotic to a chord β and negatively asymptotic to the words $\boldsymbol{\delta}_0,\boldsymbol{\delta}_1$. The boundary of the disc is transverse to $\mathbb{R}\times\{x\}$. In particular, it passes also through all $\mathbb{R}\times\{\mathrm{pt}\}$ for every points pt sufficiently close to x on Λ. If the function f is chosen so that its critical points are sufficiently close to each other, we get the equality in (b).

If $\dim\Lambda\geq 2$, we use the results in [12]. Denote by $\mathcal{M}_\Lambda^{\{*\}}(\beta;\boldsymbol{\delta}_0,\boldsymbol{\delta}_1)$ the moduli space of pseudoholomorphic discs with boundary on $\mathbb{R}\times\Lambda$, positively asymptotic to β, negatively asymptotic to the words $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_1$, and having a marked point $*$ on the boundary of the disc in the domain which is situated between the puncture mapped to the last Reeb chord of the word $\boldsymbol{\delta}_0$ and the puncture mapped to the first Reeb chord of the word $\boldsymbol{\delta}_1$. There is still an action of $\mathbb{R}$ on this moduli space. There is a smooth evaluation map

$$ev\colon \mathcal{M}_\Lambda^{\{*\}}(\beta;\boldsymbol{\delta}_0,\boldsymbol{\delta}_1)\to\Lambda$$

given by $ev(u)=u(*)$. The evaluation map takes values in Λ (instead of the general $\mathbb{R}\times\Lambda$) as we are in the cylindrical setting. By a generalization of [30, Chapter 3], every point of Λ is a regular value of the evaluation map, and so are in particular the minimum and maximum Morse critical points x and y. Now, using the transversality results in [12, Section 8], we have that the evaluation map is proper and thus $\#ev^{-1}(x)=\#ev^{-1}(y)$. Finally, the 0-dimensional moduli spaces $\mathcal{M}_\Lambda^0(\beta_{01};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1)$ and $\mathcal{M}_\Lambda^0(y_{01};\boldsymbol{\delta}_0,\beta_{10},\boldsymbol{\delta}_1)$ are, respectively, identified with $ev^{-1}(x)$ and $ev^{-1}(y)$. □

6.5 Relation to Sabloff's duality

Sabloff [33] proved that linearized Legendrian contact homology and cohomology of a Legendrian knot satisfy some duality. It was then generalized to all connected Legendrians in $P\times\mathbb{R}$ in [18]. Given an augmentation $\varepsilon\colon\mathcal{A}\to\mathbb{Z}_2$ of the C-E algebra of a Legendrian sphere of dimension n, its linearized Legendrian contact homology and cohomology fit into a duality exact sequence which gives

$$\begin{aligned}
&\mathrm{LCH}_n^\varepsilon(\Lambda)\simeq\mathrm{LCH}_\varepsilon^{-1}(\Lambda)\oplus\langle[\Lambda]\rangle,\\
&\mathrm{LCH}_\varepsilon^n(\Lambda)\simeq\mathrm{LCH}_{-1}^\varepsilon(\Lambda)\oplus\langle[\mathrm{pt}]\rangle,\\
&\mathrm{LCH}_k^\varepsilon(\Lambda)\simeq\mathrm{LCH}_\varepsilon^{n-k-1}(\Lambda)\quad\text{for } k\neq -1,n,
\end{aligned}$$

where $[\Lambda]$ corresponds to the fundamental class in $H_n(\Lambda)$, [pt] corresponds to the point class in $H_0(\Lambda)$ and the grading is the usual grading of the C-E algebra $|\gamma|=\mathrm{CZ}(\gamma)-1$.

The Calabi–Yau structure we established on the C-E algebra recovers Sabloff duality in the following way. First, for the C-E algebra $\mathcal{A}$ of Λ, define $\mathcal{A}^e=\mathcal{A}\otimes\mathcal{A}^{\mathrm{op}}$. Given an augmentation ε of $\mathcal{A}$ as above, we can see our coefficient field $\boldsymbol{k}$, which is equal to $\mathbb{Z}_2$, as a DG left $\mathcal{A}^e$-module $\boldsymbol{k}_\varepsilon$ whose structure map is given by

$$\mathcal{A}^e\otimes\boldsymbol{k}_\varepsilon\to\boldsymbol{k}_\varepsilon,\quad (\boldsymbol{a},\boldsymbol{b})\otimes m\mapsto\varepsilon(\boldsymbol{a})m\varepsilon(\boldsymbol{b}),$$

and differential vanishes. Observe now that the DG $\mathcal{A}$-bimodules $\widehat{C}_+(\Lambda_0,\Lambda_1)$ and $\check{C}_-(\Lambda_0,\Lambda_1)$ can be viewed as right DG $\mathcal{A}^e$-modules with module structure given by $c\cdot(\boldsymbol{a},\boldsymbol{b})=\boldsymbol{b}c\boldsymbol{a}$. By taking the tensor product of $\widehat{C}_+(\Lambda_0,\Lambda_1)$, resp. $\check{C}_-(\Lambda_0,\Lambda_1)$, with $\boldsymbol{k}_\varepsilon$, the Calabi–Yau structure on $\mathcal{A}$ induces a quasi-isomorphism of chain complexes

$$\widehat{C}_+(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon \xrightarrow{\simeq} \check{C}_-(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon. \tag{10}$$

On the left-hand side, the differential which we denote by $\widehat{\mathfrak{m}}_1^\varepsilon$ is given by

$$\widehat{\mathfrak{m}}_1^\varepsilon(\gamma_{10})=\sum_{\beta_{10},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{10};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\cdot\beta_{10}\varepsilon(\boldsymbol{\delta}_0)\varepsilon(\boldsymbol{\delta}_1).$$

There is no term with output x_{01} because the two bananas having positive asymptotics at x_{01} and γ_{10} give the same contribution ($x_{01}\varepsilon(\gamma)+\varepsilon(\gamma)x_{01}=0$). Then $\widehat{\mathfrak{m}}_1^\varepsilon(x_{01})=0$. Observe that $\widehat{C}_+(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon$ is the direct sum of two subcomplexes: one generated by long chords (all mixed chords from Λ_0 to Λ_1), and the other generated by x_{01} having trivial differential. Moreover, the complex generated by long chords and with differential $\widehat{\mathfrak{m}}_1^\varepsilon$ is quasi-isomorphic to the linearized Legendrian contact homology of Λ, $\mathrm{LCH}^\varepsilon_{n-*}(\Lambda)$; see [18] (a cycle consisting only of long chords of $\widehat{C}_+$-grading $*$ in $\widehat{C}_+(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon$ corresponds to a cycle in $\mathrm{LCH}^\varepsilon_{n-*}(\Lambda)$ which is graded with the standard C-E grading). So

$$\widehat{C}_+(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon\simeq\mathrm{LCH}^\varepsilon_{n-*}(\Lambda)\oplus\langle x_{01}\rangle.$$

Remember also that $\mathrm{CZ}(x_{01})=n$ so the standard C-E grading for x_{01} is $n-1$ while its $\widehat{C}$-grading is $n+1$. Let us turn to the right-hand side of (10) now. The differential $\check{\mathfrak{m}}_1^\varepsilon$ is given by

$$\check{\mathfrak{m}}_1^\varepsilon(\gamma_{01})=\sum_{\beta_{01},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1)\cdot\beta_{01}\varepsilon(\boldsymbol{\delta}_0)\varepsilon(\boldsymbol{\delta}_1)$$

and $\check{\mathfrak{m}}_1^\varepsilon(y_{01})=\sum_{\gamma\in\mathcal{R}(\Lambda)}\gamma_{01}\varepsilon(\gamma)+\varepsilon(\gamma)\gamma_{01}=0$. Again we have a direct sum of two subcomplexes: one generated by long Reeb chords from Λ_1 to Λ_0 and quasi-isomorphic to the linearized Legendrian contact cohomology $\mathrm{LCH}^{*-1}_\varepsilon(\Lambda)$ [18], and the other generated by y_{01} and with trivial differential. So

$$\check{C}_-(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\boldsymbol{k}_\varepsilon\simeq\mathrm{LCH}^{*-1}_\varepsilon(\Lambda)\oplus\langle y_{01}\rangle.$$

We get thus a quasi-isomorphism

$$\mathrm{LCH}^\varepsilon_{n-*}(\Lambda)\oplus\langle x_{01}\rangle\xrightarrow{\simeq}\mathrm{LCH}^{*-1}_\varepsilon(\Lambda)\oplus\langle y_{01}\rangle,$$

which gives $\mathrm{LCH}^\varepsilon_n(\Lambda)\simeq\mathrm{LCH}^{-1}_\varepsilon(\Lambda)\oplus\langle y_{01}\rangle$ (remember that $\mathrm{CZ}(y_{01})=0$ so the standard C-E algebra degree of y_{01} is -1). On the other side, cycles in $\mathrm{LCH}^n_\varepsilon(\Lambda)$ come from cycles in $\mathrm{LCH}^\varepsilon_{-1}(\Lambda)$ but also from x_{01} under the Calabi–Yau map CY. Finally when $*\neq 0,n+1$ we have $\mathrm{LCH}^\varepsilon_{n-*}(\Lambda)\simeq\mathrm{LCH}^{*-1}_\varepsilon(\Lambda)$ so we have recovered Sabloff duality.

7 Products and higher-order structure maps

In this section we define chain complexes $\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)$ and $\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)$ over $\mathbb{Z}_2$, obtained from $\check{C}_{-}(\Lambda_0,\Lambda_1)$ and $\widehat{C}_{+}(\Lambda_0,\Lambda_1)$ by the bimodule tensor product with the diagonal bimodule $\mathcal{A}$. We show that these complexes admit product structures and that the Calabi–Yau morphism (2) induced on them preserves the products in homology.

Notation 7.1 In the following we will consider Reeb chords with boundary on a 3-copy (and even more) of a Legendrian. Previously we wrote γ_{01} for a Reeb chord from Λ_1 to Λ_0 and γ_{10} for the corresponding Reeb chords from Λ_0 to Λ_1. Unless specified, this "correspondence" doesn't apply anymore in this section, that is, we will denote Reeb chords of the 3-copy by γ_{ij} with $1\leq i\neq j\leq 2$, but γ_{ij} is not necessarily the chord from Λ_j to Λ_i which corresponds to γ_{ji}, unless specified. But we will still use x_{ij} and y_{ij} to, respectively, denote maximum and minimum Morse Reeb chords between different 2-copies.

7.1 Chain complexes for Hochschild homology and cohomology

We start by describing the chain complex $(\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1),\widehat{\mathfrak{m}}_1)$, where by abuse of notation we also denote by $\widehat{\mathfrak{m}}_1$ the differential but this should not create any confusion. The vector space $\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)$ is infinite-dimensional, generated over $\mathbb{Z}_2$ by elements of the form $\gamma_{10}\boldsymbol{a}$ and $x_{01}\boldsymbol{a}$ where $\boldsymbol{a}=a_1\dots a_k$ denotes a word of Reeb chords of Λ. The differential is given by

$$\widehat{\mathfrak{m}}_1(\gamma_{10}\boldsymbol{a})=\sum_{\beta_{10},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{10};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1)\cdot\beta_{10}\,\boldsymbol{\delta}_0\,\boldsymbol{a}\,\boldsymbol{\delta}_1+x_{01}\gamma\boldsymbol{a}+x_{01}\boldsymbol{a}\gamma \\ +\sum_{j=1}^{k}\gamma_{10}a_1\dots a_{j-1}\partial_{\mathcal{A}}(a_j)a_{j+1}\dots a_k,$$

where γ is the pure Reeb chord corresponding to γ_{10}, and

$$\widehat{\mathfrak{m}}_1(x_{01}\boldsymbol{a})=\sum_{j=1}^{k}x_{01}a_1\dots a_{j-1}\partial_{\mathcal{A}}(a_j)a_{j+1}\dots a_k.$$

See Figure 4.

This complex computes the Hochschild homology of $\mathcal{A}$, by definition. Indeed, $\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)$ is equal to the tensor product $\widehat{C}_{+}(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\mathcal{A}$ where $\widehat{C}_{+}(\Lambda_0,\Lambda_1)$ is viewed as a DG right $\mathcal{A}^e$-module as in Section 6.5, and $\mathcal{A}$ is viewed as a DG left $\mathcal{A}^e$-module with $(a,b)\cdot a_1\dots a_k=aa_1\dots a_kb$. Remember moreover that $\widehat{C}_{+}(\Lambda_0,\Lambda_1)$ is a semifree resolution of $\mathcal{A}$, so the homology of $\widehat{C}_{+}(\Lambda_0,\Lambda_1)\otimes_{\mathcal{A}^e}\mathcal{A}$ is isomorphic to the Hochschild homology of $\mathcal{A}$. Similarly, the complex $(\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1),\check{\mathfrak{m}}_1)$ is generated over $\mathbb{Z}_2$ by elements $\gamma_{01}\boldsymbol{a}$ and $y_{01}\boldsymbol{a}$, and

$$\check{\mathfrak{m}}_1(\gamma_{01}a_1\dots a_k) \\ =\sum_{\beta_{01},\boldsymbol{\delta}_0,\boldsymbol{\delta}_1}\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1)\cdot\beta_{01}\,\boldsymbol{\delta}_0\,a_1\dots a_k\,\boldsymbol{\delta}_1+\sum_{j=1}^{k}\gamma_{01}a_1\dots a_{j-1}\partial_{\mathcal{A}}(a_j)a_{j+1}\dots a_k$$

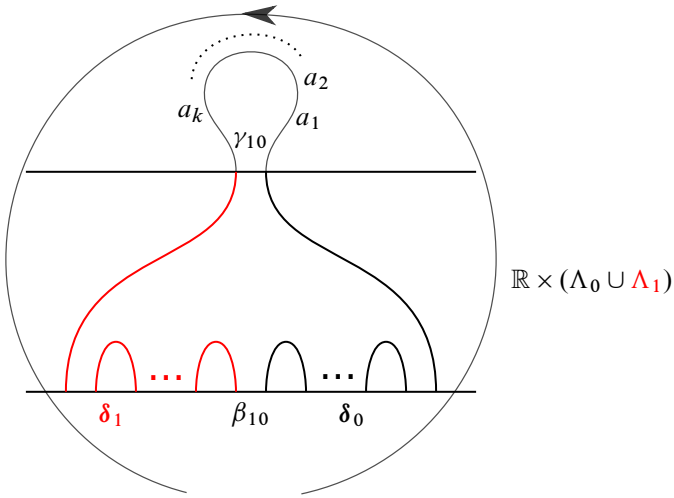


Figure 4: Pseudoholomorphic disc contributing to $\widehat{\mathfrak{m}}(\gamma_{10} a_1 \dots a_k)$. Observe that the "bubble" at the top is not a pseudoholomorphic disc but a way to write the cyclic word $\gamma_{10} a_1 \dots a_k$. The contribution $\beta_{10} \boldsymbol{\delta}_0 a_1 \dots a_k \boldsymbol{\delta}_1$ of the disc is given by the output mixed chord β_{10} followed by a word of Reeb chords as they appear along the boundary of the disc when following it counterclockwise from β_{10}.

and

$$\check{\mathfrak{m}}_1(y_{01}\boldsymbol{a}) = \sum_{\gamma \in \mathcal{R}(\Lambda)} \gamma_{01}\gamma \boldsymbol{a} + \gamma_{01}\boldsymbol{a}\gamma + \sum_{j=1}^{k} y_{01} a_1 \dots a_{j-1} \partial_{\mathcal{A}}(a_j) a_{j+1} \dots a_k,$$

where γ_{10} is the mixed chord corresponding to γ. Again, we can check that

$$\check{C}_{-}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \simeq \check{C}_{-}(\Lambda_0, \Lambda_1) \otimes_{\mathcal{A}^e} \mathcal{A} \simeq \mathrm{RHom}_{\mathcal{A}^e}(\mathcal{A}, \mathcal{A} \otimes \mathcal{A}) \otimes_{\mathcal{A}^e} \mathcal{A} \simeq \mathrm{RHom}_{\mathcal{A}^e}(\mathcal{A}, \mathcal{A})$$

and thus the complex $(\check{C}_{-}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1), \check{\mathfrak{m}}_1)$ computes the Hochschild cohomology of $\mathcal{A}$. By Theorem 6.1 we thus get an isomorphism between Hochschild homology and cohomology of the C-E algebra of a horizontally displaceable Legendrian sphere in $\mathbb{R} \times P$.

7.2 Product structures

In [29] the author defined a product structure and more generally A_∞-structure maps on the complex $\mathrm{RFC}(\Lambda_0, \Lambda_1)$ in the case when Λ_0, Λ_1 admit augmentations of their C-E algebras. When Λ_1 is a small negative push-off of Λ_0, this product extends naturally to the chain complex $\mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ (cone of the banana map induced on the cyclic model $\widehat{C}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ and with values in $\check{C}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$). This product is defined by counting the same type of pseudoholomorphic discs as the one counted to get a product on $\mathrm{RFC}(\Lambda_0, \Lambda_1)$ but keeping the negative pure Reeb chords asymptotics as coefficients instead of turning them into elements of $\mathbb{Z}_2$ with augmentations. It doesn't seem possible however to define a product directly on the $\mathcal{A}$-bimodule $\mathrm{RFC}(\Lambda_0, \Lambda_1)$, because there is no good way to deal with the

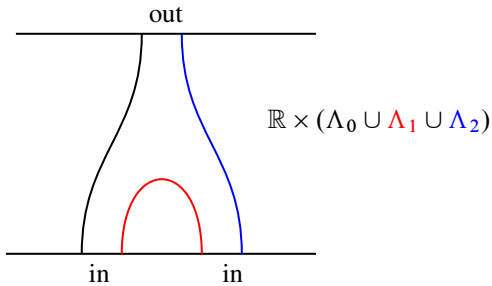


Figure 5: Pseudoholomorphic building with boundary on $\mathbb{R}\times(\Lambda_0\cup\Lambda_1\cup\Lambda_2)$ contributing to the product on $\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)$.

coefficients (see [6] for constructions of A_∞-structures with coefficients in a noncommutative algebra). The pseudoholomorphic discs we count have boundary on the cylinder over a 3-copy $\Lambda_0\cup\Lambda_1\cup\Lambda_2$ of a Legendrian Λ. This 3-copy is given by $\Lambda_0=\Lambda$ and then Λ_1 and Λ_2 are small perturbed negative push-offs of Λ (Λ_2 is a slightly more negative push-off than Λ_1) such that $\Lambda_0\cup\Lambda_1$, $\Lambda_0\cup\Lambda_2$ and $\Lambda_1\cup\Lambda_2$ are 2-copies as described in Section 6.1. Observe that $\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)$ is a subcomplex of $\mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)$, and that the restriction of the product in $\mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)$ to $\check{C}^{\mathrm{cyc}}_{-}$ takes values in $\check{C}^{\mathrm{cyc}}_{-}$; see [29]. So we get a well-defined product on $\check{C}^{\mathrm{cyc}}_{-}$. More precisely, given $\Lambda_0\cup\Lambda_1\cup\Lambda_2$ a 3-copy of Λ there is a degree-0 map $\check{\mathfrak{m}}_2\colon \check{C}^{\mathrm{cyc}}_{-}(\Lambda_1,\Lambda_2)\otimes\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)\to\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_2)$ defined by

$$\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\gamma_{01}\boldsymbol{a_0})=\sum_{\substack{\gamma_{02}\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot\gamma_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2,$$

where γ_{ij} can also be the minimum Morse Reeb chord. And this map $\check{\mathfrak{m}}_2$ satisfies the Leibniz rule $\check{\mathfrak{m}}_1\circ\check{\mathfrak{m}}_2+\check{\mathfrak{m}}_2(\mathrm{id}\otimes\check{\mathfrak{m}}_1)+\check{\mathfrak{m}}_2(\check{\mathfrak{m}}_1\otimes\mathrm{id})=0$. This product is the standard "two negative inputs one positive output" product; see Figure 5. As for the differential, the word of pure Reeb chords in the output element is obtained by following the boundary of the pseudoholomorphic disc counterclockwise from the mixed output Reeb chord. Because of the coefficients in the C-E algebra, verifying the Leibniz rule for the product on $\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)$ involves slightly more terms than for the Leibniz rule for the product on the $\mathbb{Z}_2$ vector space $C(\Lambda_0,\Lambda_1)$ as done in [29], so we detail it now. We want to prove

$$\check{\mathfrak{m}}_1\circ\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\gamma_{01}\boldsymbol{a}_0)+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\check{\mathfrak{m}}_1(\gamma_{01}\boldsymbol{a}_0))+\check{\mathfrak{m}}_2(\check{\mathfrak{m}}_1(\gamma_{12}\boldsymbol{a}_1),\gamma_{01}\boldsymbol{a}_0)=0.$$

The left-hand side can be rewritten

$$\begin{aligned}\sum_{\substack{\gamma_{02}\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}}&\#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot\check{\mathfrak{m}}_1(\gamma_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2)\\ &+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\check{\mathfrak{m}}_1(\gamma_{01})\boldsymbol{a}_0+\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0))+\check{\mathfrak{m}}_2(\check{\mathfrak{m}}_1(\gamma_{12})\boldsymbol{a}_1+\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0),\gamma_{01}\boldsymbol{a}_0)\\ =\sum\#\mathcal{M}^0_{\Lambda_{012}}&(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot(\check{\mathfrak{m}}_1(\gamma_{02})\,\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2+\gamma_{02}\partial_{\mathcal{A}}(\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2))\\ &+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\check{\mathfrak{m}}_1(\gamma_{01})\boldsymbol{a}_0)+\check{\mathfrak{m}}_2(\check{\mathfrak{m}}_1(\gamma_{12})\boldsymbol{a}_1,\gamma_{01}\boldsymbol{a}_0)\\ &+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0))+\check{\mathfrak{m}}_2(\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0),\gamma_{01}\boldsymbol{a}_0).\end{aligned}$$

We separate the terms having the boundary of an $\boldsymbol{a_i}$ from the other terms to get

$$\begin{aligned}
&= \sum \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot\big(\check{\mathfrak{m}}_1(\gamma_{02})\,\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2+\gamma_{02}\partial_{\mathcal{A}}(\boldsymbol{\delta}_0)\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2\\
&\qquad\qquad +\boldsymbol{\delta}_0\,\boldsymbol{a}_0\partial_{\mathcal{A}}(\boldsymbol{\delta}_1)\boldsymbol{a}_1\,\boldsymbol{\delta}_2+\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\partial_{\mathcal{A}}(\boldsymbol{\delta}_2)\big)\\
&\quad+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\check{\mathfrak{m}}_1(\gamma_{01})\boldsymbol{a}_0)+\check{\mathfrak{m}}_2(\check{\mathfrak{m}}_1(\gamma_{12})\boldsymbol{a}_1,\gamma_{01}\boldsymbol{a}_0)\\
&\quad+\sum \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot(\gamma_{02}\,\boldsymbol{\delta}_0\,\partial_{\mathcal{A}}(\boldsymbol{a}_0)\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2+\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\partial_{\mathcal{A}}(\boldsymbol{a}_1)\,\boldsymbol{\delta}_2)\\
&\quad+\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0))+\check{\mathfrak{m}}_2(\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0),\gamma_{01}\boldsymbol{a}_0)
\end{aligned}$$

The two first lines vanish as the algebraic contributions of pseudoholomorphic buildings in the boundary of moduli spaces of type $\mathcal{M}^1_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)$. The two last lines vanish also, because by definition of the product we have

$$\sum_{\substack{\gamma_{02}\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}} \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot\gamma_{02}\,\boldsymbol{\delta}_0\,\partial_{\mathcal{A}}(\boldsymbol{a}_0)\,\boldsymbol{\delta}_1\,\boldsymbol{a}_1\,\boldsymbol{\delta}_2=\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,\gamma_{01}\partial_{\mathcal{A}}(\boldsymbol{a}_0))$$

and

$$\sum_{\substack{\gamma_{02}\\ \boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}} \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{02};\boldsymbol{\delta}_0,\gamma_{01},\boldsymbol{\delta}_1,\gamma_{12},\boldsymbol{\delta}_2)\cdot\gamma_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{a}_0\,\boldsymbol{\delta}_1\,\partial_{\mathcal{A}}(\boldsymbol{a}_1)\,\boldsymbol{\delta}_2=\check{\mathfrak{m}}_2(\gamma_{12}\partial_{\mathcal{A}}(\boldsymbol{a}_1),\gamma_{01}\boldsymbol{a}_0),$$

and so the Leibniz rule is satisfied. Moreover, it follows from [18, Theorem 5.5] that the product $\check{\mathfrak{m}}_2$ is unital with the unit given by the minimum Morse Reeb chord, ie we have

$$\begin{aligned}
\check{\mathfrak{m}}_1(y_{01})=\check{\mathfrak{m}}_1(y_{12})&=0,\\
\check{\mathfrak{m}}_2(\gamma_{12}\boldsymbol{a}_1,y_{01})&=\gamma_{02}\boldsymbol{a}_1,\\
\check{\mathfrak{m}}_2(y_{12},\gamma_{01}\boldsymbol{a}_0)&=\gamma_{02}\boldsymbol{a}_0
\end{aligned}$$

for all $\gamma_{01}\in\mathcal{R}(\Lambda_0,\Lambda_1)$, $\gamma_{12}\in\mathcal{R}(\Lambda_1,\Lambda_2)$ and $\gamma_{02}\in\mathcal{R}(\Lambda_0,\Lambda_2)$ corresponding to the *same* chord (either a long chord or a Morse chord).

Notation 7.2 In the following, we will write only γ_{01} instead of the more general $\gamma_{01}\boldsymbol{a}$ for an element in $\check{C}^{\mathrm{cyc}}_{-}(\Lambda_0,\Lambda_1)$. This is just in order to reduce a bit the length of formulas. We will also do the same for elements in $\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)$. In particular we'll define a product $\widehat{\mathfrak{m}}_2$ only for pairs of inputs $(\gamma_{21},\gamma_{10})$, where the inputs can also be maximum Morse Reeb chords, but keeping in mind that the rule to define more generally $\widehat{\mathfrak{m}}_2(\gamma_{21}\boldsymbol{a}_1,\gamma_{10}\boldsymbol{a}_0)$ is the same as for the product $\check{\mathfrak{m}}_2$. Namely, the output will contain the words $\boldsymbol{a}_0$ and $\boldsymbol{a}_1$ in a larger word of pure Reeb chords obtained by following the boundary of pseudoholomorphic discs involved in the definition of $\widehat{\mathfrak{m}}_2(\gamma_{21},\gamma_{10})$ counterclockwise.

So let us now construct this product on $\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)$. From a 3-copy of Λ we define a map

$$\widehat{\mathfrak{m}}_2\colon\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_1,\Lambda_2)\otimes\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_1)\to\widehat{C}^{\mathrm{cyc}}_{+}(\Lambda_0,\Lambda_2)$$

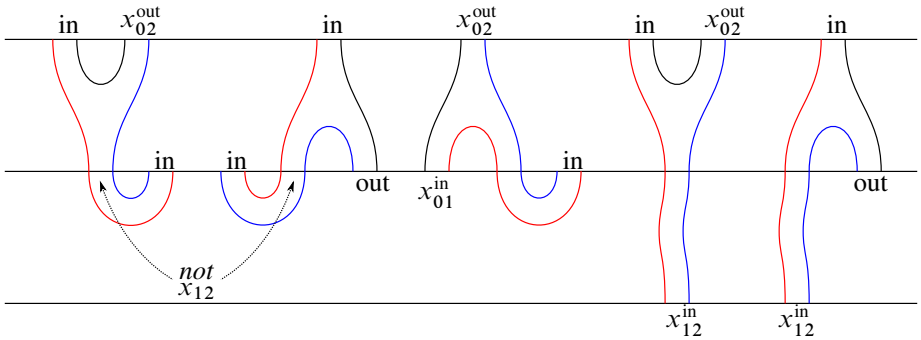


Figure 6: Pseudoholomorphic buildings contributing to the product $\widehat{\mathfrak{m}}_2$ on $\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_1)$.

by counting 2-level pseudoholomorphic buildings as shown on Figure 6. More precisely, for generators γ_{10}, x_{01} of $\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_1)$ and generators γ_{21}, x_{12} of $\widehat{C}_+^{\text{cyc}}(\Lambda_1,\Lambda_2)$ we have

$$\begin{aligned}
\widehat{\mathfrak{m}}_2(\gamma_{21},\gamma_{10}) = {}& \sum_{\beta_{12}\in\mathcal{R}^I(\Lambda_1,\Lambda_2)\cup\{y_{12}\}} \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_1'\\ \boldsymbol{\delta}_2,\boldsymbol{\delta}_2'}} \#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\cdot\#\mathcal{M}^0_{\Lambda_{12}}(\beta_{12};\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\boldsymbol{\delta}_2\boldsymbol{\delta}_2' \\
& + \sum_{\gamma_{20}} \sum_{\beta_{12}\in\mathcal{R}^I(\Lambda_1,\Lambda_2)\cup\{y_{12}\}} \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_1'\\ \boldsymbol{\delta}_2,\boldsymbol{\delta}_2'}} \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\#\mathcal{M}^0_{\Lambda_{12}}(\beta_{12};\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2) \\
& \qquad \cdot\gamma_{20}\,\boldsymbol{\delta}_0\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\boldsymbol{\delta}_2\boldsymbol{\delta}_2', \\
\widehat{\mathfrak{m}}_2(\gamma_{21},x_{01}) = {}& \sum_{\beta_{12}} \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_1'\\ \boldsymbol{\delta}_2,\boldsymbol{\delta}_2'}} \#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\#\mathcal{M}^0_{\Lambda_{12}}(\beta_{12};\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\boldsymbol{\delta}_2\boldsymbol{\delta}_2', \\
\widehat{\mathfrak{m}}_2(x_{12},\gamma_{10}) = {}& \sum_{\beta_{12}} \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_1'\\ \boldsymbol{\delta}_2,\boldsymbol{\delta}_2'}} \#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\#\mathcal{M}^0_{\Lambda_{12}}(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\boldsymbol{\delta}_2\boldsymbol{\delta}_2' \\
& + \sum_{\gamma_{20},\beta_{12}} \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_1'\\ \boldsymbol{\delta}_2,\boldsymbol{\delta}_2'}} \#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\#\mathcal{M}^0_{\Lambda_{12}}(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2)\cdot\gamma_{20}\,\boldsymbol{\delta}_0\boldsymbol{\delta}_1\boldsymbol{\delta}_1'\boldsymbol{\delta}_2\boldsymbol{\delta}_2', \\
\widehat{\mathfrak{m}}_2(x_{12},x_{01}) = {}& 0.
\end{aligned}$$

Observe that in the definition of $\widehat{\mathfrak{m}}_2(\gamma_{21},x_{01})$ and $\widehat{\mathfrak{m}}_2(x_{12},\gamma_{10})$, the "connecting" chord β_{12} will automatically be in $\mathcal{R}^I(\Lambda_1,\Lambda_2)\cup\{y_{12}\}$. In the first case, it has to be a Morse chord for action reasons, and can not be x_{12} for degree reasons. In the second case it can not be x_{12} also for degree reasons.

By [29, Proposition 2] we can check that this map $\widehat{\mathfrak{m}}_2$ is of degree 0. Then we have:

Proposition 7.3 *The map $\widehat{\mathfrak{m}}_2$ satisfies $\widehat{\mathfrak{m}}_1\circ\widehat{\mathfrak{m}}_2+\widehat{\mathfrak{m}}_2(\mathrm{id}\otimes\widehat{\mathfrak{m}}_1)+\widehat{\mathfrak{m}}_2(\widehat{\mathfrak{m}}_1\otimes\mathrm{id})=0$, ie $\widehat{\mathfrak{m}}_2$ descends to a well-defined map on homology.*

Proof We prove the proposition for each type of pair of inputs. For a pair of inputs $(\gamma_{21},\gamma_{10})$, we obtain the Leibniz rule by considering the algebraic contributions of pseudoholomorphic buildings in the

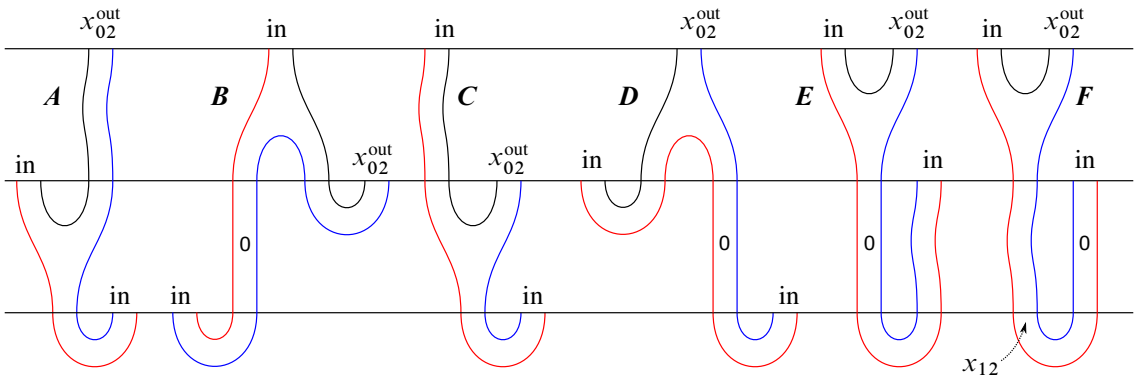


Figure 7: Pseudoholomorphic buildings in the boundary of (11) and (12).

boundary of the compactification of the following products of moduli spaces:

$$\mathcal{M}^1(x_{02};\delta_0,\gamma_{10},\delta_1,\beta_{12},\delta_2')\times\mathcal{M}^0(\beta_{12};\delta_1',\gamma_{21},\delta_2), \tag{11}$$

$$\mathcal{M}^0(x_{02};\delta_0,\gamma_{10},\delta_1,\beta_{12},\delta_2')\times\mathcal{M}^1(\beta_{12};\delta_1',\gamma_{21},\delta_2), \tag{12}$$

$$\mathcal{M}^1(\gamma_{20};\delta_0,\gamma_{10},\delta_1,\beta_{12},\delta_2')\times\mathcal{M}^0(\beta_{12};\delta_1',\gamma_{21},\delta_2), \tag{13}$$

$$\mathcal{M}^0(\gamma_{20};\delta_0,\gamma_{10},\delta_1,\beta_{12},\delta_2')\times\mathcal{M}^1(\beta_{12};\delta_1',\gamma_{21},\delta_2). \tag{14}$$

See Figure 7 for buildings in the boundary of the compactification of (11) and (12), and Figure 8 for those in the boundary of the compactification of (13) and (14). In each case (and in all the rest of the proof), we omit to draw the buildings which appear twice, ie both in the boundary of (11) and (12) or both in the boundary of (13) and (14), because they cancel each other over $\mathbb{Z}_2$.

According to Lemma 6.3(c), the building $\boldsymbol{A}$ either never appears or appears twice, so we can ignore it.

Consider now the building $\boldsymbol{B}$. Removing the banana containing the output x_{02} puncture gives a building contributing to the component of $\widehat{\mathfrak{m}}_2(\gamma_{21},\gamma_{10})$ taking values in the generators of $\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_2)$ of the form $\beta_{20}\,\delta_0\,\delta_1\,\delta_2$. Adding the banana with positive asymptotic at x_{02} and using Lemma 6.3(b), we

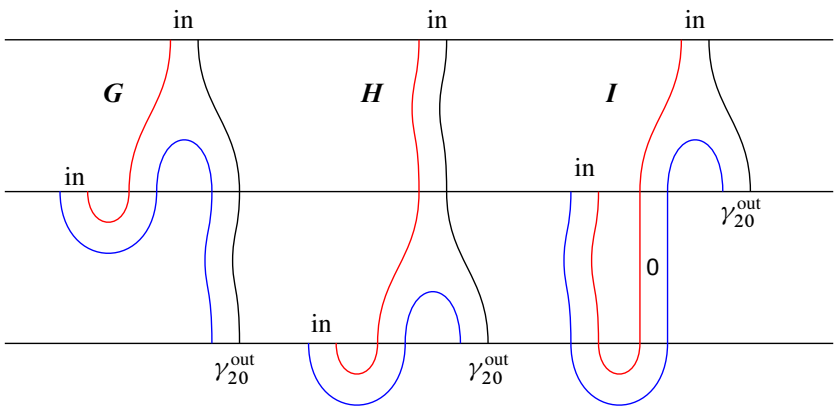


Figure 8: Pseudoholomorphic buildings in the boundary of (13) and (14).

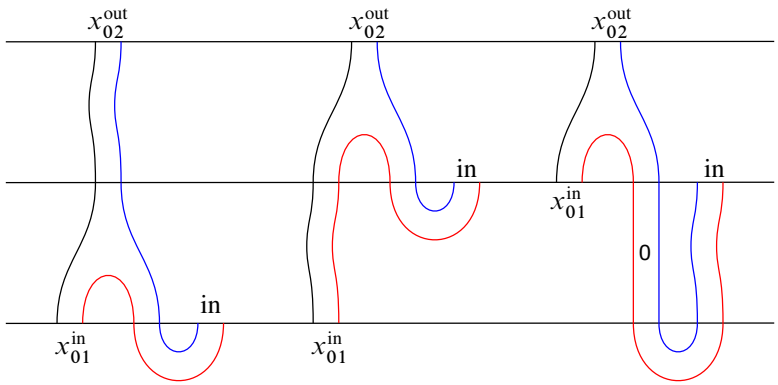


Figure 9: Pseudoholomorphic buildings in the boundary of (15) and (16).

get that the building contributes $x_{02}\beta\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2 + x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2\,\beta$, where β is the pure Reeb chord of Λ corresponding to β_{01}. Thus, this type $\boldsymbol{B}$ of buildings, together with $\boldsymbol{G}$, contribute to $\widehat{\mathfrak{m}}_1 \circ \widehat{\mathfrak{m}}_2(\gamma_{21}, \gamma_{10})$.

The buildings of type $\boldsymbol{C}$ and $\boldsymbol{D}$, together with $\boldsymbol{H}$ contribute to $\widehat{\mathfrak{m}}_2(\gamma_{21}, \widehat{\mathfrak{m}}_1(\gamma_{10}))$. Indeed the only slightly subtle thing here is about building $\boldsymbol{D}$. By assumption, the mixed connecting chord from Λ_2 to Λ_1 is not the maximum Morse Reeb chord x_{12}. But then, if such a building exists this connecting chord must be the minimum Morse chord y_{12} for action reasons, because the component of the building with three mixed asymptotics has a positive asymptotic at the Morse chord x_{02}. For degree reasons, the connecting mixed chord from Λ_1 to Λ_0 must then be x_{01} and the banana having it as a positive asymptotic contributes to the component of $\widehat{\mathfrak{m}}_1(\gamma_{01})$ taking values in $\langle x_{01}\rangle^{\text{cyc}}_{\mathcal{A}\text{-}\mathcal{A}}$ (this component actually vanishes for elements $\gamma_{10}\boldsymbol{a}$ where $\boldsymbol{a} = 1$, but it doesn't in the general case).

Finally, let us consider the building $\boldsymbol{F}$. Usually, these types of buildings cancel by pairs but because of the assumption about the connecting chord, the buildings of type $\boldsymbol{F}$ arise by degenerating the banana but will never arise by degenerating the top level with three mixed asymptotic. Observe then that the buildings $\boldsymbol{E}$, $\boldsymbol{F}$ and $\boldsymbol{I}$ contribute to $\widehat{\mathfrak{m}}_2(\widehat{\mathfrak{m}}_1(\gamma_{21}), \gamma_{10})$, and we have thus proved the Leibniz rule for the pair of inputs $(\gamma_{21}, \gamma_{10})$.

For a pair of inputs (γ_{21}, x_{01}), the Leibniz rule restricts to $\widehat{\mathfrak{m}}_2(\widehat{\mathfrak{m}}_1(\gamma_{21}), x_{01}) = 0$ because $\widehat{\mathfrak{m}}_1$ vanishes on the maximum Reeb chord. Let us consider the pseudoholomorphic buildings in the boundary of the compactification of the following products of moduli spaces:

$$\mathcal{M}^1(x_{02}; \boldsymbol{\delta}_0, x_{01}, \boldsymbol{\delta}_1, \beta_{12}, \boldsymbol{\delta}_2') \times \mathcal{M}^0(\beta_{12}; \boldsymbol{\delta}_1', \gamma_{21}, \boldsymbol{\delta}_2), \tag{15}$$

$$\mathcal{M}^0(x_{02}; \boldsymbol{\delta}_0, x_{01}, \boldsymbol{\delta}_1, \beta_{12}, \boldsymbol{\delta}_2') \times \mathcal{M}^1(\beta_{12}; \boldsymbol{\delta}_1', \gamma_{21}, \boldsymbol{\delta}_2). \tag{16}$$

See Figure 9 for a schematic picture of these buildings. As before, the first building can be ignored. The second building never appears either, for action reasons. The third building finally contributes to $\widehat{\mathfrak{m}}_2(\widehat{\mathfrak{m}}_1(\gamma_{21}), x_{01})$. For the pair of inputs (x_{12}, γ_{10}), the Leibniz rule restricts to

$$\widehat{\mathfrak{m}}_1 \circ \widehat{\mathfrak{m}}_2(x_{12}, \gamma_{10}) + \widehat{\mathfrak{m}}_2(x_{12}, \widehat{\mathfrak{m}}_1(\gamma_{10})) = 0.$$

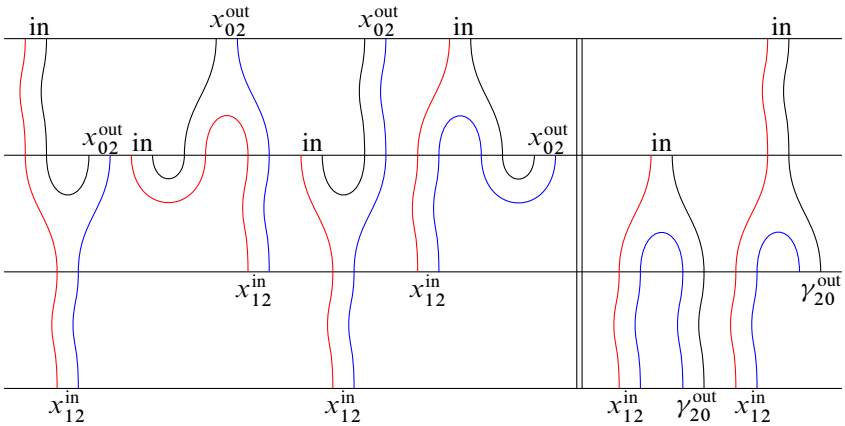


Figure 10: Left: pseudoholomorphic buildings in the boundary of (17) and (18); right: pseudoholomorphic buildings in the boundary of (19) and (20).

We consider the boundary of the compactification of

$$\mathcal{M}^1(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\times\mathcal{M}^0(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2), \tag{17}$$

$$\mathcal{M}^0(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\times\mathcal{M}^1(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2) \tag{18}$$

and of

$$\mathcal{M}^1(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\times\mathcal{M}^0(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2), \tag{19}$$

$$\mathcal{M}^0(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\beta_{12},\boldsymbol{\delta}_2')\times\mathcal{M}^1(\beta_{12};\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2), \tag{20}$$

whose different components are schematized in Figure 10. By arguments similar as before, the algebraic contributions of the second and third buildings vanish. The first and sixth contribute to $\widehat{\mathfrak{m}}_2(x_{12},\widehat{\mathfrak{m}}_1(\gamma_{10}))$, and the fourth and fifth to $\widehat{\mathfrak{m}}_1\circ\widehat{\mathfrak{m}}_2(x_{12},\gamma_{10})$. Finally, for a pair (x_{12},x_{01}) all the terms of the Leibniz rule vanish by definition. □

Remark 7.4 The product on $\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_1)$ is defined by a count of 2-level pseudoholomorphic buildings, which arise in the boundary of the compactification of 1-dimensional moduli spaces. In particular, there is no canonical choice for the buildings we choose to count to define the product, but there is a choice up to homotopy. In particular, we could choose the product to be the map

$$\widehat{\boldsymbol{d}}_2\colon\widehat{C}_+^{\text{cyc}}(\Lambda_1,\Lambda_2)\otimes\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_1)\to\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_2)$$

defined by

$$\begin{aligned}\widehat{\boldsymbol{d}}_2(\gamma_{21},\gamma_{10})=&\sum_{\beta_{01}\in\mathcal{R}^I(\Lambda_0,\Lambda_1)\cup\{y_{01}\}}\ \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_0',\\ \boldsymbol{\delta}_1,\boldsymbol{\delta}_1',\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,\beta_{01},\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0',\gamma_{10},\boldsymbol{\delta}_1)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_0'\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_1'\,\boldsymbol{\delta}_2\\ &+\sum_{\gamma_{20}}\sum_{\beta_{01}\in\mathcal{R}^I(\Lambda_0,\Lambda_1)\cup\{y_{01}\}}\ \sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_0',\\ \boldsymbol{\delta}_1,\boldsymbol{\delta}_1',\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{20};\boldsymbol{\delta}_0,\beta_{01},\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0',\gamma_{10},\boldsymbol{\delta}_1)\\ &\qquad\cdot\gamma_{20}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_0'\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_1'\,\boldsymbol{\delta}_2,\end{aligned}$$

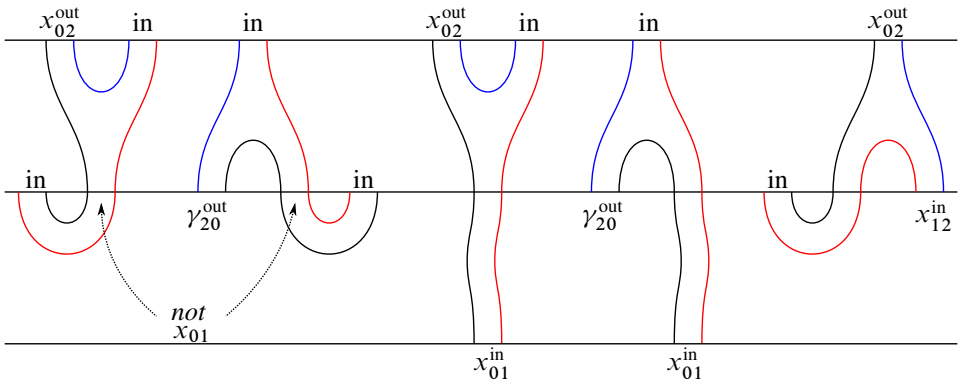


Figure 11: Pseudoholomorphic buildings contributing to $\widehat{\boldsymbol{d}}_2(x_{12},\gamma_{01})$.

$$\widehat{\boldsymbol{d}}_2(\gamma_{21},x_{01})=\sum_{\beta_{01}}\sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_0',\\ \boldsymbol{\delta}_1,\boldsymbol{\delta}_1',\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,\beta_{01},\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0',x_{01},\boldsymbol{\delta}_1)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_0'\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_1'\,\boldsymbol{\delta}_2$$
$$+\sum_{\gamma_{20},\beta_{01}}\sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_0',\\ \boldsymbol{\delta}_1,\boldsymbol{\delta}_1',\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(\gamma_{20};\boldsymbol{\delta}_0,\beta_{01},\boldsymbol{\delta}_1',\gamma_{21},\boldsymbol{\delta}_2)\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0',x_{01},\boldsymbol{\delta}_1)\cdot\gamma_{20}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_0'\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_1'\,\boldsymbol{\delta}_2,$$
$$\widehat{\boldsymbol{d}}_2(x_{12},\gamma_{10})=\sum_{\beta_{01}}\sum_{\substack{\boldsymbol{\delta}_0,\boldsymbol{\delta}_0'\\ \boldsymbol{\delta}_1,\boldsymbol{\delta}_1',\boldsymbol{\delta}_2}}\#\mathcal{M}^0_{\Lambda_{012}}(x_{02};\boldsymbol{\delta}_0,\beta_{01},\boldsymbol{\delta}_1',x_{12},\boldsymbol{\delta}_2)\#\mathcal{M}^0_{\Lambda_{01}}(\beta_{01};\boldsymbol{\delta}_0',\gamma_{10},\boldsymbol{\delta}_1)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_0'\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_1'\,\boldsymbol{\delta}_2,$$
$$\widehat{\boldsymbol{d}}_2(x_{12},x_{01})=0.$$

See Figure 11. We show that the maps $\widehat{\boldsymbol{d}}_2$ and $\widehat{\mathfrak{m}}_2$ are homotopic. Let us define a (degree -1) map $\boldsymbol{h}\colon \widehat{C}_+^{\text{cyc}}(\Lambda_1,\Lambda_2)\otimes\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_1)\to\widehat{C}_+^{\text{cyc}}(\Lambda_0,\Lambda_2)$ by

$$\boldsymbol{h}(\gamma_{21},\gamma_{10})=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}\#\mathcal{M}^0(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2$$
$$+\sum_{\gamma_{20}}\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2}\#\mathcal{M}^0(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2)\cdot\gamma_{20}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2,$$
$$\boldsymbol{h}(\gamma_{21},x_{01})=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}\#\mathcal{M}^0(x_{02};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2$$
$$+\sum_{\gamma_{20}}\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2}\#\mathcal{M}^0(\gamma_{20};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2)\cdot\gamma_{20}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2,$$
$$\boldsymbol{h}(x_{12},\gamma_{10})=\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1,\boldsymbol{\delta}_2}\#\mathcal{M}^0(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,x_{12},\boldsymbol{\delta}_2)\cdot x_{02}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2$$
$$+\sum_{\gamma_{20}}\sum_{\boldsymbol{\delta}_0,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2}\#\mathcal{M}^0(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,x_{12},\boldsymbol{\delta}_2)\cdot\gamma_{20}\,\boldsymbol{\delta}_0\,\boldsymbol{\delta}_1\,\boldsymbol{\delta}_2,$$
$$\boldsymbol{h}(x_{12},x_{01})=0.$$

See Figure 12.

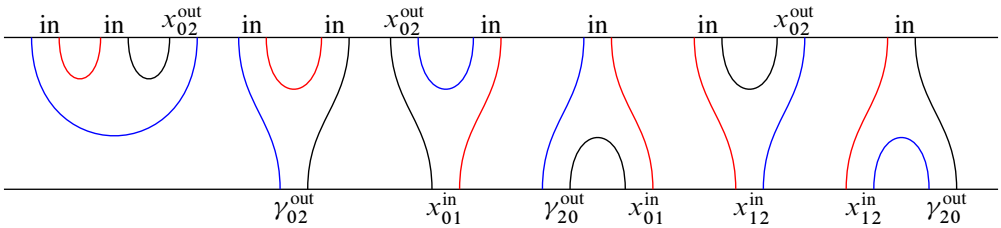


Figure 12: Pseudoholomorphic discs contributing to the map $\boldsymbol{h}$.

Now we check that

$$\widehat{\mathfrak{m}}_2 + \widehat{d}_2 = \boldsymbol{h}(\mathrm{id} \otimes \widehat{\mathfrak{m}}_1) + \boldsymbol{h}(\widehat{\mathfrak{m}}_1 \otimes \mathrm{id}) + \widehat{\mathfrak{m}}_1 \circ \boldsymbol{h} \tag{21}$$

by studying the boundary of the compactification of 1-dimensional moduli spaces of discs as the ones shown in Figure 12. Namely, in order to prove (21) for the pair of generators $(\gamma_{21}, \gamma_{10})$, we study the boundary of the compactification of

$$\mathcal{M}^1(x_{02}; \boldsymbol{\delta}_0, \gamma_{10}, \boldsymbol{\delta}_1, \gamma_{21}, \boldsymbol{\delta}_2) \tag{22}$$

and

$$\mathcal{M}^1(\gamma_{20}; \boldsymbol{\delta}_0, \gamma_{10}, \boldsymbol{\delta}_1, \gamma_{21}, \boldsymbol{\delta}_2). \tag{23}$$

The corresponding pseudoholomorphic buildings are schematized in Figures 13 and 14.

Observe that buildings of type $\boldsymbol{A}$, type $\boldsymbol{G}$ and those of type $\boldsymbol{B}$ and $\boldsymbol{H}$ where the connecting chord is x_{01} (for both types $\boldsymbol{B}$ and $\boldsymbol{H}$) are the ones contributing exactly to $\boldsymbol{h}(\gamma_{21}, \widehat{\mathfrak{m}}_1(\gamma_{10}))$. Similarly, the buildings of types $\boldsymbol{C}$ and $\boldsymbol{I}$ together with the buildings of types $\boldsymbol{D}$ and $\boldsymbol{J}$ where the connecting chord is x_{12} contribute to $\boldsymbol{h}(\widehat{\mathfrak{m}}_1(\gamma_{21}), \gamma_{10})$. The contribution of buildings of type $\boldsymbol{E}$, if such buildings exist, vanishes according to Lemma 6.3(c). The buildings of types $\boldsymbol{F}$ and $\boldsymbol{K}$ contribute to $\widehat{\mathfrak{m}}_1 \circ \boldsymbol{h}(\gamma_{21}, \gamma_{10})$. We are thus counting buildings which algebraically contribute to the right-hand side of (21) for the pair of generators $(\gamma_{21}, \gamma_{10})$. Finally, observe that the buildings of types $\boldsymbol{B}$ and $\boldsymbol{H}$, when the connecting chord is not the Morse chord x_{01}, are exactly the ones contributing to $\widehat{\boldsymbol{d}}_2(\gamma_{21}, \gamma_{10})$, and that those of types $\boldsymbol{D}$

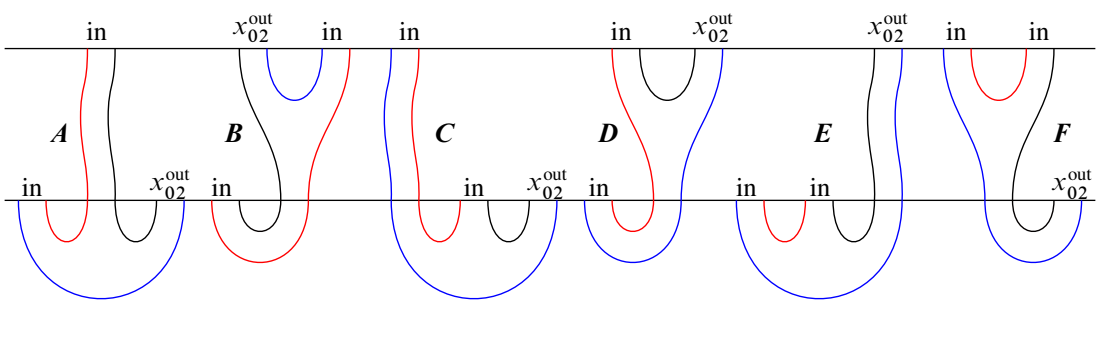


Figure 13: Pseudoholomorphic buildings in the boundary of (22).

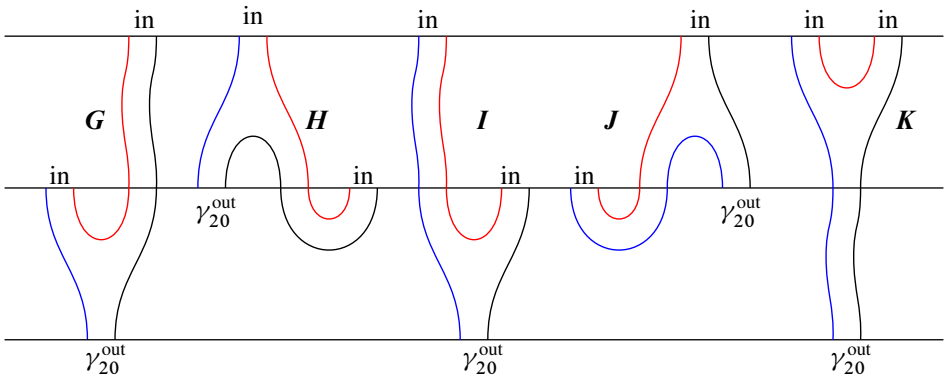


Figure 14: Pseudoholomorphic buildings in the boundary of (23).

and $\boldsymbol{J}$ with the connecting chord not being x_{21} contribute to $\widehat{\mathfrak{m}}_2(\gamma_{21},\gamma_{10})$. These give the left-hand side of (21) and thus we have proved this equation for the pair $(\gamma_{21},\gamma_{10})$.

In order to prove this equation for a pair (γ_{21},x_{01}), we consider the pseudoholomorphic buildings in the boundary of the compactification of

$$\mathcal{M}^1(x_{02};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2), \tag{24}$$

$$\mathcal{M}^1(\gamma_{20};\boldsymbol{\delta}_0,x_{01},\boldsymbol{\delta}_1,\gamma_{21},\boldsymbol{\delta}_2), \tag{25}$$

which we schematized in Figure 15.

The buildings of types $\boldsymbol{A}'$ and $\boldsymbol{F}'$ contribute to $\widehat{\boldsymbol{d}}_2(\gamma_{21},x_{01})$ and the buildings of type $\boldsymbol{C}'$ contribute to $\widehat{\mathfrak{m}}_2(\gamma_{21},x_{01})$ (observe that the connecting chord in this case cannot be x_{12}, for example, for degree reasons). This gives the left-hand side of (21). The other buildings contribute algebraically to the right-hand side. Indeed, buildings of types $\boldsymbol{B}'$and $\boldsymbol{G}'$ contribute to $\boldsymbol{h}(\widehat{\mathfrak{m}}_1(\gamma_{21}),x_{01})$, the contribution of the buildings of type $\boldsymbol{D}'$ vanishes according to Lemma 6.3(c), and buildings of types $\boldsymbol{E}'$ and $\boldsymbol{H}'$ contribute to $\widehat{\mathfrak{m}}_1\circ\boldsymbol{h}(\gamma_{21},x_{01})$. The remaining term in the relation, $\boldsymbol{h}(\gamma_{21},\widehat{\mathfrak{m}}(x_{01}))$, vanishes because $\widehat{\mathfrak{m}}_1(x_{01})=0$.

We continue by proving (21) for a pair (x_{12},γ_{10}). We study the boundary of the compactification of

$$\mathcal{M}^1(x_{02};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,x_{12},\boldsymbol{\delta}_2), \tag{26}$$

$$\mathcal{M}^1(\gamma_{20};\boldsymbol{\delta}_0,\gamma_{10},\boldsymbol{\delta}_1,x_{12},\boldsymbol{\delta}_2) \tag{27}$$

for which the pseudoholomorphic buildings are schematized in Figure 16.

Observe first that the contribution of the buildings of type $\boldsymbol{B}''$ vanishes as before. Then, buildings of type $\boldsymbol{C}''$, $\boldsymbol{D}''$ and $\boldsymbol{G}''$ contribute to $(\widehat{\mathfrak{m}}_2+\widehat{d}_2)(x_{12},\gamma_{10})$. The buildings of type $\boldsymbol{A}''$ and $\boldsymbol{F}''$ are exactly the ones contributing to $\boldsymbol{h}(x_{12},\widehat{\mathfrak{m}}_1(\gamma_{10}))$ (indeed there is also a component of $\widehat{\mathfrak{m}}_1(\gamma_{10})$ with output the Morse chord x_{01} but then $h(x_{12},x_{01})=0$ so these do not contribute). The buildings of type $\boldsymbol{E}''$ and $\boldsymbol{H}''$ are the ones contributing to $\widehat{\mathfrak{m}}_1\circ h(x_{12},\gamma_{10})$, and we get thus (21) because $\boldsymbol{h}(\widehat{\mathfrak{m}}(x_{12}),\gamma_{10})=0$.

Finally, (21) is trivially satisfied for the pair of inputs (x_{12},x_{01}) because both sides vanish.

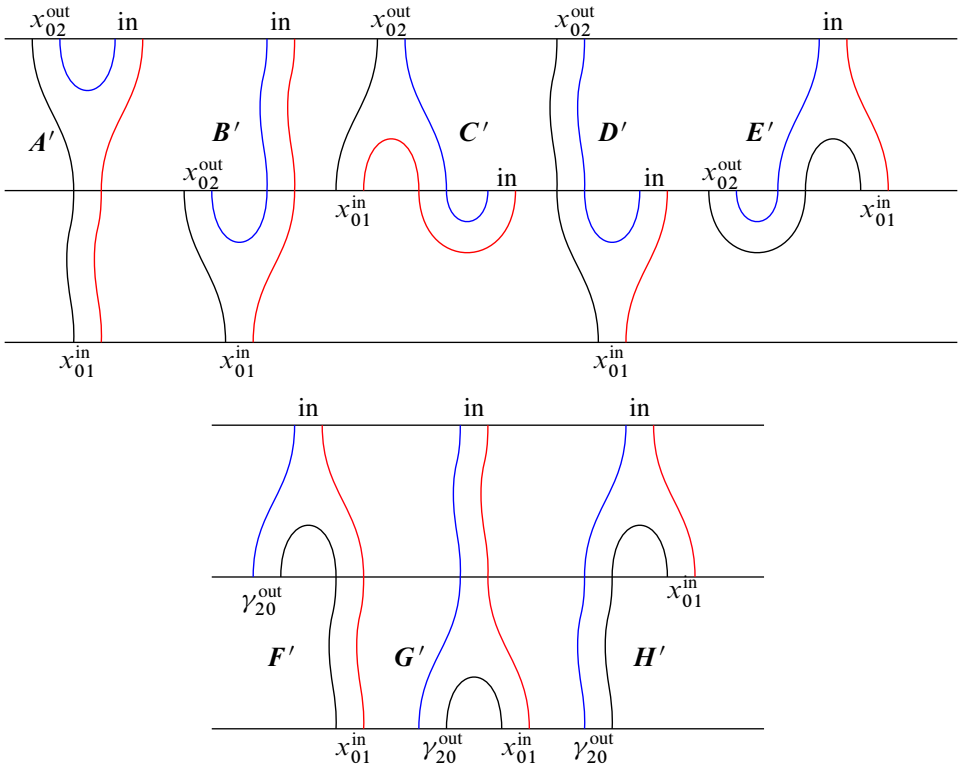


Figure 15: Pseudoholomorphic buildings in the boundary of (24) (top) and (25) (bottom).

Remark 7.5 The pseudoholomorphic buildings contributing to the component of the product $\widehat{\mathfrak{m}}_2$ with long chords as inputs and output can also be used to build a product structure on the linearized Legendrian contact *homology* complex, which will be unital when the Legendrian submanifold is horizontally displaceable.

It is well known that there is a (nonunital) product structure (even an A_∞-structure) on the linearized Legendrian contact cohomology, defined first in [11] and generalized to the bilinearized case in [1]. This product can be computed directly from the C-E algebra, and equivalently by a count of pseudoholomorphic discs with boundary on a 3-copy $\Lambda_0 \cup \Lambda_1 \cup \Lambda_2$ of Λ negatively asymptotic to a chord from Λ_0 to Λ_1 and to a chord from Λ_1 to Λ_2 (the inputs), and positively asymptotic to a chord from Λ_0 to Λ_2 (the output), and potentially having pure negative asymptotics which are augmented. In this case, the mixed chords considered are never Morse chords.

In [32], for knots in $\mathbb{R}^3$ the authors consider the complex generated by mixed chords from Λ_1 to Λ_0 in a 2-copy of Λ, which they denote by $\mathrm{Hom}_+(\varepsilon_0, \varepsilon_1)$. They define then a product (as well as an A_∞-structure) on this complex by a count of similar curves as above (two negative mixed inputs and one positive mixed output) but the main difference is that in this case the Morse chords can be inputs and outputs. The homology of the complex $\mathrm{Hom}_+(\varepsilon_0, \varepsilon_1)$ is isomorphic to the Legendrian contact homology of Λ. This is implied by the acyclicity of the complex of the 2-copy which holds for knots

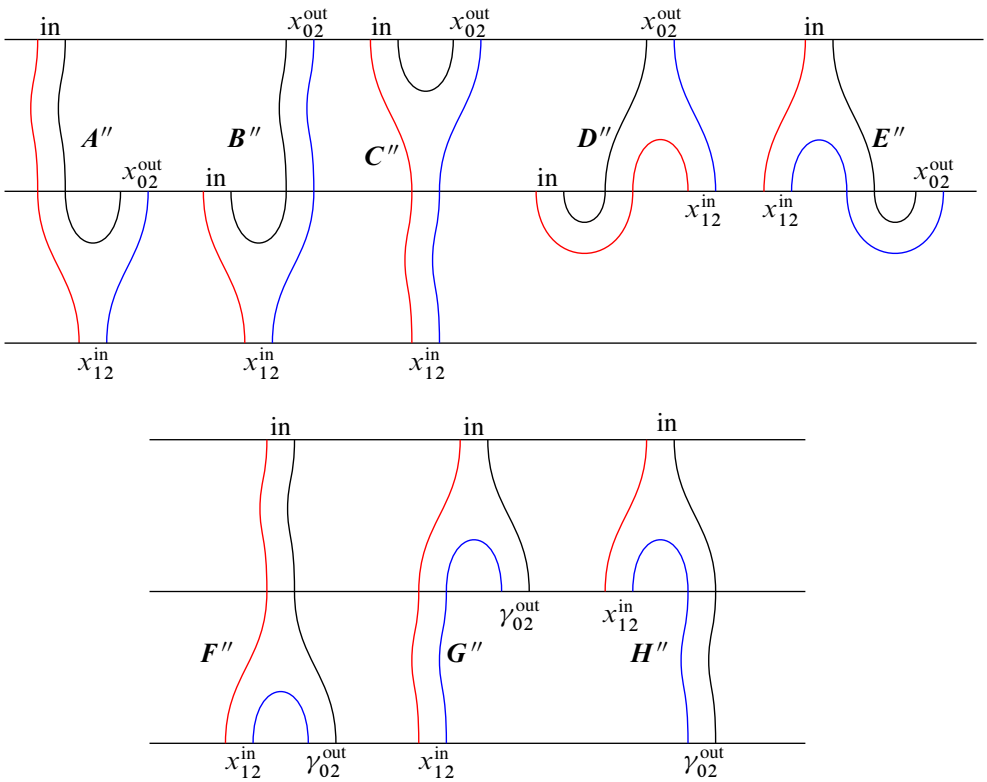


Figure 16: Pseudoholomorphic buildings in the boundary of (26) (top) and (27) (bottom).

in $\mathbb{R}^3$; see [32, Proposition 5.4]. In particular, the product on Hom_+ doesn't give canonically a product on Legendrian contact homology.

By [4, Proposition 2.7], the (bi)linearization of the DG bimodule $(C_+(\Lambda_0,\Lambda_1), \mathbf{\Delta}_1)$ by augmentations ε_0 and ε_1 is canonically isomorphic to the bilinearized Legendrian contact homology complex $\mathrm{LCC}_*^{\varepsilon_0,\varepsilon_1}(\Lambda)$. We claim that the only-long-chords-asymptotics component of the product $\widehat{\mathfrak{m}}_2$, when linearized by augmentations, gives a product on the Legendrian contact homology complex, the proof of this being schematized in Figure 8. In the presence of a filling of Λ and under the hypothesis of horizontal displaceability, it is possible to prove that this new product on Legendrian contact homology is isomorphic to the product on the Hom_+ complex, because both are isomorphic to the product on the singular cohomology of the filling (through the Ekholm–Seidel isomorphism [13; 16]). This will be investigated more precisely in a forthcoming paper with Georgios Dimitroglou Rizell, where we describe a *relative Calabi–Yau structure* carried by the C-E algebra. This relative structure allows in particular to show the isomorphism between the products on LCH_* and Hom_+ without the presence of a filling.

7.3 Products under the Calabi–Yau morphism

Let us consider the Calabi–Yau map CY induced on the $\mathbb{Z}_2$-modules, denoted by

$$\mathrm{CY}_1 \colon \widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1) \to \check{C}_-^{\mathrm{cyc}}(\Lambda_0,\Lambda_1).$$

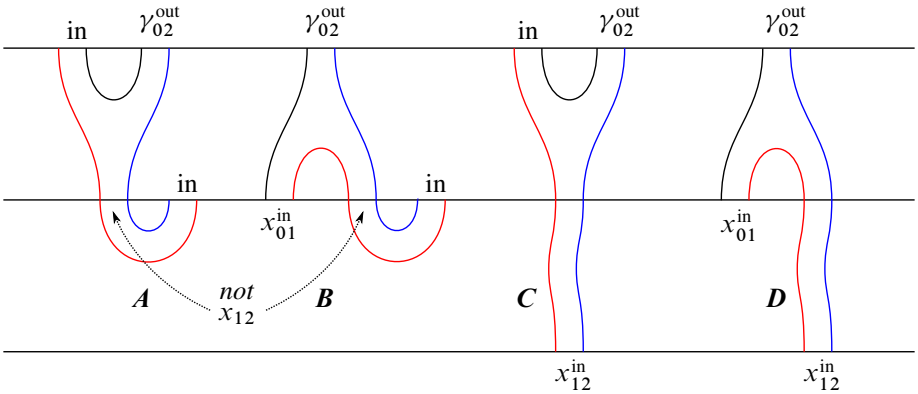


Figure 17: Pseudoholomorphic buildings contributing to the map CY_2.

Theorem 7.6 *The map* CY_1 *preserves the product structures in homology, ie*

$$\mathrm{CY}_1 \circ \widehat{\mathfrak{m}}_2 + \check{\mathfrak{m}}_2(\mathrm{CY}_1, \mathrm{CY}_1) = 0$$

is satisfied in homology.

Proof In this proof we will use the notation $\widehat{\mathfrak{m}}_2^+$ and $\widehat{\mathfrak{m}}_2^x$ to denote the components of $\widehat{\mathfrak{m}}_2$ with values in $C_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ and $\langle x_{01} \rangle_{\mathcal{A}\text{-}\mathcal{A}}^{\mathrm{cyc}}$, respectively.

Given a 3-copy $\Lambda_0 \cup \Lambda_1 \cup \Lambda_2$ of Λ, we define a (degree -1) map

$$\mathrm{CY}_2 \colon \widehat{C}_+^{\mathrm{cyc}}(\Lambda_1, \Lambda_2) \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_2)$$

by a count of pseudoholomorphic buildings as shown in Figure 17. Similarly as for the product $\widehat{\mathfrak{m}}_2$ we require that the connecting chord from Λ_2 to Λ_1 in the buildings is not the maximum Morse Reeb chord x_{12}. Observe however that this is automatically satisfied for action reasons for the buildings of types $\boldsymbol{C}$ and $\boldsymbol{D}$.

By considering buildings in the boundary of the compactification of one-dimensional products of moduli spaces of types $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{D}$ in Figure 17, one proves

$$\check{\mathfrak{m}}_1 \circ \mathrm{CY}_2 + \check{\mathfrak{m}}_2(\mathrm{CY}_1 \otimes \mathrm{CY}_1) + \mathrm{CY}_1 \circ \widehat{\mathfrak{m}}_2 + \mathrm{CY}_2(\mathrm{id} \otimes \widehat{\mathfrak{m}}_1) + \mathrm{CY}_2(\widehat{\mathfrak{m}}_1 \otimes \mathrm{id}) = 0, \tag{28}$$

which shows that CY_1 preserves the products in homology. In Figure 18 we depicted the different types of buildings in the boundary of the compactification of 1-dimensional products of moduli spaces of type $\boldsymbol{A}$. The buildings of type $\boldsymbol{A1}$ together with those of type $\boldsymbol{A3}$, where the connecting chord from Λ_1 to Λ_0 is x_{01}, contribute to $\mathrm{CY}_2(\gamma_{21}, \widehat{\mathfrak{m}}_1(\gamma_{10}))$. Those of type $\boldsymbol{A2}$ contribute to $\mathrm{CY}_1 \circ \widehat{\mathfrak{m}}_2^x(\gamma_{21}, \gamma_{10})$ when the connecting chord from Λ_2 to Λ_0 is x_{02}, and to $\check{\mathfrak{m}}_1 \circ \mathrm{CY}_2(\gamma_{21}, \gamma_{10})$ otherwise. The buildings of type $\boldsymbol{A3}$ when the connecting chord from Λ_1 to Λ_0 is not x_{01} contribute to $\check{\mathfrak{m}}_2(\mathrm{CY}_1(\gamma_{21}), \mathrm{CY}_1(\gamma_{10}))$. Those of type $\boldsymbol{A4}$ contribute to $\mathrm{CY}_1 \circ \widehat{\mathfrak{m}}_2^+(\gamma_{21}, \gamma_{10})$ and finally those of types $\boldsymbol{A5}$ and $\boldsymbol{A6}$ contribute to $\mathrm{CY}_2(\widehat{\mathfrak{m}}_1(\gamma_{21}), \gamma_{10})$. This gives (28) for the pair of inputs $(\gamma_{21}, \gamma_{10})$. In Figure 19 we consider the broken discs in the boundary of 1-dimensional products of moduli spaces of type $\boldsymbol{B}$. The buildings $\boldsymbol{B1}$ contribute to $\mathrm{CY}_1 \circ \widehat{\mathfrak{m}}_2^x(\gamma_{21}, x_{01})$

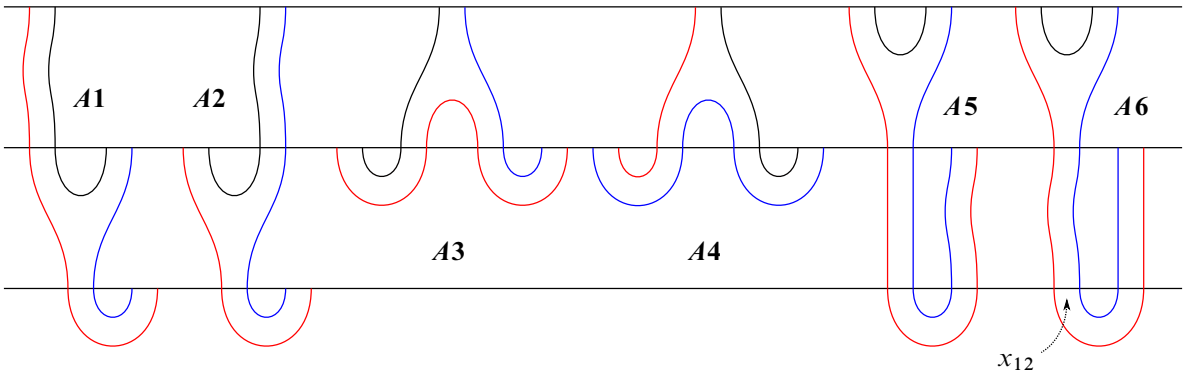


Figure 18: Pseudoholomorphic buildings in the boundary of 1-dimensional products of moduli spaces of type $\boldsymbol{A}$.

when the connecting chord from Λ_2 to Λ_0 is the maximum Morse chord x_{02}, and to $\check{\mathfrak{m}}_1 \circ \mathrm{CY}_2(\gamma_{21}, x_{01})$ otherwise. Those of type $\boldsymbol{B2}$ contribute to $\check{\mathfrak{m}}_2(\mathrm{CY}_1(\gamma_{21}), \mathrm{CY}_1(x_{01}))$ (none of the connecting chord is a maximum).The buildings of types $\boldsymbol{B3}$ and $\boldsymbol{B4}$ finally contribute to $\mathrm{CY}_2(\widehat{\mathfrak{m}}_1(\gamma_{21}), x_{01})$. The sum of these contributions gives (28) for the pair of inputs (γ_{21}, x_{01}) (observe that some terms in the relation vanish by definition). Similarly this relation can be checked for pairs of inputs (x_{12}, γ_{10}) and (x_{12}, x_{01}) by considering broken discs in the boundary of 1-dimensional products of moduli spaces of type $\boldsymbol{C}$ and $\boldsymbol{D}$, respectively; see Figure 20. □

From Theorem 7.6, we deduce that the product $\widehat{\mathfrak{m}}_2$ has a unit in homology represented by any cycle in $\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ which is sent to the minimum Morse Reeb chord y_{01} by the map CY_1.

7.4 A_∞-structure

We can go further and define for each $d \geq 3$ maps $\check{\mathfrak{m}}_d$ and $\widehat{\mathfrak{m}}_d$ of degree $2-d$ by counting pseudoholomorphic discs with boundary on a $(d+1)$-copy of Λ (defined in an analogous way as the 2- and 3-copies).

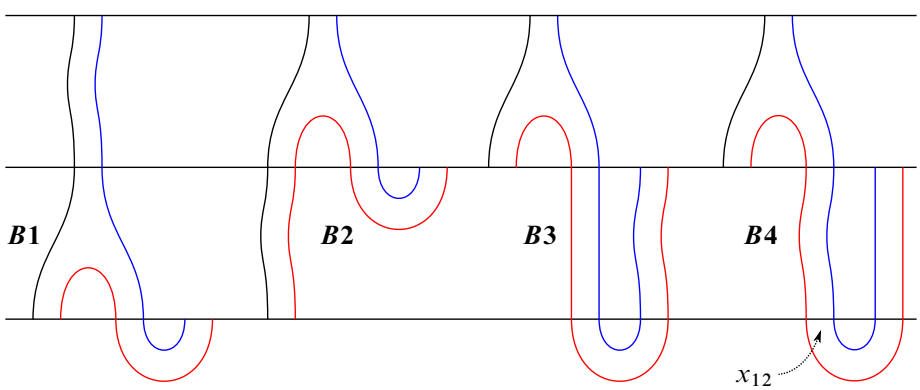


Figure 19: Pseudoholomorphic buildings in the boundary of 1-dimensional products of moduli spaces of type $\boldsymbol{B}$.

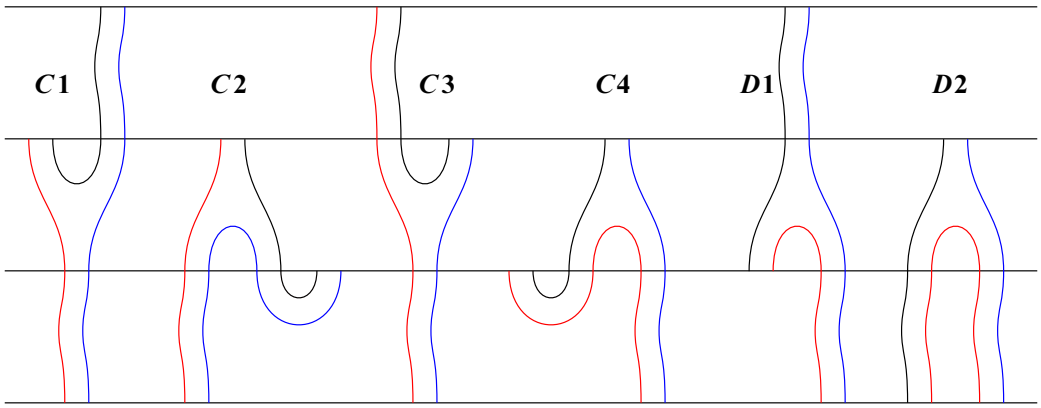


Figure 20: Pseudoholomorphic buildings in the boundary of 1-dimensional products of moduli spaces of type $\boldsymbol{C}$ and $\boldsymbol{D}$.

Given a $(d+1)$-copy $\Lambda_0 \cup \cdots \cup \Lambda_d$ of Λ, the maps

$$\check{\mathfrak{m}}_d \colon \check{C}_-^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_d)$$

are given by a count of pseudoholomorphic discs having mixed negative asymptotics corresponding to inputs and one positive asymptotic which is the output. The map $\check{\mathfrak{m}}_d$ has degree $2-d$. The fact that these maps satisfy the A_∞-equations goes back to [1; 11]. In our case we have to take extra care of the coefficients in the C-E algebra but it works exactly the same as in the case $d = 2$ we treated in Section 7.2.

Now let's define the maps $\widehat{\mathfrak{m}}_d$. First, we extend the maps $\boldsymbol{b}_1$ and $\boldsymbol{\Delta}_1$ to higher-order maps

$$\boldsymbol{b}_d, \boldsymbol{\Delta}_d \colon \mathrm{RFC}^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1) \to \mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0, \Lambda_d)$$

for $d \geq 1$ as follows. These maps have degree $2-d$ and were considered in [29, Section 8]; we recall the definitions. For a d-tuple of elements $(c_{d-1}\boldsymbol{a}_{d-1}, \ldots, c_0\boldsymbol{a}_0) \in \mathrm{RFC}^{\mathrm{cyc}}(\Lambda_{d-1}, \Lambda_d) \otimes \cdots \otimes \mathrm{RFC}^{\mathrm{cyc}}(\Lambda_0, \Lambda_1)$ where c_j are mixed chords and $\boldsymbol{a}_j$ words of pure Reeb chords, set

$$\boldsymbol{b}_d(c_{d-1}\boldsymbol{a}_{d-1}, \ldots, c_0\boldsymbol{a}_0) = \sum_{\substack{\gamma_{0d} \\ \boldsymbol{\delta}_0, \ldots, \boldsymbol{\delta}_d}} \#\mathcal{M}^0_{\Lambda_{0\ldots d}}(\gamma_{0d}; \boldsymbol{\delta}_0, c_0, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_{d-1}, c_{d-1}, \boldsymbol{\delta}_d) \cdot \gamma_{0d}\, \boldsymbol{\delta}_0 \boldsymbol{a}_0 \boldsymbol{\delta}_1 \ldots \boldsymbol{a}_{d-1} \boldsymbol{\delta}_d,$$

$$\boldsymbol{\Delta}_d(c_{d-1}\boldsymbol{a}_{d-1}, \ldots, c_0\boldsymbol{a}_0) = \sum_{\substack{\gamma_{d0} \\ \boldsymbol{\delta}_0, \ldots, \boldsymbol{\delta}_d}} \#\mathcal{M}^0_{\Lambda_{0\ldots d}}(\gamma_{d0}; \boldsymbol{\delta}_0, c_0, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_{d-1}, c_{d-1}, \boldsymbol{\delta}_d) \cdot \gamma_{d0}\, \boldsymbol{\delta}_0 \boldsymbol{a}_0 \boldsymbol{\delta}_1 \ldots \boldsymbol{a}_{d-1} \boldsymbol{\delta}_d.$$

Observe that for the map $\boldsymbol{b}_d$, the mixed chord in the output is a positive asymptotic of the pseudoholomorphic discs considered to define it, while for $\boldsymbol{\Delta}_d$ it is a negative asymptotic. Thus, for energy reasons, $\boldsymbol{\Delta}_d(c_{d-1}\boldsymbol{a}_{d-1}, \ldots, c_0\boldsymbol{a}_0)$ is automatically 0 if for all $0 \leq j \leq d-1$ we have $c_j \in C(\Lambda_j, \Lambda_{j+1})$.

Notation 7.7 We will denote by $\boldsymbol{b}_d^x$ the component of $\boldsymbol{b}_d$ which takes values in $\langle x_{0d} \rangle^{\mathrm{cyc}}_{\mathcal{A}\text{-}\mathcal{A}}$, and $\boldsymbol{b}_d^\vee$ for the component which takes values in $\check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_d)$.

We make the following observations:

- As ungraded maps, the map CY_1 is equal to the restriction to $\widehat{C}_+^{\mathrm{cyc}}$ of the map $\boldsymbol{b}_1^\vee$, namely it is defined by a count of bananas with two positive asymptotics.
- The maps $\check{\mathfrak{m}}_d$ are equal to the restriction of $\boldsymbol{b}_d^\vee$ to $\check{C}_-^{\mathrm{cyc}}(\Lambda_{d-1},\Lambda_d)\otimes\cdots\otimes\check{C}_-^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)$.

With this notation we can also rewrite the differential $\widehat{\mathfrak{m}}_1$ and the product $\widehat{\mathfrak{m}}_2$ as follows (we now drop again the full notation for elements, removing the words of pure Reeb chords):

$$\begin{aligned}
\widehat{\mathfrak{m}}_1(\gamma_{10}) &= \boldsymbol{\Delta}_1(\gamma_{10}) + \boldsymbol{b}_1^x(\gamma_{10}),\\
\widehat{\mathfrak{m}}_2(\gamma_{21},\gamma_{10}) &= \boldsymbol{b}_2^x(\mathrm{CY}_1(\gamma_{21}),\gamma_{10}) + \boldsymbol{\Delta}_2(\mathrm{CY}_1(\gamma_{21}),\gamma_{10}),\\
\widehat{\mathfrak{m}}_2(\gamma_{21},x_{01}) &= \boldsymbol{b}_2^x(\mathrm{CY}_1(\gamma_{21}),x_{01}),\\
\widehat{\mathfrak{m}}_2(x_{12},\gamma_{10}) &= \boldsymbol{b}_2^x(\mathrm{CY}_1(x_{12}),\gamma_{10}) + \boldsymbol{\Delta}_2(\mathrm{CY}_1(x_{12}),\gamma_{10}).
\end{aligned}$$

Note that $\boldsymbol{\Delta}_2(\mathrm{CY}_1(\gamma_{21}),x_{01})=0$ for action reasons, as well as

$$\boldsymbol{b}_2^x(x_{12},\mathrm{CY}_1(x_{01})) = \boldsymbol{\Delta}_2(\mathrm{CY}_1(x_{12}),x_{01}) = 0.$$

So we can write a compact formula for the product

$$\widehat{\mathfrak{m}}_2 = (\boldsymbol{b}_2^x + \boldsymbol{\Delta}_2)(\mathrm{CY}_1\otimes\mathrm{id}).$$

Finally, observe that the map CY_2 can be rewritten $\mathrm{CY}_2 = \boldsymbol{b}_2^\vee(\mathrm{CY}_1\otimes\mathrm{id})$.

Remark 7.8 Very rigorously, the map $\widehat{\mathfrak{m}}_1$ has domain and target the $\mathbb{Z}_2$-module $\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)$, so it is actually equal to the sum of the *shifted by* 1 restrictions of the maps $\boldsymbol{\Delta}_1$ and $\boldsymbol{b}_1^x$ (these restrictions have domain $\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)[-1]$). So for the formulas for $\widehat{\mathfrak{m}}_1,\widehat{\mathfrak{m}}_2,\mathrm{CY}_1,\mathrm{CY}_2$ as well as for the maps we define below, the reader should consider these equalities as equalities of ungraded maps.

We extend now these formulas, ie we define maps

$$\begin{aligned}
\widehat{\mathfrak{m}}_d &: \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{d-1},\Lambda_d)\otimes\cdots\otimes\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)\to\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_d),\\
\mathrm{CY}_d &: \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{d-1},\Lambda_d)\otimes\cdots\otimes\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)\to\check{C}_-^{\mathrm{cyc}}(\Lambda_0,\Lambda_d)
\end{aligned}$$

for a $(d+1)$-copy of Λ by

$$\widehat{\mathfrak{m}}_d = \sum_{j=2}^{d}\sum_{\substack{1\le i_2,\dots,i_j\le d-1\\ i_2+\cdots+i_j=d-1}} (\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}), \tag{29}$$

$$\mathrm{CY}_d = \sum_{j=2}^{d}\sum_{\substack{1\le i_2,\dots,i_j\le d-1\\ i_2+\cdots+i_j=d-1}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}). \tag{30}$$

Observe that $\widehat{\mathfrak{m}}_d$ is of degree $2-d$ while CY_d is of degree $1-d$.

Theorem 7.9 *Let $\Lambda_0 \cup \cdots \cup \Lambda_d$ be a $(d+1)$-copy of Λ_0. Then for any $1 \le k \le d$ and any $(k+1)$-tuple of integers $0 \le s_0 < \cdots < s_k \le d$ we have*

$$\sum_{m=1}^{k} \sum_{n=0}^{k-m} \widehat{\mathfrak{m}}_{k-m+1}(\mathrm{id}^{\otimes k-m-n} \otimes \widehat{\mathfrak{m}}_m \otimes \mathrm{id}^{\otimes n}) = 0, \tag{31}$$

$$\sum_{r=1}^{k} \sum_{\substack{t_1,\dots,t_r \\ t_1+\cdots+t_r=k}} \check{\mathfrak{m}}_r(\mathrm{CY}_{t_r} \otimes \cdots \otimes \mathrm{CY}_{t_1}) + \sum_{m=1}^{k} \sum_{n=0}^{k-m} \mathrm{CY}_{k-m+1}(\mathrm{id}^{\otimes k-m-n} \otimes \widehat{\mathfrak{m}}_m \otimes \mathrm{id}^{\otimes n}) = 0, \tag{32}$$

where

- *$\widehat{\mathfrak{m}}_m$ has domain $\widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{n+m-1}}, \Lambda_{s_{n+m}}) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_n}, \Lambda_{s_{n+1}})$,*
- *$\widehat{\mathfrak{m}}_{k-m+1}$ and CY_{k-m+1} have domain*

$$\widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{k-1}}, \Lambda_{s_k}) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{n+m}}, \Lambda_{s_{n+m+1}}) \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{n-1}}, \Lambda_{s_n}) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_1}, \Lambda_{s_0}),$$

- *if we define $\tau_j = \sum_{i=1}^{j} t_i$, then CY_{t_j} has domain*

$$\widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{\tau_j-1}}, \Lambda_{s_{\tau_j}}) \otimes \cdots \otimes \widehat{C}_+^{\mathrm{cyc}}(\Lambda_{s_{\tau_{j-1}}}, \Lambda_{s_{\tau_{j-1}+1}}),$$

and $\check{\mathfrak{m}}_r$ has domain

$$\check{C}_-^{\mathrm{cyc}}(\Lambda_{s_{\tau_{r-1}}}, \Lambda_{s_{\tau_r}}) \otimes \cdots \otimes \check{C}_-^{\mathrm{cyc}}(\Lambda_{s_{\tau_1}}, \Lambda_{s_{\tau_2}}) \otimes \check{C}_-^{\mathrm{cyc}}(\Lambda_{s_{\tau_1}}, \Lambda_{s_0}).$$

To simplify notation, in the following we will assume that the $(k+1)$-tuple of integers $(s_0, \dots, s_k)$ is $(0, \dots, k)$. In order to prove Theorem 7.9 we will use results proved in [29] about the maps $\boldsymbol{b}_d$ and $\boldsymbol{\Delta}_d$ that we recall now.

Lemma 7.10 [29, Lemmas 3 and 5] *Let $\Lambda_0 \cup \cdots \cup \Lambda_d$ be a $(d+1)$-copy of Λ_0. Then for any $1 \le k \le d$ we have*

(a) $$\sum_{m=1}^{k} \sum_{n=0}^{k-m} \boldsymbol{b}_{k-m+1}(\mathrm{id}^{\otimes d-m-n} \otimes (\boldsymbol{b}_m + \boldsymbol{\Delta}_m) \otimes \mathrm{id}^{\otimes n}) = 0,$$

(b) $$\sum_{m=1}^{k} \sum_{n=0}^{k-m} \boldsymbol{\Delta}_{k-m+1}(\mathrm{id}^{\otimes d-m-n} \otimes (\boldsymbol{b}_m + \boldsymbol{\Delta}_m) \otimes \mathrm{id}^{\otimes n}) = 0.$$

Proof of Theorem 7.9 In order to avoid any useless complicated notation and use Lemma 7.10 as it is, we prove the theorem ignoring the grading of maps. This means that we will write $\boldsymbol{b}_1^{\vee}$ (restricted to the appropriate module) for CY_1 and $\boldsymbol{\Delta}_1 + \boldsymbol{b}_1^x$ for $\widehat{\mathfrak{m}}_1$. We start by proving (32) which is a proof by induction. For $k = 1, 2$ we have already shown the relation in Lemma 6.5 and the proof of Theorem 7.6, respectively. Now let us prove (32) for $k \ge 3$ inputs, assuming that the relation holds for any number

of inputs less or equal to $k-1$. We denote by LHS(32) the left-hand side of (32). Using the fact that $\check{\mathfrak{m}}_r = \boldsymbol{b}_r^\vee$ and the formula for the Calabi–Yau map, we have

$$\mathrm{LHS}(32) = \sum_{j=1}^{k} \sum_{\substack{t_1,\dots,t_j\geq 1\\ t_1+\cdots+t_j=k}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_1}) + \boldsymbol{b}_1^\vee\circ\widehat{\mathfrak{m}}_k + \sum_{m=1}^{k-1}\sum_{n=0}^{k-m}\sum_{j=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})(\mathrm{id}^{\otimes k-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n}).$$

We separate the case $n=0$ from the others in the second line and get

$$\sum_{j=1}^{k} \sum_{\substack{t_1,\dots,t_j\geq 1\\ t_1+\cdots+t_j=k}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_1}) + \boldsymbol{b}_1^\vee\circ\widehat{\mathfrak{m}}_k + \sum_{m=1}^{k-1}\sum_{j=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\widehat{\mathfrak{m}}_m)$$
$$+ \sum_{m=1}^{k-1}\sum_{n=1}^{k-m}\sum_{j=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})(\mathrm{id}^{\otimes k-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n}).$$

We now separate the cases $t_1=1$ and $m=1$ from the others in the first line, and apply a change of variables in the second line (note that for any $2\leq s\leq j$, when $1+\sum_{v=2}^{s-1} i_v \leq n \leq 1+\sum_{v=2}^{s} i_v$, the "inner" $\widehat{\mathfrak{m}}_m$ will be an argument of CY_{i_s}; thus instead of summing over $1\leq n\leq k-m$ we can sum over the variables s, n with $2\leq s\leq j$ and $0\leq n\leq i_s-1$) to obtain

$$\sum_{j=2}^{k} \sum_{\substack{t_2,\dots,t_j\geq 1\\ t_2+\cdots+t_j=k-1}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{CY}_1) + \boldsymbol{b}_1^\vee\circ\widehat{\mathfrak{m}}_k + \sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\widehat{\mathfrak{m}}_1)$$
$$+ \sum_{j=1}^{k} \sum_{\substack{t_1,\dots,t_j\geq 1\\ t_1\geq 2\\ t_1+\cdots+t_j=k}} \boldsymbol{b}_j^\vee(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_1}) + \sum_{m=2}^{k-1}\sum_{r=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_r\geq 1\\ i_2+\cdots+i_r=k-m}} \boldsymbol{b}_r^\vee(\mathrm{CY}_{i_r}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\widehat{\mathfrak{m}}_m)$$
$$+ \sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{s=2}^{j}\sum_{m=1}^{i_s}\sum_{n=0}^{i_s-m} \boldsymbol{b}_j^\vee(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_s-m+1}(\mathrm{id}^{\otimes i_s-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n})\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}).$$

On the first line the summations are on the same sets of parameters, so we combine them together using that

$$\mathrm{CY}_1 + \widehat{\mathfrak{m}}_1 = \boldsymbol{b}_1^\vee + \boldsymbol{\Delta}_1 + \boldsymbol{b}_1^x = \boldsymbol{b}_1 + \boldsymbol{\Delta}.$$

Then we rewrite CY_{t_1} and $\widehat{\mathfrak{m}}_m$ on the second line, as well as $\widehat{\mathfrak{m}}_k$ on the first line, using (29) and (30). Finally, we use (32) in the third line, which is assumed to hold by induction hypothesis. After all these

changes we get

$$\sum_{j=2}^{k} \sum_{\substack{t_2,\dots,t_j\geq 1\\ t_2+\cdots+t_j=k-1}} \boldsymbol{b}_j^{\vee}(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes(\boldsymbol{b}_1+\boldsymbol{\Delta}_1))$$

$$+\sum_{r=2}^{k} \sum_{\substack{t_2,\dots,t_r\geq 1\\ t_2+\cdots+t_r=k-1}} \boldsymbol{b}_1^{\vee}\circ(\boldsymbol{b}_r^{x}+\boldsymbol{\Delta}_r)(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{id})$$

$$+\sum_{j=1}^{k} \sum_{\substack{t_1,\dots,t_j\geq 1\\ t_1\geq 2\\ t_1+\cdots+t_j=k}} \sum_{r=2}^{t_1} \sum_{\substack{i_2,\dots,i_r\geq 1\\ i_2+\cdots+i_r=t_1-1}} \boldsymbol{b}_j^{\vee}(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\boldsymbol{b}_r^{\vee}(\mathrm{CY}_{i_r}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}))$$

$$+\sum_{m=2}^{k-1}\sum_{j=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}} \sum_{r=2}^{m} \sum_{\substack{t_2,\dots,t_r\geq 1\\ t_2+\cdots+t_r=m-1}} \boldsymbol{b}_j^{\vee}(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes(\boldsymbol{b}_r^{x}+\boldsymbol{\Delta}_r)(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{id}))$$

$$+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{s=2}^{j}\sum_{r=1}^{i_s} \sum_{\substack{t_1,\dots,t_r\geq 1\\ t_1+\cdots+t_r=i_s}} \boldsymbol{b}_j^{\vee}(\mathrm{CY}_{i_j}\otimes\cdots\otimes\boldsymbol{b}_r^{\vee}(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_1})\otimes\mathrm{CY}_{i_{s-1}}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}).$$

We can sum the three middle lines together, and change variables on the resulting sum as well as on the last line to get

$$\sum_{j=2}^{k} \sum_{\substack{t_2,\dots,t_j\geq 1\\ t_2+\cdots+t_j=k-1}} \boldsymbol{b}_j^{\vee}(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes(\boldsymbol{b}_1+\boldsymbol{\Delta}_1))$$

$$+\sum_{j=2}^{k} \sum_{\substack{t_2,\dots,t_j\geq 1\\ t_2+\cdots+t_j=k-1}} \sum_{r=2}^{j} \boldsymbol{b}_{j-r+1}^{\vee}(\mathrm{CY}_{t_j}\otimes\cdots\otimes\mathrm{CY}_{t_{r+1}}\otimes(\boldsymbol{b}_r+\boldsymbol{\Delta}_r)(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{id}))$$

$$+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{r=1}^{j-1}\sum_{n=1}^{j-r} \boldsymbol{b}_{j-r+1}^{\vee}(\mathrm{CY}_{i_j}\otimes\cdots\otimes\boldsymbol{b}_r^{\vee}(\mathrm{CY}_{i_{n+r}}\otimes\cdots\otimes\mathrm{CY}_{i_{n+1}})\otimes\mathrm{CY}_{i_n}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}).$$

Finally, note that adding $\boldsymbol{b}_r^{x}+\boldsymbol{\Delta}_r$ to $\boldsymbol{b}_r^{\vee}$ on the last line doesn't change anything because these terms vanish for degree and energy reasons (observe that curves contributing to $\boldsymbol{b}_r^{x}(\mathrm{CY}_{i_{n+r}}\otimes\cdots\otimes\mathrm{CY}_{i_{n+1}})$ would have a unique positive asymptotic at a maximum Morse chord and negative asymptotic which have to be Morse chords for action reasons. These negative asymptotics are in the image of the CY map so can only be minimum Morse chords. For index reasons, such rigid discs do not exist). Then we observe that the first line is the case $r=1, n=0$ of the third line while the second line is the case $r\geq 1, n=0$

of the third line, so we have

$$\mathrm{LHS}(32)=\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}}\sum_{r=1}^{j}\sum_{n=0}^{j-r}\boldsymbol{b}^{\vee}_{j-r+1}(\mathrm{id}^{\otimes j-r-n}\otimes(\boldsymbol{b}_r+\boldsymbol{\Delta}_r)\otimes\mathrm{id}^{\otimes n})(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}),$$

which vanishes by Lemma 7.10(a) and thus we have proven (32). In order to prove (31), we will use (32), which simplifies slightly the notation in the computations below. We have

$$\mathrm{LHS}(31)=\widehat{\mathfrak{m}}_1\circ\widehat{\mathfrak{m}}_k+\widehat{\mathfrak{m}}_k(\mathrm{id}^{\otimes k-1}\otimes\widehat{\mathfrak{m}}_1)+\sum_{m=2}^{k-1}\widehat{\mathfrak{m}}_{k-m+1}(\mathrm{id}^{\otimes k-m}\otimes\widehat{\mathfrak{m}}_m)+\sum_{m=1}^{k-1}\sum_{n=1}^{k-m}\widehat{\mathfrak{m}}_{k-m+1}(\mathrm{id}^{\otimes k-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n})$$

and we use (29) to rewrite $\widehat{\mathfrak{m}}_j$ where it appear, except the "inner one" in the last line. We get

$$\begin{aligned}\mathrm{LHS}(31)=&\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}}(\boldsymbol{b}^x_1+\boldsymbol{\Delta}_1)(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})\\&+\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes(\boldsymbol{b}^x_1+\boldsymbol{\Delta}_1))\\&+\sum_{m=2}^{k-1}\sum_{j=2}^{k-m+1}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}}\sum_{r=2}^{m}\sum_{\substack{t_2,\dots,t_r\geq 1\\ t_2+\cdots+t_r=m-1}}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)\big(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\\&\qquad\otimes(\boldsymbol{b}^x_r+\boldsymbol{\Delta}_r)(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{id})\big)\\&+\sum_{m=1}^{k-1}\sum_{n=1}^{k-m}\sum_{j=2}^{k-m+1}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})(\mathrm{id}^{\otimes k-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n}).\end{aligned}$$

Observe that the last line can be written

$$\begin{aligned}&\sum_{m=1}^{k-1}\sum_{j=2}^{k-m+1}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}}\sum_{s=2}^{j}\sum_{n=0}^{i_s}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_s}(\mathrm{id}^{\otimes i_s-1-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n})\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})\\&\quad=\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}}\sum_{s=2}^{j}\sum_{m=1}^{i_s}\sum_{n=0}^{i_s-m}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)\big(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_s-m+1}(\mathrm{id}^{\otimes i_s-m-n}\otimes\widehat{\mathfrak{m}}_m\otimes\mathrm{id}^{\otimes n})\\&\qquad\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}\big),\end{aligned}$$

where the right-hand side of the equality is obtained after a change of variables. On this last line we can now apply (32) to get

$$\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}}\sum_{s=2}^{j}\sum_{r=1}^{i_s}\sum_{\substack{t_1,\dots,t_r\geq 1\\ t_1+\cdots+t_r=i_s}}(\boldsymbol{b}^x_j+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_{s+1}}\otimes\check{\mathfrak{m}}_r(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_1})\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}).$$

Using that $\check{\mathfrak{m}}_r = \boldsymbol{b}_r^\vee$, we have

$$\begin{aligned}
\mathrm{LHS}(31) = &\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} (\boldsymbol{b}_1^x+\boldsymbol{\Delta}_1)(\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})\\
&+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} (\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes(\boldsymbol{b}_1^x+\boldsymbol{\Delta}_1))\\
&+\sum_{m=2}^{k-1}\sum_{j=2}^{k-m+1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-m}} \sum_{r=2}^{m} \sum_{\substack{t_2,\dots,t_r\geq 1\\ t_2+\cdots+t_r=m-1}} (\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)\big(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\\
&\qquad\otimes(\boldsymbol{b}_r^x+\boldsymbol{\Delta}_r)(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_2}\otimes\mathrm{id})\big)\\
&+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{s=2}^{j}\sum_{r=1}^{i_s} \sum_{\substack{t_1,\dots,t_r\geq 1\\ t_1+\cdots+t_r=i_s}} (\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)\big(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_{s+1}}\\
&\qquad\otimes\boldsymbol{b}_r^\vee(\mathrm{CY}_{t_r}\otimes\cdots\otimes\mathrm{CY}_{t_1})\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}\big).
\end{aligned}$$

By a change of variables in the third and fourth lines we obtain

$$\begin{aligned}
\mathrm{LHS}(31) = &\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} (\boldsymbol{b}_1^x+\boldsymbol{\Delta}_1)(\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id})\\
&+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} (\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j)(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes(\boldsymbol{b}_1^x+\boldsymbol{\Delta}_1))\\
&+\sum_{j=2}^{k-1} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{r=2}^{j-1} (\boldsymbol{b}_{j-r+1}^x+\boldsymbol{\Delta}_{j-r+1})(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_{r+1}}\otimes(\boldsymbol{b}_r^x+\boldsymbol{\Delta}_r)(\mathrm{CY}_{i_r}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}))\\
&+\sum_{j=2}^{k} \sum_{\substack{i_2,\dots,i_j\geq 1\\ i_2+\cdots+i_j=k-1}} \sum_{r=1}^{j-1}\sum_{n=1}^{j-r} (\boldsymbol{b}_{j-r+1}^x+\boldsymbol{\Delta}_{j-r+1})\big(\mathrm{CY}_{i_j}\otimes\cdots\\
&\qquad\otimes\boldsymbol{b}_r^\vee(\mathrm{CY}_{i_{n+r}}\otimes\cdots\otimes\mathrm{CY}_{i_{n+1}})\otimes\mathrm{CY}_{i_n}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}\big).
\end{aligned}$$

Finally, observe that:

- Adding $\boldsymbol{b}_j^\vee$ to $\boldsymbol{b}_j^x+\boldsymbol{\Delta}_j$ on the first line does nothing as it vanishes for energy reasons (when the output is a long chord) or by canceling pairs of discs (when the output of $\boldsymbol{b}_j^\vee$ is y and the dimension of the Legendrian is 1).

- Adding $\boldsymbol{b}_1^\vee$ to $\boldsymbol{b}_1^x+\boldsymbol{\Delta}_1$ on the second line contributes nothing more when the output of $\boldsymbol{b}_1^\vee$ is a long chord (for energy reasons). Let us check that it also vanishes when the output of $\boldsymbol{b}_1^\vee$ is the Morse chord y.

In such a case, energy arguments imply that the terms

$$\boldsymbol{\Delta}_j(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\boldsymbol{b}_1^y)$$

vanish because all inputs for $\boldsymbol{\Delta}_j$ would be negative Reeb chords asymptotics; see the definition of $\boldsymbol{\Delta}_j$ at the beginning of Section 7.4. Then we consider the terms

$$\boldsymbol{b}_j^x(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\boldsymbol{b}_1^y).$$

Pseudoholomorphic buildings contributing to such terms should contain a rigid disc with a unique positive asymptotic to x. For energy reasons, the j negative asymptotics must be Morse chords and in particular can only be y's (by definition of the CY maps). For index reasons, such a rigid disc doesn't exist.

- Adding $\boldsymbol{b}_r^\vee$ to $\boldsymbol{b}_r^x+\boldsymbol{\Delta}_r$ on the third line doesn't contribute either for the same reasons as the previous point.

- Adding $\boldsymbol{b}_r^x+\boldsymbol{\Delta}_r$ to $\boldsymbol{b}_r^\vee$ on the fourth line does not change anything either: $\boldsymbol{\Delta}_r(\mathrm{CY}_{i_{n+r}}\otimes\cdots\otimes\mathrm{CY}_{i_{n+1}})$ vanishes for energy reasons and then as before the term $\boldsymbol{b}_r^x(\mathrm{CY}_{i_{n+r}}\otimes\cdots\otimes\mathrm{CY}_{i_{n+1}})$ vanishes also because there is an even number of discs, or no disc at all, with positive asymptotic at x and negative asymptotics only at y's chords.

Moreover, the first line is the case $r=j, n=0$ of the last line, the second line is the case $r=1, n=0$ of the last line, and the third line is the case $2\le r\le j-1, n=0$ of the last line. So we have

$$\mathrm{LHS}(31)=\sum_{j=2}^{k}\sum_{\substack{i_2,\dots,i_j\ge 1\\ i_2+\cdots+i_j=k-1}}\sum_{r=1}^{j}\sum_{n=0}^{j-r}(\boldsymbol{b}_{j-r+1}^x+\boldsymbol{\Delta}_{j-r+1})\\ (\mathrm{id}^{\otimes j-r-n}\otimes(\boldsymbol{b}_r+\boldsymbol{\Delta}_r)\otimes\mathrm{id}^{\otimes n})(\mathrm{CY}_{i_j}\otimes\cdots\otimes\mathrm{CY}_{i_2}\otimes\mathrm{id}),$$

which vanishes by Lemma 7.10(a)–(b). □

8 Example: the unknot

The computation done here is a subcase of the computation done in [2, Section 5]. We nevertheless detail it using our notation. Let Λ be the standard $\mathrm{TB}=-1$ unknot and consider a 2-copy and a 3-copy; see Figure 21. We have

$$\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)=\langle a_{10},x_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}^{\mathrm{cyc}}\quad\text{and}\quad\check{C}_-^{\mathrm{cyc}}(\Lambda_0,\Lambda_1)=\langle a_{01},y_{01}\rangle_{\mathcal{A}\text{-}\mathcal{A}}^{\mathrm{cyc}}$$

with $|a_{10}|_{\widehat{C}_+}=0$, $|x_{01}|_{\widehat{C}_+}=2$, $|a_{01}|_{\check{C}_-}=2$ and $|y_{01}|_{\check{C}_-}=0$. Denote by $\boldsymbol{a}^j=a\dots a$ the word consisting of j times the chord a which is the only Reeb chord of Λ. We have, for all $j\ge 0$, $\widehat{\mathfrak{m}}_1(a_{10}\boldsymbol{a}^j)=x_{01}a\boldsymbol{a}^j+x_{01}\boldsymbol{a}^ja=0$, and $\widehat{\mathfrak{m}}_1(x_{01})=0$ as well. On the other side $\check{\mathfrak{m}}_1(a_{01}\boldsymbol{a}^j)=0$, for example, for degree reasons, and $\check{\mathfrak{m}}_1(y_{01}\boldsymbol{a}^j)=a_{01}a\boldsymbol{a}^j+a_{01}\boldsymbol{a}^ja=0$. So both differentials $\widehat{\mathfrak{m}}_1$ and $\check{\mathfrak{m}}_1$ vanish implying that the homologies of $(\widehat{C}_+^{\mathrm{cyc}}(\Lambda_0,\Lambda_1),\widehat{\mathfrak{m}}_1)$ and $(\check{C}_-^{\mathrm{cyc}}(\Lambda_0,\Lambda_1),\check{\mathfrak{m}}_1)$ are infinite-dimensional generated

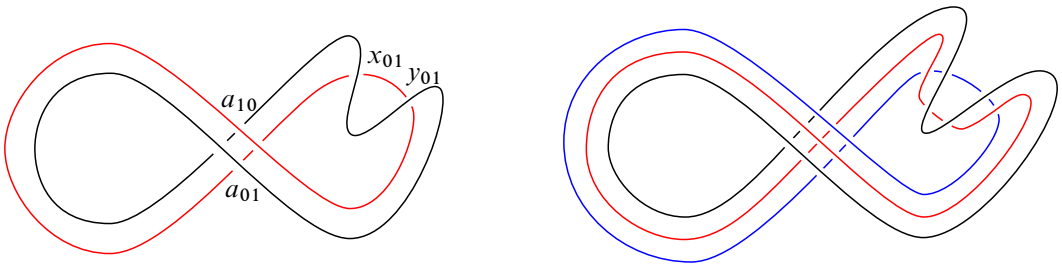


Figure 21: Lagrangian projections of the 2-copy $\Lambda_0 \cup \Lambda_1$ on the left, and of the 3-copy $\Lambda_0 \cup \Lambda_1 \cup \Lambda_2$ on the right.

by all words $a_{10}\boldsymbol{a}^j, x_{01}\boldsymbol{a}^j$ and $a_{01}\boldsymbol{a}^j, y_{01}\boldsymbol{a}^j$, respectively. In this simple case there is a unique way to define the Calabi–Yau map by degree reasons, but one can also easily see on the figure which are the bananas with a positive asymptotic at a_{10} and the strips with a negative asymptotic at x_{01}. This gives

$$\mathrm{CY}_1(a_{10}\boldsymbol{a}^j) = y_{01}\boldsymbol{a}^j \quad \text{and} \quad \mathrm{CY}_1(x_{01}\boldsymbol{a}^j) = a_{01}\boldsymbol{a}^j.$$

We can then use the Lagrangian projection of the 3-copy to compute the product structures on $\widehat{C}_+^{\mathrm{cyc}}$ and $\check{C}_-^{\mathrm{cyc}}$. The rigid discs asymptotic to generators of $\check{C}_-^{\mathrm{cyc}}$ with one positive asymptotic and two negative asymptotics contribute to the product $\check{\mathfrak{m}}_2$ and one can see that the only ones are those giving

$$\check{\mathfrak{m}}_2(a_{12}\boldsymbol{a}^j, y_{01}\boldsymbol{a}^i) = a_{02}\boldsymbol{a}^{i+j}, \quad \check{\mathfrak{m}}_2(y_{12}\boldsymbol{a}^j, a_{01}\boldsymbol{a}^i) = a_{02}\boldsymbol{a}^{i+j}, \quad \check{\mathfrak{m}}_2(y_{12}\boldsymbol{a}^j, y_{01}\boldsymbol{a}^i) = y_{02}\boldsymbol{a}^{i+j},$$

which expresses the fact that the minimum Morse Reeb chord acts as a unit, ie induces a quasi-isomorphism $\check{C}_-^{\mathrm{cyc}}(\Lambda_1, \Lambda_2) \cong \check{C}_-^{\mathrm{cyc}}(\Lambda_0, \Lambda_2)$. For the product $\widehat{\mathfrak{m}}_2$ we have to find buildings as the one pictured in Figure 6. The only ones are those we depicted on Figure 22. They give the following nontrivial components of the product:

$$\widehat{\mathfrak{m}}_2(a_{21}\boldsymbol{a}^j, a_{10}\boldsymbol{a}^i) = a_{20}\boldsymbol{a}^{i+j}, \quad \widehat{\mathfrak{m}}_2(a_{21}\boldsymbol{a}^j, x_{01}\boldsymbol{a}^i) = x_{02}\boldsymbol{a}^{i+j}, \quad \widehat{\mathfrak{m}}_2(x_{12}\boldsymbol{a}^j, a_{10}\boldsymbol{a}^i) = x_{02}\boldsymbol{a}^{i+j},$$

which translates the fact that the mixed a generator acts as a unit for the product $\widehat{\mathfrak{m}}_2$. The higher-order operations $\check{\mathfrak{m}}_j$ and $\widehat{\mathfrak{m}}_j$ for $j \geq 3$ vanish.

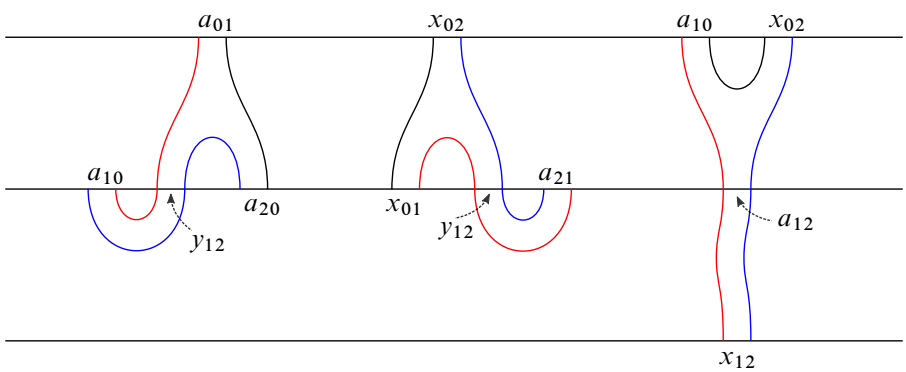


Figure 22: Pseudoholomorphic buildings contributing to the product $\widehat{\mathfrak{m}}_2$ for the unknot.

References

[1] **F Bourgeois**, **B Chantraine**, *Bilinearized Legendrian contact homology and the augmentation category*, J. Symplectic Geom. 12 (2014) 553–583 MR

[2] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Symplectic homology product via Legendrian surgery*, Proc. Natl. Acad. Sci. USA 108 (2011) 8114–8121 MR

[3] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR

[4] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Floer homology and Lagrangian concordance*, from "Proceedings of the Gökova Geometry–Topology Conference 2014", GGT, Gökova (2015) 76–113 MR

[5] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Floer theory for Lagrangian cobordisms*, preprint (2015) arXiv 1511.09471v1

[6] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Noncommutative augmentation categories*, from "Proceedings of the Gökova Geometry-Topology Conference 2015", GGT, Gökova (2016) 116–150 MR

[7] **B Chantraine**, **G Dimitroglou Rizell**, **P Ghiggini**, **R Golovko**, *Geometric generation of the wrapped Fukaya category of Weinstein manifolds and sectors*, Ann. Sci. École Norm. Sup. 57 (2024) 1–85 MR

[8] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR

[9] **K Cieliebak**, **U A Frauenfelder**, *A Floer homology for exact contact embeddings*, Pacific J. Math. 239 (2009) 251–316 MR

[10] **K Cieliebak**, **A Oancea**, *Symplectic homology and the Eilenberg–Steenrod axioms*, Algebr. Geom. Topol. 18 (2018) 1953–2130 MR

[11] **G Civan**, **P Koprowski**, **J Etnyre**, **J M Sabloff**, **A Walker**, *Product structures for Legendrian contact homology*, Math. Proc. Cambridge Philos. Soc. 150 (2011) 291–311 MR

[12] **G Dimitroglou Rizell**, *Legendrian ambient surgery and Legendrian contact homology*, J. Symplectic Geom. 14 (2016) 811–901 MR

[13] **G Dimitroglou Rizell**, *Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications*, Quantum Topol. 7 (2016) 29–105 MR

[14] **G Dimitroglou Rizell**, **T Ekholm**, **D Tonkonog**, *Refined disk potentials for immersed Lagrangian surfaces*, J. Differential Geom. 121 (2022) 459–539 MR

[15] **T Ekholm**, *Morse flow trees and Legendrian contact homology in 1-jet spaces*, Geom. Topol. 11 (2007) 1083–1224 MR

[16] **T Ekholm**, *Rational SFT, linearized Legendrian contact homology, and Lagrangian Floer cohomology*, from "Perspectives in analysis, geometry, and topology", Progr. Math. 296, Birkhäuser, New York (2012) 109–145 MR

[17] **T Ekholm**, *Holomorphic curves for Legendrian surgery*, preprint (2019) arXiv 1906.07228

[18] **T Ekholm**, **J B Etnyre**, **J M Sabloff**, *A duality exact sequence for Legendrian contact homology*, Duke Math. J. 150 (2009) 1–75 MR

[19] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 177–305 MR

[20] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Legendrian contact homology in $P \times \mathbb{R}$*, Trans. Amer. Math. Soc. 359 (2007) 3301–3335 MR

[21] **T Ekholm**, **Y Lekili**, *Duality between Lagrangian and Legendrian invariants*, Geom. Topol. 27 (2023) 2049–2179 MR

[22] **Y Eliashberg**, *Invariants in contact topology*, from "Proceedings of the International Congress of Mathematicians, II" (G Fischer, U Rehmann, editors), Deutsche Mathematiker Vereinigung, Berlin (1998) 327–338 MR

[23] **Y Félix**, **S Halperin**, **J-C Thomas**, *Differential graded algebras in topology*, from "Handbook of algebraic topology", North-Holland, Amsterdam (1995) 829–865 MR

[24] **S Ganatra**, *Symplectic cohomology and duality for the wrapped Fukaya category*, PhD thesis, Massachusetts Institute of Technology (2012)

[25] **S Ganatra**, **J Pardon**, **V Shende**, *Sectorial descent for wrapped Fukaya categories*, J. Amer. Math. Soc. 37 (2024) 499–635 MR

[26] **V Ginzburg**, *Calabi–Yau algebras*, preprint (2006) arXiv math/0612139

[27] **B Keller**, *Deformed Calabi–Yau completions*, J. Reine Angew. Math. 654 (2011) 125–180 MR

[28] **M Kontsevich**, **Y Soibelman**, *Notes on A_∞-algebras, A_∞-categories and non-commutative geometry*, from "Homological mirror symmetry", Lecture Notes in Phys. 757, Springer (2009) 153–219 MR

[29] **N Legout**, *A_∞-category of Lagrangian cobordisms in the symplectization of $P \times \mathbb{R}$*, Quantum Topol. 14 (2023) 101–200 MR

[30] **D McDuff**, **D Salamon**, *J-holomorphic curves and symplectic topology*, Amer. Math. Soc. Colloq. Publ. 52, Amer. Math. Soc., Providence, RI (2004) MR

[31] **W J Merry**, *Lagrangian Rabinowitz Floer homology and twisted cotangent bundles*, Geom. Dedicata 171 (2014) 345–386 MR

[32] **L Ng**, **D Rutherford**, **V Shende**, **S Sivek**, **E Zaslow**, *Augmentations are sheaves*, Geom. Topol. 24 (2020) 2149–2286 MR

[33] **J M Sabloff**, *Duality for Legendrian contact homology*, Geom. Topol. 10 (2006) 2351–2381 MR

Department of Mathematical Sciences, Chalmers University of Technology
Göteborg, Sweden

noelet@hig.se

Received: 24 August 2023 Revised: 28 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3679–3706
DOI: 10.2140/agt.2025.25.3679
Published: 1 October 2025

On the resolution of kinks of curves on punctured surfaces

CHRISTOF GEISS
DANIEL LABARDINI-FRAGOSO

Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a surface with marked points $\mathbb{M} \subseteq \partial\Sigma \neq \varnothing$ and punctures $\mathbb{P} \subseteq \Sigma \setminus \partial\Sigma$. We show that for every curve γ on $\Sigma \setminus \mathbb{P}$, the curve obtained by resolving the kinks of γ in any order is uniquely determined, up to homotopy in $\Sigma \setminus \mathbb{P}$, by the 2-orbifold homotopy class of γ, in which the punctures are interpreted to be orbifold points of order 2. Our proof relies on an application of the diamond lemma.

57K20; 13F60, 18B40

1. Introduction 3679
2. Background 3682
3. The (leafy) dual graph of a triangulation of signature zero 3685
4. Kinks of walks on graphs 3689
5. Kinks of curves via kinks of walks on graphs 3699
6. Equivalence of orbifold fundamental groupoids 3702
7. Main result: uniqueness of kink-free representative curves 3704
References 3705

1 Introduction

Surfaces with marked points are classical mathematical objects that, after the appearance of works by Assem, Brüstle, Charbonneau-Jodoin, and Plamondon [6], Fock and Goncharov [8], Fomin, Shapiro, and Thurston [9], and Labardini-Fragoso [16], have suffused both cluster algebras and the representation theory of algebras during the last 15 years, with remarkable connections between geometry and representation theory discovered in works by Haiden, Katzarkov, and Kontsevich [11], Lekili and Polishchuk [18], and Opper, Plamondon, and Schroll [21]. Recent developments by Amiot and Brüstle [2], Amiot and Plamondon [5], and Labardini-Fragoso, Schroll, and Valdivieso [17] have discovered that, somewhat mysteriously, it is sometimes necessary to interpret the punctures not as holes, but as orbifold points of order 2.

When a puncture is regarded as an orbifold point of order 2 there are certain loops, originally of infinite order, that are declared to have order 2, namely each loop closely wrapping around such puncture. The

© 2025 The Authors, under license to MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

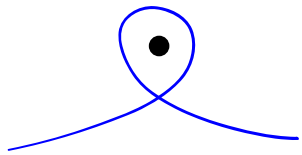


Figure 1

2-*orbifold fundamental groupoid* of a surface $(\Sigma, \mathbb{M}, \mathbb{P})$ with marked points on the boundary $\mathbb{M} \subseteq \partial\Sigma$ and punctures $\mathbb{P} \subseteq \Sigma \setminus \partial\Sigma$ is thus defined as the quotient groupoid of the topological fundamental groupoid of $(\Sigma, \mathbb{M}, \mathbb{P})$ obtained by treating all punctures as orbifold points of order 2.

Intuitively, a *kink* of a curve γ on $\Sigma \setminus \mathbb{P}$ is a segment κ of γ that turns out to be one of the aforementioned loops, and that cannot be dissolved with any homotopy rel $\{0, 1\}$ of curves on $\Sigma \setminus \mathbb{P}$; see Figure 1. However, the notion of a *kink* of an arbitrary morphism in the topological fundamental groupoid of $(\Sigma, \mathbb{M}, \mathbb{P})$, that is, of an arbitrary homotopy class rel $\{0, 1\}$ of curves connecting two given points on $\Sigma \setminus \mathbb{P}$, is not as easy to define in terms of arbitrary representatives of the morphism as one would like, because an arbitrary representative can be a very complicated curve even if the morphism possesses some nice representative — consider, for instance, a curve that has not only self-crossings and self-tangencies, but also different segments that it traverses multiple times, perhaps with immediate backtrackings of some of them.

Here we define the notion of a kink of a curve γ in terms of the associated walk on the (leafy) dual graph of an arbitrary triangulation of signature zero. We show that γ having kinks is independent of the triangulation of signature zero taken, and that the curve obtained by resolving the kinks of γ in any order is uniquely determined, up to homotopy in $\Sigma \setminus \mathbb{P}$, by the 2-orbifold homotopy class of γ, in which the punctures are interpreted to be orbifold points of order 2. Our proof uses a nontrivial application of the diamond lemma.

Let us describe the contents of the paper in some detail. A *triangulation of signature zero* is an ideal triangulation with the property that every puncture is enclosed by a self-folded triangle. Such a triangulation τ has its associated *dual graph* $G(\tau)$, whose vertices are the triangles of τ and the boundary segments of $(\Sigma, \mathbb{M}, \mathbb{P})$, with an edge connecting two triangles each time they share an arc of τ, and with an edge between a boundary segment and the unique triangle containing it. To simplify the treatment of morphisms in the fundamental groupoid of the graph, we introduce the *leafy dual graph* $G^\circlearrowright(\tau)$, which is a slight enlargement of $G(\tau)$. An important feature of $G^\circlearrowright(\tau)$ is that every nonidentity morphism f in the fundamental groupoid $\pi_1(G^\circlearrowright(\tau))$ can be uniquely represented as a backtrack-free walk on $G^\circlearrowright(\tau)$; we call such a walk the *standard form* of f.

The orientation of Σ provides $G(\tau)$ with a natural structure of a *ribbon graph*, which we extend to a ribbon graph structure for $G^\circlearrowright(\tau)$. We use this ribbon structure of $G^\circlearrowright(\tau)$ to define the notion of a *kink* of any backtrack-free walk on $G^\circlearrowright(\tau)$, ie of any nonidentity morphism in $\pi_1(G^\circlearrowright(\tau))$ written in standard form. We define the *resolution* of a kink of a backtrack-free walk w on $G^\circlearrowright(\tau)$ as the result of applying a purely combinatorial operation that mimics the topological replacement sketched in Figure 2 and produces another backtrack-free walk w' representing the same morphism as w in the 2-orbifold fundamental

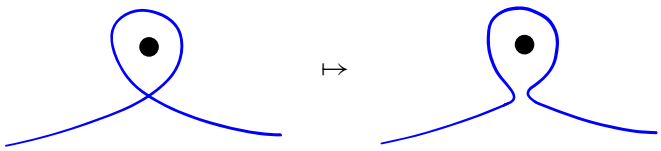


Figure 2

groupoid $\pi_1^{\mathrm{orb}}(G^{\circlearrowleft}(\tau))$, but whose class in the fundamental groupoid $\pi_1(G^{\circlearrowleft}(\tau))$ may differ from that of w. We prove that the number of kinks decreases every time the resolution of a kink is applied, and that if two distinct kinks of a backtrack-free walk are resolved, producing backtrack-free walks w' and w'', then it is always possible to resolve some sequence of kinks of w' and some sequence of kinks of w'' so that the resulting backtrack-free walks coincide. This enables us to apply the *diamond lemma* and deduce that, no matter the order in which the kinks of w are resolved, one always arrives at the same kink-free backtrack-free walk without kinks after finitely many steps.

To establish the corresponding result for curves on $\Sigma \setminus \mathbb{P}$ we proceed as follows. The *ribbon surface* $\Sigma(G^{\circlearrowleft}(\tau))$ can be naturally embedded in Σ. As usual, there are strong deformation retractions $\Sigma \setminus \mathbb{P} \to \Sigma(G^{\circlearrowleft}(\tau)) \to G^{\circlearrowleft}(\tau)$, whose composition we denote by ρ. We use these retractions to define the notion of a *kink of a curve γ with respect to τ* as a kink of the standard form of $\rho(\gamma)$. We show that for any two triangulations τ and σ of signature zero, γ has a kink with respect to τ if and only if γ has a kink with respect to σ. Furthermore, the constructions and definition make it transparent that ρ induces a commutative diagram of groupoids

$$\begin{array}{ccc} \pi_1(\Sigma \setminus \mathbb{P}, E) & \xrightarrow[\cong]{\rho_\#} & \pi_1(G^{\circlearrowleft}(\tau), E) \\ \downarrow{\scriptstyle \mathfrak{p}} & & \downarrow{\scriptstyle \mathfrak{p}} \\ \pi_1^{\mathrm{orb}}(\Sigma \setminus \mathbb{P}, E) & \xrightarrow[\bar{\rho}_\#]{\cong} & \pi_1^{\mathrm{orb}}(G^{\circlearrowleft}(\tau), E) \end{array}$$

whose horizontal arrows are isomorphisms, where the vertical arrows are the canonical projections of quotient groupoids. From this, the fact that the resolution of kinks of a backtrack-free walk w on $G^{\circlearrowleft}(\tau)$ does not affect the 2-orbifold homotopy class of w, and the uniqueness result from the last line of the previous paragraph, we deduce the main result of this paper:

Theorem 1.1 *Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a surface with nonempty boundary, and let $E \subseteq \partial\Sigma$ be a set containing exactly one point from the relative interior of each boundary segment of $(\Sigma, \mathbb{M}, \mathbb{P})$.*

- *Given $u_0, v_0 \in E$, there is exactly one function*

$$\iota\colon \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u_0, v_0) \to \pi_1(\Sigma \setminus \mathbb{P}, E)(u_0, v_0)$$

such that

(1) $\mathfrak{p} \circ \iota = \mathbb{1}$,

(2) *for every $f \in \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u_0, v_0)$ there exists a representative curve $\gamma \in \iota(f)$ that has no kinks.*

- *There is exactly one function*

$$\iota\colon \pi_{1,\mathbb{P}}^{\mathrm{orb,free}}(\Sigma, E)/\!\sim \to \pi_1^{\mathrm{free}}(\Sigma \setminus \mathbb{P}, E)/\!\sim$$

such that

(1) $\mathfrak{p} \circ \iota = \mathbb{1}$,

(2) *for every* $f \in \pi_{1,\mathbb{P}}^{\mathrm{orb,free}}(\Sigma, E)/\!\sim$ *there exists a representative curve* $\gamma \in \iota(f)$ *that has no kinks, where* $\sim$ *is the equivalence relation that identifies each closed curve with its opposite orientation.*

The paper is organized as follows. In Section 2 we present the formal definition of a *quotient subgroupoid* of a groupoid by what we call a *normal multilocular subgroup*. We also recall the definition of the fundamental groupoid of a graph G, and that, provided G is loop-free, every nonidentity morphism f in the fundamental groupoid of G can be uniquely represented as a walk on G without backtrackings; we call this walk the *standard form* of f. In Section 3 we recall the definition of the dual graph $G(\tau)$ of a triangulation τ of signature zero, and introduce the *leafy dual graph* $G^\circledcirc(\tau)$. In Section 4 we define the combinatorial notions of a *kink* of a morphism belonging to the fundamental groupoid $\pi_1(G^\circledcirc(\tau))$, and of the *resolution* of a kink, and show that the resolution of kinks satisfies the confluence conditions that allow us to apply the diamond lemma. In Section 5 we introduce the notion of a *kink of a curve* with respect to a triangulation of signature zero, and show that the absence of kinks in a curve is independent of the triangulation with respect to which it is considered. In Section 6 we see that the (2-orbifold) fundamental groupoid of $(\Sigma, \mathbb{M}, \mathbb{P})$ is isomorphic to the (2-orbifold) fundamental groupoid of $G^\circledcirc(\tau)$. In Section 7 we present our main result.

2 Background

2.1 Groupoids, normal multilocular subgroups, and quotient groupoids

Recall that a *groupoid* is a category in which every morphism is invertible. Thus, for every object x of a groupoid Γ, the endomorphism set $\Gamma(x,x)$ is a group under composition.

Definition 2.1 A *multilocular subgroup* of a groupoid Γ is a collection $H = (H(x))_{x\in \mathrm{obj}(\Gamma)}$ of subgroups $H(x) \subseteq \Gamma(x,x)$. A *normal multilocular subgroup* is a multilocular subgroup $H = (H(x))_{x\in \mathrm{obj}(\Gamma)}$ with the property that for every two objects x and y of Γ, every $h \in H(x)$, and every morphism $g \in \Gamma(x,y)$, we have $ghg^{-1} \in H(y)$.

Suppose that H is a normal multilocular subgroup of the groupoid Γ. For each pair of objects x and y of Γ, let $\equiv_H \subseteq \Gamma(x,y) \times \Gamma(x,y)$ be the relation defined by the rule

$$f \equiv_H g \iff \text{there exists } h \in H(x) \text{ such that } f = gh.$$

A routine exercise shows that $\equiv_H$ is an equivalence relation. Moreover, if x, y, and z are objects of Γ and $f_1, g_1 \in \Gamma(x,y)$ and $f_2, g_2 \in \Gamma(y,z)$ are morphisms such that $f_1 \equiv_H g_1$ and $f_2 \equiv_H g_2$, then, taking $h_1 \in H(x)$ and $h_2 \in H(y)$ such that $f_1 = g_1h_1$ and $f_2 = g_2h_2$, we have

$$f_2f_1 = g_2h_2g_1h_1 = g_2g_1(g_1^{-1}h_2g_1)h_1,$$

with $(g_1^{-1}h_2g_1)h_1 \in H(x)$, which shows that $f_2f_1 \equiv_H g_2g_1$ too, just as in the case of groups and normal subgroups.

Definition 2.2 Suppose Γ is a groupoid and H is a normal multilocular subgroup of Γ. The *quotient groupoid* Γ/H is the category having the same objects as Γ, with

$$\begin{aligned}(\Gamma/H)(x,y) &:= \Gamma(x,y)/\equiv_H && \text{for } x,y \in \operatorname{obj}(\Gamma),\\ [f_2][f_1] &:= [f_2f_1] && \text{for } x,y,z \in \operatorname{obj}(\Gamma),\ f_1 \in \Gamma(x,y), \text{ and } f_2 \in \Gamma(y,z).\end{aligned}$$

It is obvious that Γ/H is indeed a groupoid, and that the canonical projection $\mathfrak{p}\colon \Gamma \to \Gamma/H$ is a full covariant functor, essentially surjective. Furthermore, the first isomorphism theorem is satisfied:

Theorem 2.3 *If Γ_1 and Γ_2 are groupoids and $F\colon \Gamma_1 \to \Gamma_2$ is a full covariant functor, then, setting $H := \big(\ker\big(\Gamma_1(x,x) \xrightarrow{F} \Gamma_2(F(x),(F(x)))\big)\big)_{x\in\operatorname{obj}\Gamma_1}$, there is a unique covariant functor $\overline{F}\colon \Gamma_1/H \to \Gamma_2$ such that the diagram*

$$\begin{array}{ccc}\Gamma_1 & \xrightarrow{F} & \Gamma_2\\ \downarrow & \nearrow_{\overline{F}} & \\ \Gamma_1/H & & \end{array}$$

commutes. The functor $\overline{F}$ is fully faithful, and if F is bijective on objects (*resp. essentially surjective*)*, then $\overline{F}$ is an isomorphism of categories* (*resp. an equivalence of categories*)*.*

It is straightforward to verify that if Γ is a groupoid and $\mathcal{F} = \{H_\alpha\}_{\alpha\in I}$ is a nonempty class of normal multilocular subgroups of Γ, then $\big(\bigcap_{\alpha\in I} H_\alpha(x)\big)_{x\in\operatorname{obj}(\Gamma)}$ is again a normal multilocular subgroup of Γ. Hence any collection $S = (S(x))_{x\in\operatorname{obj}(\Gamma)}$ of subsets $S(x) \subseteq \Gamma(x,x)$ generates a normal multilocular subgroup H of Γ. Such H has the property that for each $x \in \operatorname{obj}(\Gamma)$, the normal subgroup of $\Gamma(x,x)$ generated by $S(x)$ is contained in $H(x)$, but the containment may be proper if $H(x)$ is not a characteristic subgroup of $\Gamma(x,x)$. In any case, it is readily seen that $H(x)$ coincides with the subgroup of $\Gamma(x,x)$ generated by $\bigcup_{y\in\operatorname{obj}(\Gamma)}\{g^{-1}hg \mid h \in S(y),\ g \in \Gamma(x,y)\}$.

2.2 The fundamental groupoid of a graph

Let G be a finite graph. As is common practice, whenever we want to think of G as a topological space, we shall identify G with a 1-dimensional CW-complex having the vertices of G as 0-cells, and the edges of G as 1-cells.

Define a category $\pi_1(G)$ as follows. Its objects are the vertices of G. Given two objects u and v of $\pi_1(G)$, we set

$$\pi_1(G)(u,v) := P(u,v)/\simeq,$$

where $P(u,v)$ is the set of all continuous curves from u to v in G seen as a topological space, and $\simeq \subseteq P(u,v)\times P(u,v)$ is the equivalence relation of homotopy relative to extreme points. Composition in $\pi_1(G)$ is induced by concatenation of curves. It is well known, and easy to see, that $\pi_1(G)$ is a groupoid, the *fundamental groupoid of* G.

Definition 2.4 Let G be a loop-free graph.

(1) A *backtrack-free walk* on G is a finite sequence

$$f = (u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, u_n),$$

where $u_0, u_1, \dots, u_n$ are vertices of G, and $e_1, \dots, e_n$ are edges of G, such that

- for $l = 1, \dots, n$, the edge e_l connects the vertices u_l and u_{l-1},
- for $l = 1, \dots, n$, the edges e_l and e_{l-1} are not equal.

(2) A *closed backtrack-free walk* is a backtrack-free walk $f = (u_0, e_1, \dots, e_n, u_n)$ such that $u_0 = u_n$ and $e_1 \neq e_n$.

(3) We say that two closed backtrack-free walks

$$f = (u_0, e_1, u_1, \dots, u_{n-1}, e_n, u_n) \quad \text{and} \quad g = (v_0, d_1, v_1, \dots, v_{m-1}, d_m, v_m)$$

are *rotationally equivalent*, and write $f \sim_{\mathrm{rot}} g$, if $n = m$ and for some index $k \in \{0, \dots, n-1\}$ we have

$$g = (u_k, e_{k+1}, u_{k+1}, \dots, u_{n-1}, e_n, u_0, e_1, u_1, \dots, u_{k-1}, e_k, u_k).$$

Remark 2.5 Let G be a loop-free graph. Rotational equivalence is an equivalence relation on the set of closed backtrack-free walks. The equivalence class of such an f will be denoted by $[F]_{\mathrm{rot}}$.

Let G be a loop-free graph. A backtrack-free walk $(u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, u_n)$ determines a morphism from u_0 to u_n in the fundamental groupoid $\pi_1(G)$ by concatenating the curves obtained by parametrizing each edge e_l as a continuous curve from u_{l-1} to u_l. Similarly, each rotational equivalence class C determines an element of the so-called *free-homotopy fundamental group* $\pi_1^{\mathrm{free}}(G)$ by taking, for any representative $f \in C$ (thus $C = [f]_{\mathrm{rot}}$), the free-homotopy class of the closed curve $\mathbb{S}^1 \to G$ defined by concatenating the curves obtained by parametrizing each edge e_l as a continuous curve from u_{l-1} to u_l.

The next result is a consequence of eg [19, Chapter 4].

Theorem 2.6 *Suppose that G is a loop-free graph.*

(1) *For every pair of vertices u_0 and v_0 of G, every morphism $f : u_0 \to v_0$ in the fundamental groupoid $\pi_1(G)$ can be represented uniquely as a backtrack-free walk $(u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, v_0)$.*

(2) *Every free-homotopy class belonging to $\pi_1^{\mathrm{free}}(G)$ can be represented uniquely as a rotational equivalence class of closed backtrack-free walks on G.*

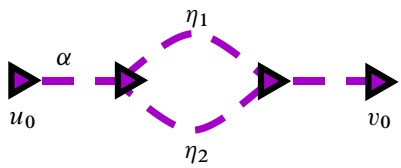


Figure 3

Definition 2.7 Let G be a loop-free graph, and let f be a morphism in $\pi_1(G)$ (resp. a free-homotopy class belonging to $\pi_1^{\text{free}}(G)$). The unique backtrack-free walk (resp. the unique rotational equivalence class of closed backtrack-free walks) representing f in Theorem 2.6 will be called the *standard form* of f and written as $\phi(f)$.

Thus, if $\phi(f) = (u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, v_0)$, then in the fundamental groupoid $\pi_1(G)$ we have the equality

$$f = (u_0, e_1, u_1) * (u_1, e_2, u_2) * \cdots * (u_{n-2}, e_{n-1}, u_{n-1}) * (u_{n-1}, e_n, v_0),$$

where $*$ is the operation induced by concatenation of paths (written from left to right, that is, opposite to the usual way of composing morphisms in a category, ie $\circ = *^{\text{op}}$ for $\pi_1(G)$). In particular, $\mathbb{1}_{u_0} = \phi(\mathbb{1}_{u_0}) = (u_0)$.

Notice that in the situation of Theorem 2.6, if $(u_0, e_1, \dots, e_n, v_0)$ is the standard form of the nonidentity morphism $f : u_0 \to v_0$, then $(v_0, e_n, \dots, e_1, u_0)$ is the standard form of the inverse $f^{-1} : v_0 \to u_0$.

By abuse of notation, we denote any given backtrack-free walk $(u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, u_n)$ or rotation class of backtrack-free walks $[(u_0, e_1, u_1, e_2, \dots, e_{n-1}, u_{n-1}, e_n, u_n)]_{\text{rot}}$ simply by

$$(u_0, e_1, e_2, \dots, e_{n-1}, e_n, u_n).$$

Example 2.8 The graph G depicted in Figure 3 does not have loops, so Theorem 2.6 can be applied to it. The endomorphism group $\pi_1(G)(u_0, u_0)$ is infinite cyclic, generated by either $f = (u_0, \alpha, \eta_1, \eta_2, \alpha, u_0)$ or $f^{-1} = (u_0, \alpha, \eta_2, \eta_1, \alpha, u_0)$. For $n > 0$, the standard forms of f^n and f^{-n} are, respectively,

$$(u_0, \alpha, \underbrace{\eta_1, \eta_2, \eta_1, \eta_2, \dots, \eta_1, \eta_2}_{\text{length}=2n}, \alpha, u_0) \quad \text{and} \quad (u_0, \alpha, \underbrace{\eta_2, \eta_1, \eta_2, \eta_1, \dots, \eta_2, \eta_1}_{\text{length}=2n}, \alpha, u_0).$$

3 The (leafy) dual graph of a triangulation of signature zero

The following notion is in sync with [9, Definition 9.1 and Section 9.2].

Definition 3.1 Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a (possibly punctured) surface with nonempty boundary. An ideal triangulation τ of $(\Sigma, \mathbb{M}, \mathbb{P})$ is said to have *signature zero* if every puncture is enclosed by a self-folded triangle of τ.

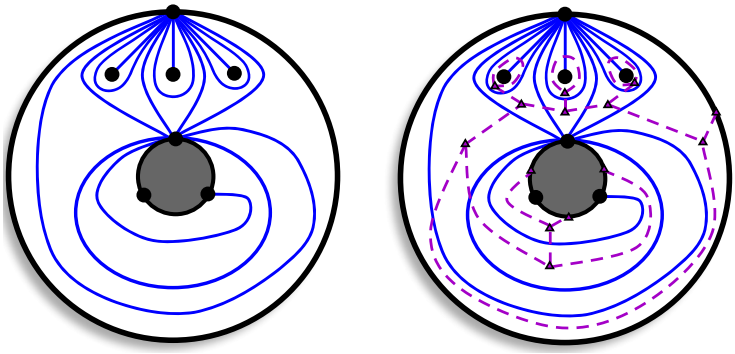


Figure 4

For each ideal triangulation τ of signature zero, the *dual graph* $G(\tau)$ is defined as follows. The vertices of $G(\tau)$ are the triangles of τ and the boundary segments of $(\Sigma, \mathbb{M}, \mathbb{P})$. For each arc k of τ we put an edge connecting the triangles that share k, and for each boundary segment s of $(\Sigma, \mathbb{M}, \mathbb{P})$ we put an edge connecting s to the unique triangle of τ that contains it. Notice that

- the self-folded triangles of τ are precisely the vertices incident to loops of the graph $G(\tau)$,
- every triangle of τ has valency 3 as a vertex of $G(\tau)$, whereas each boundary segment has valency 1.

We turn $G(\tau)$ into a ribbon graph (or fat graph) in a natural way by letting the cyclic order on the edges incident to each vertex v of $G(\tau)$ be the clockwise order around v, according to the orientation of Σ.

Example 3.2 In Figure 4 we can see a triangulation τ of an annulus with one marked point on one boundary component, three on the other, and three punctures. We can also see the dual graph $G(\tau)$ drawn on the surface.

Definition 3.3 Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a (possibly punctured) surface with nonempty boundary, and τ a signature-zero ideal triangulation of $(\Sigma, \mathbb{M}, \mathbb{P})$. The *leafy dual graph* of τ is the graph $G^{\circledcirc}(\tau)$ obtained from $G(\tau)$ after applying the following combinatorial procedure. For each self-folded triangle v of τ,

(1) split its corresponding loop η_v into two distinct edges $\eta_{v,1}$ and $\eta_{v,2}$, each connecting v to a newly introduced vertex w_v,

(2) introduce a leaf ℓ_v incident to w_v; call z_v the vertex of ℓ_v distinct from w_v.

We extend the ribbon graph structure of $G(\tau)$ to a ribbon graph structure of $G^{\circledcirc}(\tau)$ as indicated in Figure 5.

Remark 3.4 We stress the fact that all the vertices of $G(\tau)$ are vertices of $G^{\circledcirc}(\tau)$ as well, and that every boundary segment of $(\Sigma, \mathbb{M}, \mathbb{P})$ is a vertex of τ.

Every edge of the leafy dual graph $G^{\circledcirc}(\tau)$ is incident to a vertex of valency 3. Given a valency-3 vertex u of $G^{\circledcirc}(\tau)$ and an edge e of $G^{\circledcirc}(\tau)$ containing u, we use the ribbon graph structure of $G^{\circledcirc}(\tau)$ to define

$$\begin{aligned} e^{+,u} &:= \text{the edge incident to } u \text{ which is preceded by } e \text{ around } u, \\ e^{-,u} &:= \text{the edge incident to } u \text{ which is followed by } e \text{ around } u. \end{aligned} \tag{3-1}$$

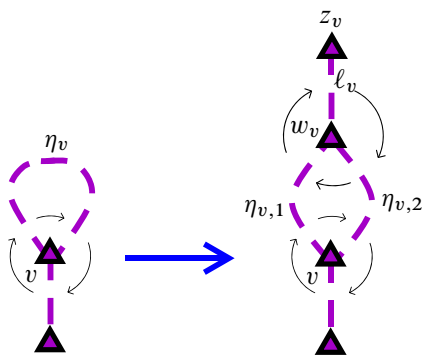


Figure 5

Example 3.5 Consider the triangulation τ from Example 3.2. On the left-hand side of Figure 6 we can see the ribbon graph $G(\tau)$. On the right-hand side of the figure we can see the ribbon graph $G^{\Upsilon}(\tau)$.

Although the next result is quite standard, we briefly sketch its proof, essentially following [1, Section 2.3; 3, Section 2.2; 4, Section 3.2]. Along the way, we emphasize that the standard construction of the ribbon surface $\Sigma(G^{\Upsilon}(\tau))$ associated to $G^{\Upsilon}(\tau)$ allows us to see it as the result of gluing very rigid hexagonal ribbons. This will be very useful later on, to fix a very specific class of curves for which the notion of a "kink" is easy to define, and with the property that every morphism in the fundamental groupoid $\pi_1(\Sigma(G^{\Upsilon}(\tau)))$ is represented by at least one curve in the class.

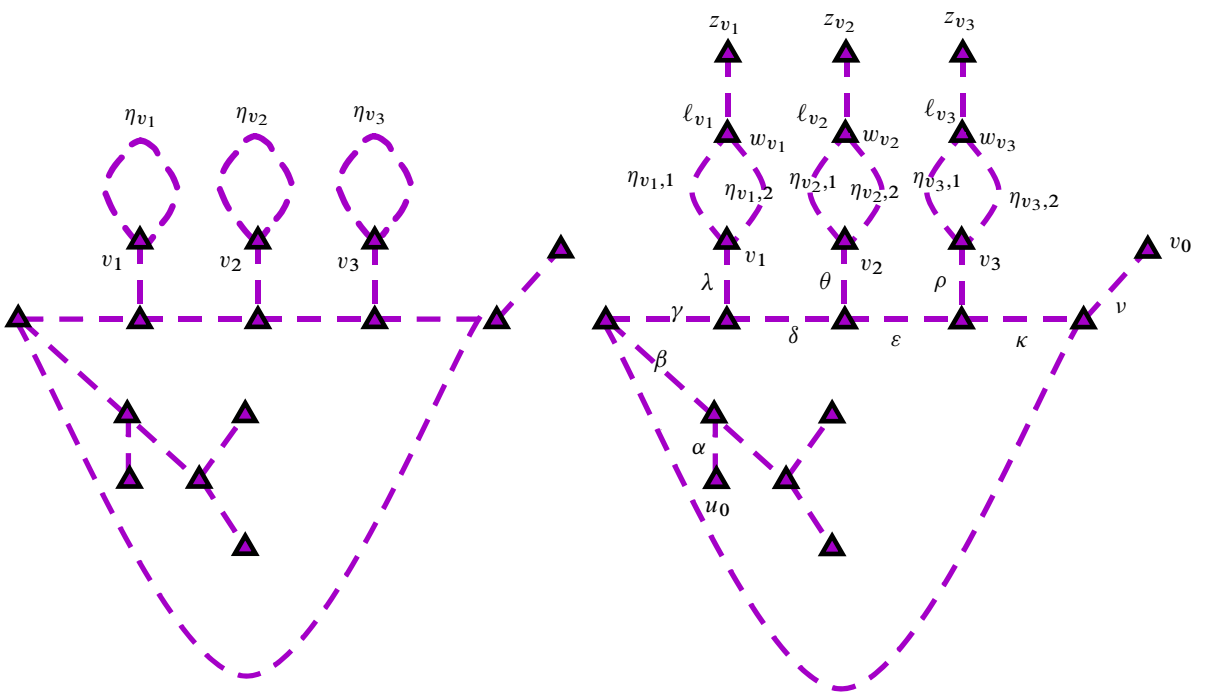


Figure 6

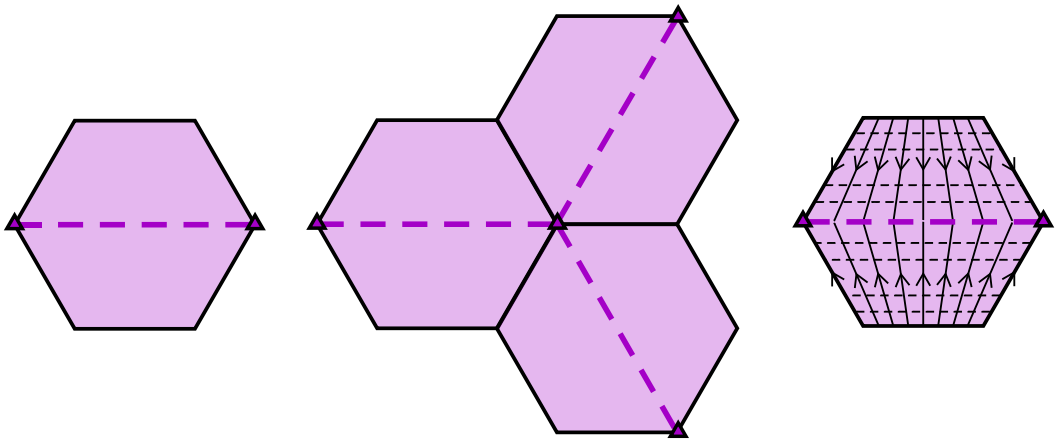


Figure 7

Theorem 3.6 *The leafy dual graph $G^\dagger(\tau)$ is a strong deformation retract of $\Sigma \setminus \mathbb{P}$.*

Proof For each edge e of $G^\dagger(\tau)$, let R_e be an open regular oriented Euclidean hexagon inscribed in the unit circle in the complex plane, and take an embedding of e into R_e as a Euclidean straight line segment joining a pair of radially opposite vertices. We fix such an R_e and such an embedding of e into R_e once and for all; see Figure 7, left.

For each valency-3 vertex u of $G^\dagger(\tau)$, take an edge e of $G^\dagger(\tau)$ containing u, and glue the hexagons R_e, $R_{e^{+,u}}$, and $R_{e^{-,u}}$ as indicated in Figure 7, center; see also (3-1). Denote by $\Sigma(G^\dagger(\tau))$ the result of performing this gluing over all valency-3 vertices of $G^\dagger(\tau)$. We shall refer to $\Sigma(G^\dagger(\tau))$ as the *ribbon surface* of $G^\dagger(\tau)$. We can take a natural open piecewise-differentiable continuous injective function

$$\iota_\tau \colon \Sigma(G^\dagger(\tau)) \to \Sigma \setminus (\mathbb{P} \cup \partial\Sigma) \tag{3-2}$$

that embeds $\Sigma(G^\dagger(\tau))$ as an open subsurface of $\Sigma \setminus \mathbb{P}$. Similarly to [1, Section 2.3; 3, Section 2.2; 4, Section 3.2], it is not hard to see that this embedding admits a strong deformation retraction $\Sigma \setminus \mathbb{P} \to \Sigma(G^\dagger(\tau))$.

For each edge e of $G^\dagger(\tau)$, let $\varrho_e \colon R_e \to e$ be the piecewise-linear strong deformation retraction sketched in Figure 7, right. Define $\varrho \colon \Sigma(G^\dagger(\tau)) \to G^\dagger(\tau)$ by setting $\varrho|_{R_e} := \varrho_e$ for every edge e. Then ϱ is a strong deformation retraction. Composing ϱ with a strong deformation retraction $\Sigma \setminus \mathbb{P} \to \Sigma(G^\dagger(\tau))$ from the previous paragraph, we obtain a strong deformation retraction $\rho \colon \Sigma \setminus \mathbb{P} \to G^\dagger(\tau)$. □

Example 3.7 For the ribbon graph $G^\dagger(\tau)$ from Figure 6, the ribbon surface $\Sigma(G^\dagger(\tau))$ can be visualized in Figure 8. Strictly speaking, the picture is incorrect, since the hexagonal ribbons R_e are not drawn as regular hexagons congruent to each other. Drawing them as such would interfere with an intuitive visualization of $\Sigma(G^\dagger(\tau))$.

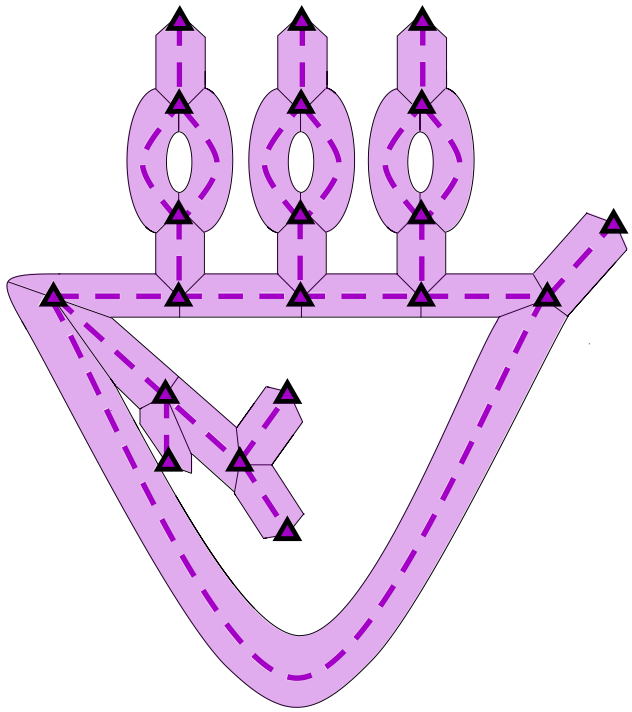


Figure 8

4 Kinks of walks on graphs

From this point on we will work only with triangulations of signature zero. By construction, no edge of $G^{\circ}(\tau)$ is a loop. Hence Theorem 2.6 can be applied. Let f be either

(4-1) a nonidentity morphism in $\pi_1(G^{\circ}(\tau))$ between vertices of $G(\tau)$ distinct from all the self-folded triangles of τ, or

(4-2) a noncontractible free-homotopy class belonging to $\pi_1^{\mathrm{free}}(G^{\circ}(\tau))$.

Define a finite sequence $\boldsymbol{s}(f)=(\varepsilon_1,\varepsilon_2,\dots)$ of signs $\varepsilon_j\in\{+,-\}$ as follows. Let $\phi(f)=(u_0,e_1,\dots,e_n,v_0)$ be the standard form of f. If $n=1$, we set $\boldsymbol{s}(f)$ to be the empty sequence. Otherwise, for $j=1,\dots,n-1$, set ε_j to be the sign with the property that

$$e_{j+1}=e_j^{\varepsilon_j,u_j}.$$

If f belongs to $\pi_1^{\mathrm{free}}(G^{\circ}(\tau))$, we further set ε_n to be the sign with the property that

$$e_1=e_n^{\varepsilon_n,u_n}.$$

We define

$$\boldsymbol{s}(f):=\begin{cases}(\varepsilon_1,\dots,\varepsilon_{n-1}) & \text{if } f \text{ is a morphism in } \pi_1(G),\\ (\varepsilon_1,\dots,\varepsilon_{n-1},\varepsilon_n) & \text{if } f \text{ belongs to } \pi_1^{\mathrm{free}}(G^{\circ}(\tau)).\end{cases}$$

Notice that if $f\in\pi_1^{\mathrm{free}}(G^{\circ}(\tau))$, then $\boldsymbol{s}(f)$ is defined up to a cyclic rotation of its entries.

Example 4.1 Consider the vertices u_0 and v_0 of the graph $G^\circlearrowright(\tau)$ shown in Figure 6. For the following morphisms $f : u_0 \to v_0$ (written in standard form) in $\pi_1(G^\circlearrowright(\tau))$, we can see the sign sequence $\boldsymbol{s}(f)$ on the right (the commas have been omitted for space reasons):

$$\begin{aligned}
f_1 &= (u_0\alpha\beta\gamma\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_1) &= (+----+),\\
f_2 &= (u_0\alpha\beta\gamma\lambda\eta_{v_1,1}\eta_{v_1,2}\lambda\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_2) &= (+-++-++--+),\\
f_3 &= (u_0\alpha\beta\gamma\lambda\eta_{v_1,1}\eta_{v_1,2}\eta_{v_1,1}\eta_{v_1,2}\lambda\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_3) &= (+-++---++--+),\\
f_4 &= (u_0\alpha\beta\gamma\lambda\eta_{v_1,2}\eta_{v_1,1}\lambda\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_4) &= (+-+-+-+--+).
\end{aligned}$$

Definition 4.2 Let τ be a triangulation of signature zero, and $G^\circlearrowright(\tau)$ its leafy dual graph.

(1) If f is a morphism as in (4-1), then a *kink* of f is a segment $(e_j, e_{j+1}, \dots, e_m)$ of the standard form $\phi(f) = (u_0, e_1, \dots, e_n, u_n)$, with the property that $r := \frac{1}{2}(m-j-3)$ is a positive integer and there exists a self-folded triangle v of τ, and distinct indices $i, k \in \{1,2\}$, such that either

(a) $e_{j+r+1} = \eta_{v,i}$ and $e_{j+r+2} = \eta_{v,k}$,

(b) for $t = 1, \dots, r$, we have $e_{j+t} = e_{m-t}$,

(c) $e_{j+r} = e_{j+r+3} \notin \{\eta_{v,k}, \eta_{v,i}\}$, and

(d) $\varepsilon_j = \varepsilon_{j+r+1} = \varepsilon_{m-1}$,

or

(i) $e_{j+1} = e_{j+3} = \cdots = e_{j+2r+1} = \eta_{v,i}$ and $e_{j+2} = e_{j+4} = \cdots = e_{j+2r+2} = \eta_{v,k}$, and

(ii) $e_j = e_m \notin \{\eta_{v,i}, \eta_{v,k}\}$.

(2) If f is a free-homotopy class as in (4-2), then a *kink* of f is a kink of any representative of the rotational equivalence class of the standard form $\phi(f)$.

We say that a kink has *multiplicity* 1 or $r+1$ according to whether it satisfies the first or second set of conditions above. Furthermore, the *core* of κ is its segment (e_{j+r+1}, e_{j+r+2}) if the multiplicity of κ is 1, and its segment $(e_{j+1}, e_{j+2}, \dots, e_{j+2r+1}, e_{j+2r+2})$ if the multiplicity of κ is greater than 1.

Example 4.3 With reference to Example 4.1, the morphisms f_1 and f_2 do not have kinks whatsoever. The morphisms f_3 and f_4 do have kinks (again, for space reasons we omit the commas):

$$\begin{aligned}
f_3 &= (u_0\alpha\beta\underbrace{\gamma\lambda\eta_{v_1,1}\eta_{v_1,2}\eta_{v_1,1}\eta_{v_1,2}\lambda}_{\text{mult}=2,\, j=4, m=9, r=1}\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_3) &= (+-++---++--+),\\
f_4 &= (u_0\alpha\beta\underbrace{\gamma\lambda\eta_{v_1,2}\eta_{v_1,1}\lambda}_{\text{mult}=1,\, j=3, m=8, r=1}\delta\varepsilon\kappa\nu v_0), & \boldsymbol{s}(f_4) &= (+-\boxed{+}-\boxed{+}-\boxed{+}--+).
\end{aligned}$$

For the morphism f_4, the signs $\varepsilon_j = \varepsilon_{j+r+1} = \varepsilon_{m-1}$ from Definition 4.2(d) appear enclosed in squares.

Definition 4.4 Let τ be a triangulation of signature zero, and $G^{\Upsilon}(\tau)$ its leafy dual graph.

(1) Suppose that f is a morphism as in (4-1) and that $\kappa = (e_j, e_{j+1}, \dots, e_m)$ is a kink of f, and set $r := \frac{1}{2}(m - j + 3)$ as in Definition 4.2. A *partial resolution* of κ in f is the sequence $\tilde{\rho}_\kappa(f)$ obtained after applying one of the following two combinatorial operations to f:

- If κ has multiplicity 1, switch the $(j+r+1)^{\text{th}}$ and $(j+r+2)^{\text{th}}$ entries of f.
- If κ has multiplicity $r+1 > 1$, choose an index $s \in \{j+1, \dots, j+2r+1\}$ and switch the s^{th} and $(s+1)^{\text{th}}$ entries of f.

(2) Suppose that f is a free-homotopy class as in (4-2), and that $\kappa = (e_j, e_{j+1}, \dots, e_m)$ is a kink of f. Suppose further that $\phi(f) = \{f_1, \dots, f_n\}$ (see Definitions 2.7 and 4.2, as well as Remark 2.5), and that κ appears as a kink of $f_j \in \phi(f)$. A *partial resolution* of κ in f is the sequence $\tilde{\rho}_\kappa(f_j)$ obtained by applying one of the two combinatorial operations just described to f_j.

Suppose that f is as in (4-1) or (4-2), and κ is a kink of f. If f is a morphism and κ has multiplicity 1, then the sequence $\tilde{\rho}_\kappa(f)$ is well defined, ie uniquely determined by f and κ, whereas if k has multiplicity $r+1 > 1$, then there is ambiguity in the definition of $\tilde{\rho}_\kappa(f)$ as a sequence. Furthermore, if f is a free-homotopy class, then κ does not appear as a kink of all the representatives $f_1, \dots, f_n$, as rotation of walks eventually breaks the appearance of κ as a subsequence. However, we do have the following result, whose proof is immediate:

Proposition 4.5 *Let τ be a triangulation of signature zero and $G^{\Upsilon}(\tau)$ its leafy dual graph, and let f be as in (4-1) or (4-2).*

(1) *If f is a morphism, then for any kink κ of f, the sequence $\phi(\tilde{\rho}_\kappa(f))$ is uniquely determined by f and κ. Thus, for any kink κ of multiplicity greater than 1, the sequence $\phi(\tilde{\rho}_\kappa(f))$ is independent of the index s chosen in Definition 4.4.*

(2) *If f is a free-homotopy class, with standard form $\phi(f) = \{f_1, \dots, f_n\}$, and κ appears as a kink of f_i and f_j, with such appearances being brought to each other by the corresponding rotations that bring f_i and f_j to each other, then $[\phi(\tilde{\rho}_\kappa(f_i))]_{\text{rot}} = [\phi(\tilde{\rho}_\kappa(f_j))]_{\text{rot}}$.*

Furthermore, the number of kinks of $\phi(\tilde{\rho}_\kappa(f))$ or $[\phi(\tilde{\rho}_\kappa(f_j))]_{\text{rot}}$, counted with multiplicity, is strictly less than the number of kinks of f, counted with multiplicity.

Definition 4.6 In the situation of Proposition 4.5,

$$\rho_\kappa(f) := \begin{cases} \phi(\tilde{\rho}_\kappa(f)) & \text{if } f \text{ is a morphism,} \\ [\phi(\tilde{\rho}_\kappa(f_j))]_{\text{rot}} & \text{if } f \text{ is a free-homotopy class,} \end{cases}$$

is the *resolution* of κ in f.

Lemma 4.7 *Let τ be a triangulation of signature zero, and $G^{\Upsilon}(\tau)$ its leafy dual graph, and let f be as in (4-1) or (4-2).*

- *Resolving a kink of multiplicity 1 makes it disappear but does not affect the length of f.*

- *Resolving a kink of multiplicity* 2 *makes it disappear and makes the length of* f *drop by at least* 4.
- *Resolving a kink* κ *of* f *of multiplicity* 3 *makes the multiplicity of* κ *drop by* 2 *or* 3 *and the length of* f *drop by* 4.
- *Resolving a kink* κ *of* f *of multiplicity greater than* 3 *makes the multiplicity of* κ *drop by* 2 *and the length of* f *drop by* 4.

The proof of Lemma 4.7 is left to the reader.

Lemma 4.8 *Let* u_0, v_0, w_0 *be vertices of* $G(\tau)$ *distinct from all the self-folded triangles of* τ. *Suppose* $f = (u_0, e_1, e_2, \dots, e_n, v_0)$ *and* $g = (v_0, e_n, \dots, e_2, d_1, w_0)$ *are nonidentity morphisms in* $\pi_1(G^\circlearrowleft(\tau))$, *both written in standard form, with* $e_1 \neq d_1$, *and that* $\kappa = (e_j, \dots, e_m)$ *is a kink of* $(e_2, \dots, e_n)$.

(1) *If the multiplicity of* κ *is* 1, *then for* $r := \frac{1}{2}(m - j - 3)$ *we have*

$$\phi(\rho_\kappa(f) * g) = (u_0, e_1, \dots, e_{j+r-1}, \underbrace{e_{j+r}, e_{j+r+2}, e_{j+r+1}, e_{j+r+2}, e_{j+r+1}, e_{j+r}}_{\text{entries } j+r \text{ to } j+r+5}, e_{j+r-1}, \dots, d_1, w_0).$$

(2) *If the multiplicity of* κ *is greater than* 1, *then*

$$\phi(\rho_\kappa(f) * g) = (u_0, e_1, e_2, \dots, e_{j-1}, \underbrace{e_j, e_{j+1}, e_{j+2}, e_{j+1}, e_{j+2}, e_j}_{\text{entries } j \text{ to } j+5}, e_{j-1}, \dots, e_2, d_1, w_0).$$

In any case, $\phi(\rho_\kappa(f) * g)$ *has a kink* κ' *of multiplicity* 2, *such that*

$$\phi(f * g) = (u_0, e_1, d_1, w_0) = \rho_{\kappa'}(\phi(\rho_\kappa(f) * g)).$$

Proof In the case where the multiplicity of κ is 1, the computation of $\phi(\rho_\kappa(f) * g)$ can be seen in Figure 9, whereas in Figure 10 we have included the computation of $\phi(\rho_\kappa(f) * g)$ in the case where the multiplicity of κ is exactly 3. The computation of $\phi(\rho_\kappa(f) * g)$ in the case where the multiplicity of κ is an arbitrary integer greater than 1 is completely analogous.

With $\phi(\rho_\kappa(f) * g)$ explicitly computed, the assertions of the lemma can be directly verified. □

Proposition 4.9 *Let* τ *be a triangulation of signature zero and* $G^\circlearrowleft(\tau)$ *its leafy dual graph, and let* φ *be as in (4-1) or (4-2). If* κ_0 *and* κ_0' *are distinct kinks of* φ, *then there exist finite sequences of kinks* $(\kappa_1, \dots, \kappa_l)$ *and* $(\kappa_1', \dots, \kappa_{l'}')$ *such that*

(1) *for each* $t = 1, \dots, l$, κ_t *is a kink of* $\rho_{\kappa_{t_1}} \rho_{\kappa_{t-2}} \cdots \rho_{\kappa_0}(\varphi)$,

(2) *for each* $t = 1, \dots, l'$, κ_t' *is a kink of* $\rho_{\kappa'_{t_1}} \rho_{\kappa'_{t-2}} \cdots \rho_{\kappa'_0}(\varphi)$,

(3) $\rho_{\kappa_l} \cdots \rho_{\kappa_1} \rho_{\kappa_0}(\varphi) = \rho_{\kappa'_{l'}} \cdots \rho_{\kappa'_1} \rho_{\kappa'_0}(\varphi)$ *if* φ *is a morphism,*

(3′) $\rho_{\kappa_l} \cdots \rho_{\kappa_1} \rho_{\kappa_0}(\varphi)$ *is equal to* $\rho_{\kappa'_{l'}} \cdots \rho_{\kappa'_1} \rho_{\kappa'_0}(\varphi)$ *or its opposite orientation if* φ *is a free-homotopy class as in (4-2).*

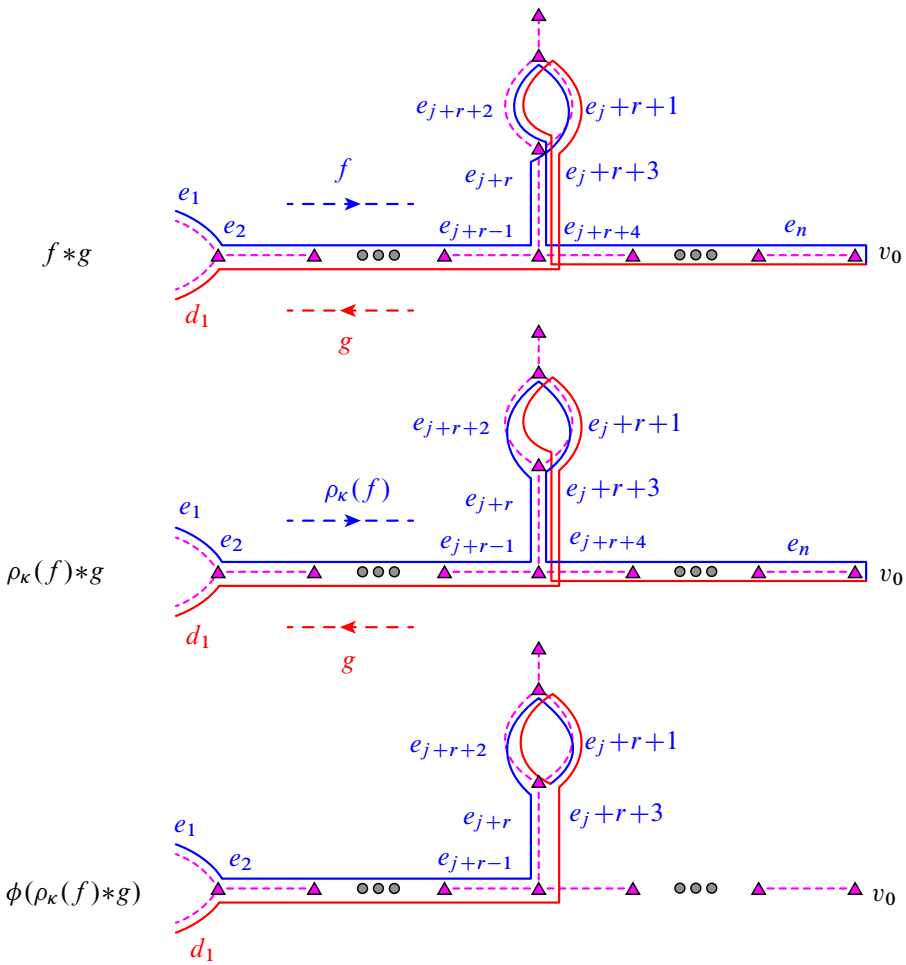


Figure 9

Proof Throughout the proof we will assume, without loss of generality, that φ is already written in standard form.

Suppose first that φ is a morphism as in (4-1). We distinguish five cases:

(i) when each of κ_0 and κ_0' has multiplicity greater than 2,

(ii) when one of κ_0 and κ_0' has multiplicity greater than 2 and the other one has multiplicity equal to 2,

(iii) when each of κ_0 and κ_0' has multiplicity equal to 2,

(iv) when one of κ_0 and κ_0' has multiplicity equal to 2 and the other one has multiplicity equal to 1,

(v) when each of κ_0 and κ_0' has multiplicity equal to 1.

Case 1 Using Lemma 4.7, it is easy to verify that if each of κ_0 and κ_0' has multiplicity greater than 2, then κ_0 is a kink of $\rho_{\kappa_0'}(\varphi)$, κ_0' is a kink of $\rho_{\kappa_0}(\varphi)$, and $\rho_{\kappa_0}(\rho_{\kappa_0'}(\varphi)) = \rho_{\kappa_0'}(\rho_{\kappa_0}(\varphi))$, so the proposition follows in this case.

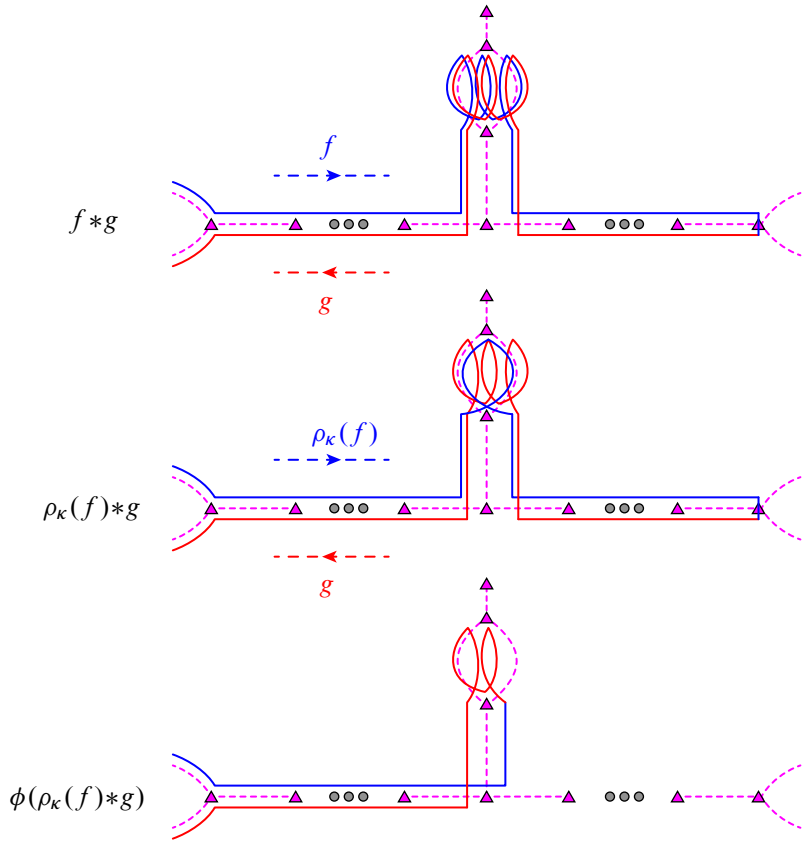


Figure 10

Case 2 Suppose that the multiplicity of κ_0 is 2 and the multiplicity of κ_0' is greater than 2. Then κ_0 is a kink of $\rho_{\kappa_0'}(\varphi)$ by Lemma 4.7, and φ can be written as a concatenation

$$\varphi = \alpha * f * \kappa_0 * g * \beta,$$

where α, f, g, and β are morphisms in $G^{\diamond}(\tau)$ between vertices of $G(\tau)$ distinct from all the self-folded triangles of τ, with $f = (u_0, e_1, e_2, \ldots, e_n, v_0)$ and $g = (v_0, e_n, \ldots, e_2, d_1, w_0)$, and either

$$\begin{cases} e_1 = d_1 \text{ and } \alpha = \mathbb{1}_{u_0} = \beta, & \text{or} \\ e_1 \neq d_1. \end{cases}$$

If $e_1 \neq d_1$, then either κ_0' is disjoint from f and from g, or it overlaps one of them. If κ_0' is entirely contained in f, then we can apply Lemma 4.8 to obtain a kink κ_1' of $\rho_{\kappa_0}\rho_{\kappa_0'}(\varphi) = \phi(\alpha * \rho_{\kappa_0'}(f) * g * \beta)$ such that

$$\rho_{\kappa_0}(\varphi) = \phi(\alpha * f * g * \beta) = \rho_{\kappa_1'}\rho_{\kappa_0}\rho_{\kappa_0'}(\varphi),$$

which is the assertion of the proposition. Essentially the same argument shows that the proposition holds also when κ_0' is entirely contained in g, or is disjoint from f and from g, or overlaps one of them without being entirely contained in either.

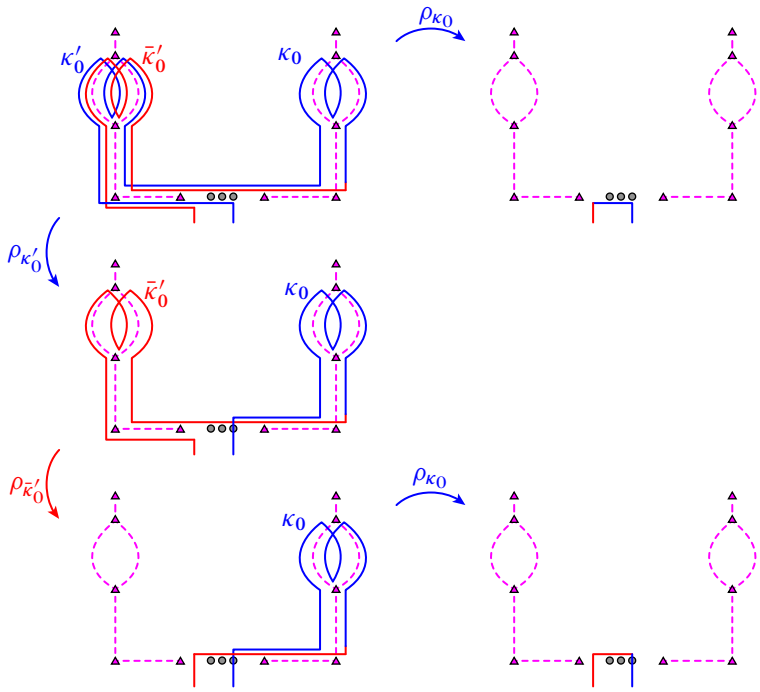


Figure 11

The same argument works also if $e_1 = d_1$ and $\alpha = \mathbb{1}_{u_0} = \beta$; notice that in this case $\rho_{\kappa_0}(\varphi) = \mathbb{1}_{u_0}$.

Case 3 Suppose that each of κ_0 and κ_0' has multiplicity equal to 2. Then φ can be written as concatenations

$$\alpha * f * \kappa_0 * g * \beta = \varphi = \alpha' * f' * \kappa_0' * g' * \beta',$$

where α, α', f, f', g, g', β, and β' are morphisms in $G^{\circ}(\tau)$ between vertices of $G(\tau)$ that are distinct from all the self-folded triangles of τ, with $f = (u_0, e_1, e_2, \dots, e_n, v_0)$, $g = (v_0, e_n, \dots, e_2, d_1, w_0)$, $f' = (u_0', e_1', e_2', \dots, e_n', v_0')$, $g' = (v_0', e_n', \dots, e_2', d_1', w_0')$, and either

$$\begin{cases} e_1 = d_1 \text{ and } \alpha = \mathbb{1}_{u_0} = \beta, & \text{or} \\ e_1 \neq d_1. \end{cases}$$

If $e_1 \neq d_1$, then κ_0' either is disjoint from f and from g, or it overlaps one of them. We analyze the situation where κ_0' is entirely contained in f, and leave in the hands of the reader the analysis of the remaining possibilities.

So, suppose κ_0' is entirely contained in f. Then κ_0 either is disjoint from f' and from g', or it overlaps one of them. If κ_0 is disjoint from both f' and g', then Figure 11 shows us a kink $\bar\kappa_0'$ of $\rho_{\kappa_0'}(\varphi)$ such that κ_0 is a kink of $\rho_{\bar\kappa_0'}\rho_{\kappa_0'}(\varphi)$ and

$$\rho_{\kappa_0}(\varphi) = \rho_{\kappa_0}\rho_{\bar\kappa_0'}\rho_{\kappa_0'}(\varphi).$$

If κ_0 is entirely contained in f, then Figure 12 shows us kinks $\bar\kappa_0$ and $\bar\kappa_0'$ such that

$$\rho_{\bar\kappa_0'}\rho_{\kappa_0'}(\varphi) = \rho_{\bar\kappa_0}\rho_{\kappa_0}(\varphi).$$

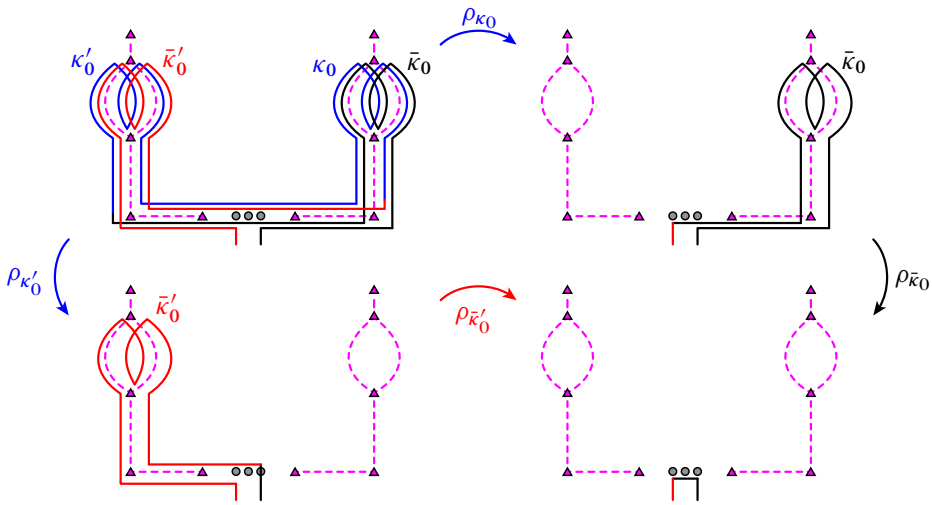


Figure 12

We leave the analysis of the cases where κ_0 is entirely contained in g', or it overlaps f' or g' without being entirely contained in either, to the reader.

Case 4 If one of κ_0 and κ_0' has multiplicity equal to 2 and the other one has multiplicity equal to 1, one can resort to Lemma 4.8 as we did in Case 2 to obtain the assertion of the proposition.

Case 5 If each of κ_0 and κ_0' has multiplicity equal to 1, then $\rho_{\kappa_0'}\rho_{\kappa_0}(\varphi) = \rho_{\kappa_0}\rho_{\kappa_0'}(\varphi)$.

Suppose now that φ is a free-homotopy class as in (4-2). If both κ_0 and κ_0' appear as kinks of at least one of the closed backtrack-free walks f representing φ, considering these walks as morphisms in $\pi_1(G^{\circlearrowright}(\tau))$, and if the standard forms of the morphisms $\rho_{\kappa_0}(f)$ and $\rho_{\kappa_0'}(f)$ are closed backtrack-free walks (see Definition 2.4), then the result follows from the proof we have just given above. Otherwise, φ is represented by one of the closed backtrack-free walks depicted in Figures 13 and 14 (in Figure 13, κ_0 and κ_0' are images of antipodal segments of $\mathbb{S}^1$, whereas in Figure 14 they are not; see Remark 4.10 below).

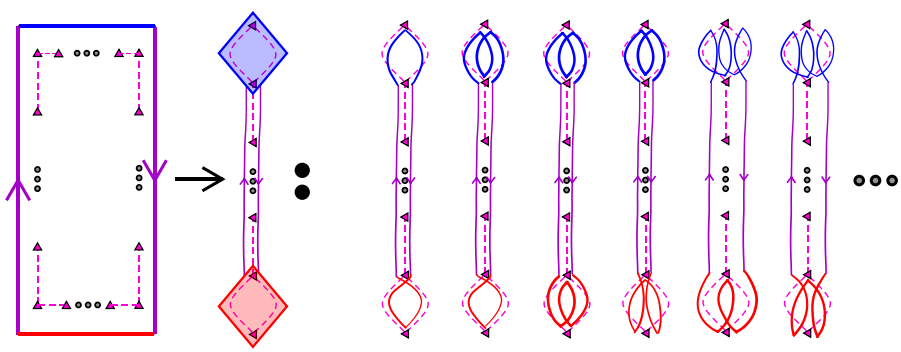


Figure 13

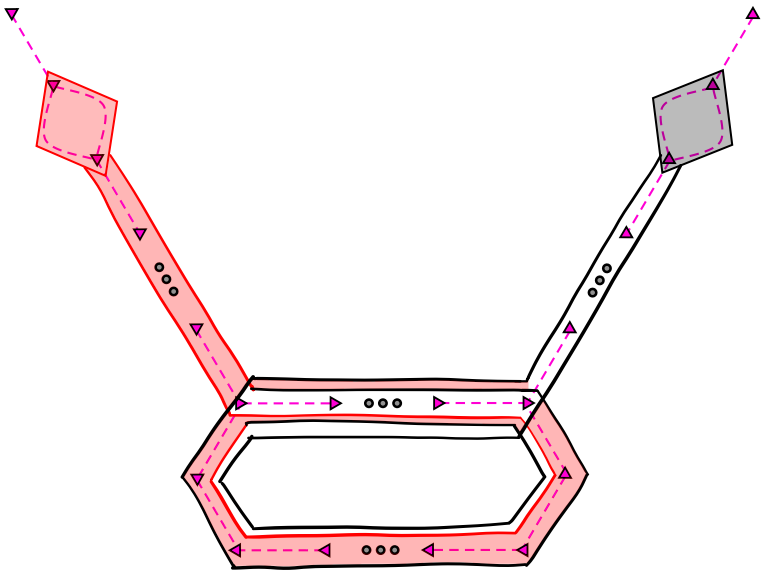


Figure 14

In the situation of Figure 13, the assertion of the proposition follows easily (it is here that one may have to take opposite orientations of the closed curves). The situation of Figure 14 is dealt with in Figure 15, where the assertion of the proposition is proved in the cases where the multiplicities of κ_0 and κ_0' are less than 3. The proposition is very easy to verify in the cases where the multiplicities are at least 3 (see Lemma 4.7). □

Remark 4.10 Consider a rectangle Q, made into a graph in such a way that the number of edges constituting the left vertical side of Q is equal to the number of edges constituting the right vertical side of Q, and each of the horizontal sides of Q is conformed by an even number of edges (the number of edges in the top side of Q is not necessarily equal to the number of edges in the bottom side). What we mean by "κ_0 and κ_0' are images of antipodal segments of $\mathbb{S}^1$" in the final part of the proof of Proposition 4.9 is that the backtrack-free walk representing φ is given by a graph homomorphism $Q \to G^{\circledcirc}(\tau)$ as depicted in Figure 13, ie sending the horizontal sides of Q to the respective subsequences $(e_{j+1}, \dots, e_{j+2r+2}) = (\eta_{v,i}, \eta_{v,k}, \dots, \eta_{v_i}, \eta_{v,k})$ of κ_0 and κ_0', and the vertical sides of Q to opposite orientations of the same walk on $G^{\circledcirc}(\tau)$. Depending on the parities of κ_0 and κ_0', this sometimes has the effect that resolving κ_0 first and resolving κ_0' first result in opposite orientations of the same closed walk on $G^{\circledcirc}(\tau)$.

Theorem 4.11 *Let τ be a triangulation of signature zero, $G^{\circledcirc}(\tau)$ be its leafy dual graph, and f be as in (4-1) or (4-2). Suppose that $(\kappa_1, \dots, \kappa_l)$ and $(\kappa_1', \dots, \kappa_{l'}')$ are finite sequences of kinks such that*

(1) *for each $t = 1, \dots, l$, κ_t is a kink of $\rho_{\kappa_{t_1}} \rho_{\kappa_{t-2}} \cdots \rho_{\kappa_0}(f)$,*

(2) *for each $t = 1, \dots, l'$, κ_t' is a kink of $\rho_{\kappa'_{t_1}} \rho_{\kappa'_{t-2}} \cdots \rho_{\kappa'_0}(f)$,*

(3) *neither $\rho_{\kappa_l} \cdots \rho_{\kappa_1} \rho_{\kappa_0}(f)$ nor $\rho_{\kappa'_{l'}} \cdots \rho_{\kappa'_1} \rho_{\kappa'_0}(f)$ has kinks.*

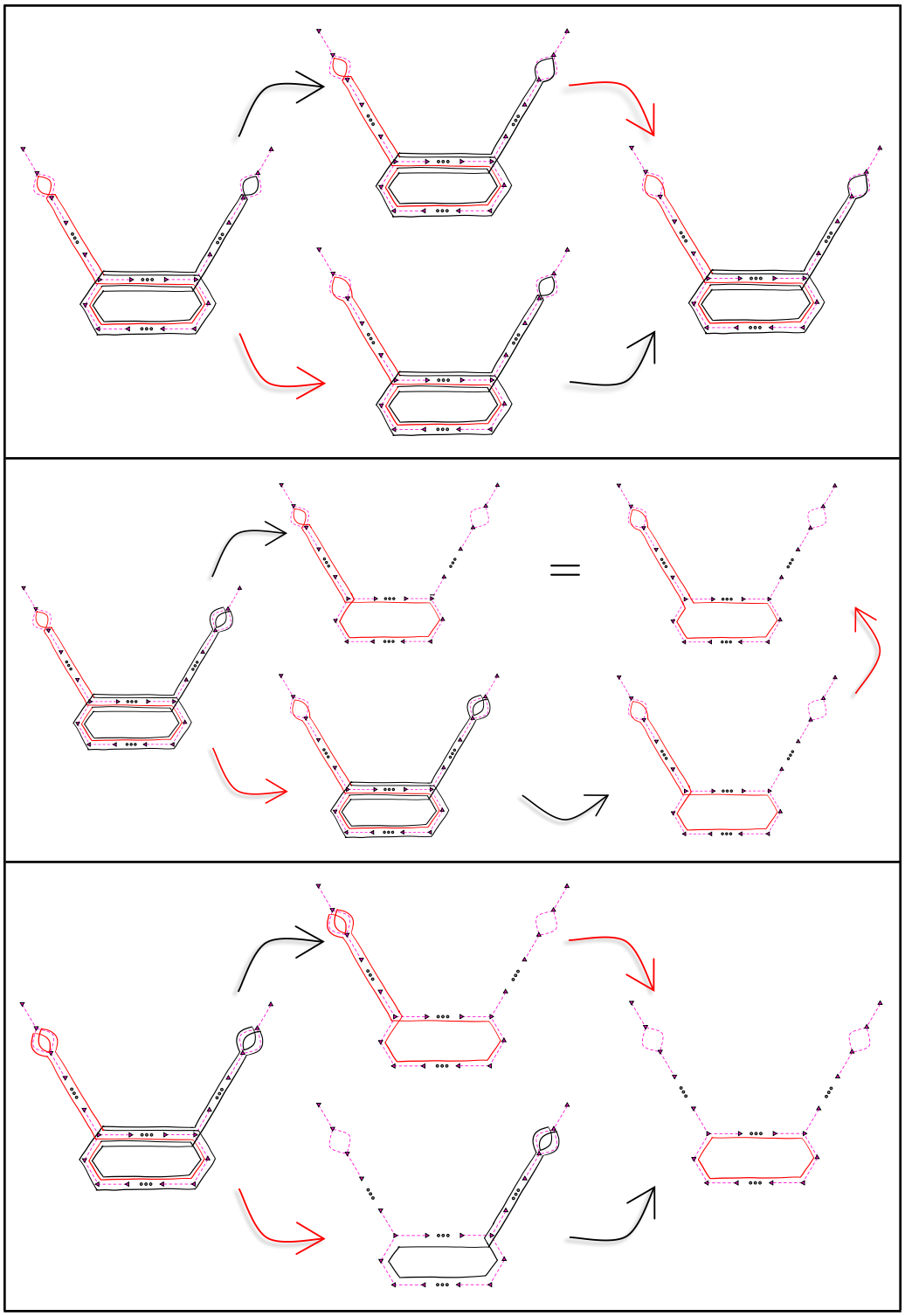


Figure 15

Then

- $\rho_{\kappa_l}\cdots\rho_{\kappa_1}\rho_{\kappa_0}(f)=\rho_{\kappa'_{l'}}\cdots\rho_{\kappa'_1}\rho_{\kappa'_0}(f)$ *if* φ *is a morphism*,
- $\rho_{\kappa_l}\cdots\rho_{\kappa_1}\rho_{\kappa_0}(\varphi)$ *is equal to* $\rho_{\kappa'_{l'}}\cdots\rho_{\kappa'_1}\rho_{\kappa'_0}(\varphi)$ *or its opposite orientation if* φ *is a free-homotopy class as in (4-2).*

Proof By Propositions 4.5 and 4.9, resolutions of kinks satisfy the hypotheses of the diamond lemma; see [14, Lemma 2.4], [20, Theorem 3] or [23, Theorem 1.7.10]. The result follows. □

5 Kinks of curves via kinks of walks on graphs

Definition 5.1 A continuous function $\gamma\colon[0,1]\to\Sigma(G^{\circledcirc}(\tau))$ connecting valency-1 vertices of $G(\tau)$ (resp. continuous closed curve $\gamma\colon\mathbb{S}^1\to\Sigma(G^{\circledcirc}(\tau))$) is called a *curve adapted to the ribbon structure of* $\Sigma(G^{\circledcirc}(\tau))$ if there is a finite partition $0=t_0<t_1<\cdots<t_l=1$ of the interval $[0,1]$ such that

(1) for $i=1,\dots,l$, there is an edge e_i of $G^{\circledcirc}(\tau)$ such that the image of the open interval (t_{i-1},t_i) (resp. the circle segment $e^{2\pi i(t_{i-1},t_i)}$) under γ is entirely contained in the interior of R_{e_i},

(2) for $i=2,\dots,l$, the edges e_{i-1} and e_i are distinct,

(3) for $i=1,\dots,l$, the image $\gamma([t_{i-1},t_i])\subseteq R_{e_i}$ (resp. $\gamma(e^{2\pi i[t_{i-1},t_i]})\subseteq R_{e_i}$) has exactly one element in common with each fiber of the map $\varrho_{e_i}\colon R_{e_i}\to e_i$ introduced in the proof of Theorem 3.6,

(4) for $i,j=1,\dots,l$, if $i\neq j$, then $\gamma([t_{i-1},t_i])\cap\gamma([t_{j-1},t_j])$ (resp. $\gamma(e^{2\pi i[t_{i-1},t_i]})\cap\gamma(e^{2\pi i[t_{j-1},t_j]})$) is a finite set.

Notice that if $\gamma\colon[0,1]\to\Sigma(G^{\circledcirc}(\tau))$ is a curve adapted to the ribbon structure of $\Sigma(G^{\circledcirc}(\tau))$, then the sequence $(\gamma(0),e_1,e_2,\dots,e_l,\gamma(1))$ is a morphism written in standard form in the fundamental groupoid $\pi_1(G^{\circledcirc}(\tau))$.

The next result can be seen to be a consequence of Theorems 2.6 and 3.6. Alternatively, it follows from [13].

Theorem 5.2 *Let* τ *be a signature-zero triangulation, and let* $\varrho\colon\Sigma(G^{\circledcirc}(\tau))\to G^{\circledcirc}(\tau)$ *be the strong deformation retraction defined in the proof of Theorem 3.6.*

(1) *Every continuous curve* $\delta\colon[0,1]\to\Sigma(G^{\circledcirc}(\tau))$ *connecting valency-1 vertices of* $G(\tau)$ *is homotopic rel* $\{0,1\}$ *to* $\gamma\colon[0,1]\to\Sigma(G^{\circledcirc}(\tau))$, *a curve adapted to the ribbon structure of* $\Sigma(G^{\circledcirc}(\tau))$, *say, with respect to a partition* $0=t_0<t_1<\cdots<t_l=1$ *of the interval* $[0,1]$. *Furthermore, the sequence* $(\gamma(0),e_1,e_2,\dots,e_l,\gamma(1))$ *from Definition 5.1 is a backtrack-free walk that coincides with the standard form of* $\varrho\delta$, *and is thus uniquely determined by* δ.

(2) *Every continuous closed curve* $\delta\colon\mathbb{S}^1\to\Sigma(G^{\circledcirc}(\tau))$ *is freely homotopic to* $\gamma\colon\mathbb{S}^1\to\Sigma(G^{\circledcirc}(\tau))$, *a closed curve adapted to the ribbon structure of* $\Sigma(G^{\circledcirc}(\tau))$, *say, with respect to a partition* $0=t_0<t_1<\cdots<t_l=1$ *of the interval* $[0,1]$. *Furthermore, the sequence* $(\gamma(1),e_1,e_2,\dots,e_l,\gamma(1))$ *from Definition 5.1 is a closed backtrack-free walk whose rotational equivalence class coincides with the standard form of* $\varrho\delta$, *and is thus uniquely determined by* δ.

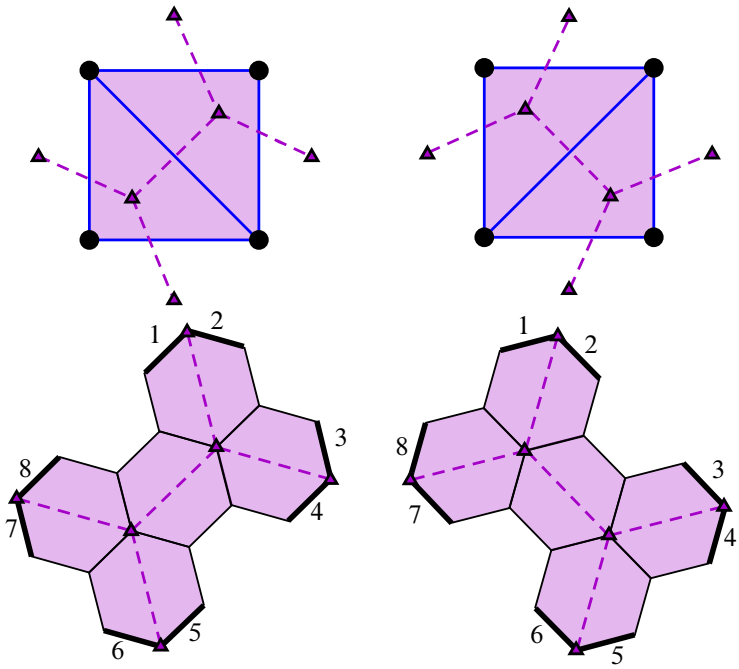


Figure 16

Definition 5.3 Let τ be a signature-zero triangulation and $\varrho\colon \Sigma(G^\circledcirc(\tau)) \to G^\circledcirc(\tau)$ the strong deformation retraction defined in the proof of Theorem 3.6. Let $\gamma\colon [0,1] \to \Sigma(G^\circledcirc(\tau))$ (resp. $\gamma\colon \mathbb{S}^1 \to \Sigma(G^\circledcirc(\tau))$) be a curve adapted to the ribbon structure of $\Sigma(G^\circledcirc(\tau))$ with respect to a partition $0 = t_0 < t_1 < \cdots < t_l = 1$ of the interval $[0,1]$. A *kink of γ with respect to τ* is a segment $\gamma|_{[t_j,t_m]}\colon [t_j,t_m] \to \Sigma(G^\circledcirc(\tau))$ (resp. $\gamma|_{e^{2\pi i[a,b]}}\colon e^{2\pi i[a,b]} \to \Sigma(G^\circledcirc(\tau))$) such that the subsequence $(e_j,\dots,e_m)$ of $\phi(\varrho\delta)$ it determines is a kink of $\phi(\varrho\delta)$.

Proposition 5.4 *Suppose that τ and σ are triangulations of signature zero. Let γ_1 and γ_2 be curves adapted to the ribbon structures of $\Sigma(G^\circledcirc(\tau))$ and $\Sigma(G^\circledcirc(\sigma))$, respectively. If $\iota_\tau(\gamma_1)$ and $\iota_\sigma(\gamma_2)$ are homotopic rel $\{0,1\}$ (resp. freely homotopic) in Σ (see (3-2)), then γ_1 has at least one kink if and only if γ_2 has at least one kink.*

Proof Any two triangulations of $(\Sigma,\mathbb{M},\mathbb{P})$ of signature zero are related by a finite sequence of $\diamondsuit$-flips; see [9, Definition 9.11]. Hence it suffices to prove the proposition in the case where τ and σ are related by a single $\diamondsuit$-flip. Furthermore, it is enough to show that, *given* γ_1, *there exists* a γ_2 with the properties that $\iota_\tau(\gamma_1)$ and $\iota_\sigma(\gamma_2)$ are homotopic rel $\{0,1\}$ (resp. freely homotopic) in Σ, and γ_1 has at least one kink if and only if γ_2 has at least one kink.

There are two types of $\diamondsuit$-flips, namely, the ones sketched in Figures 16 and 17. In each of these figures, we have drawn the subgraphs of $G^\circledcirc(\tau)$ and $G^\circledcirc(\sigma)$ conformed by the edges incident to the triangles of τ

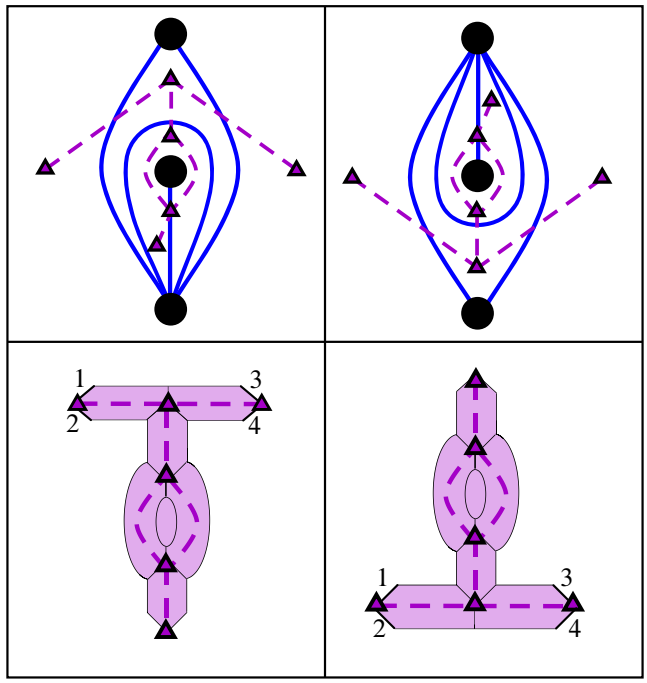


Figure 17

and σ that contain the arcs involved in the $\diamondsuit$-flip. We have also drawn the ribbon subsurfaces of $\Sigma(G^{\circ}(\tau))$ and $\Sigma(G^{\circ}(\sigma))$ induced by these subgraphs. We shall denote these ribbon subsurfaces by $S(\tau)$ and $S(\sigma)$.

Outside the portions depicted in Figures 16 and 17, $G^{\circ}(\tau)$ and $G^{\circ}(\sigma)$ are identical, and $\Sigma(G^{\circ}(\tau))$ and $\Sigma(G^{\circ}(\sigma))$ are identical. Thus, in order to exhibit a γ_2 if we are given γ_1, it suffices to replace each segment s of γ_1 contained in $S(\tau)$ by a segment s' contained in $S(\sigma)$, connecting the endpoints of s, and satisfying the conditions in Definition 5.1, in such a way that $\iota_\tau(s)$ and $\iota_\sigma(s')$ are homotopic rel $\{0,1\}$ in $\Sigma \setminus (\mathbb{P} \cup \partial\Sigma)$. This replacement can be carried out combinatorially very easily, and has been sketched in Figure 18.

Letting γ_2 be the result of performing all of the above replacements, it becomes routine to verify that γ_1 has a kink with respect to τ if and only if γ_2 has a kink with respect to σ. □

Remark 5.5 With the current definition of the multiplicity of a kink on a (closed) backtrack-free walk on $G^{\circ}(\tau)$ (see Definition 4.2), for kinks of positive multiplicity it is not always true that their multiplicity is invariant under $\diamondsuit$-flips. However, a suitable refinement of the notion of multiplicity can be given that is invariant under $\diamondsuit$-flips. It would also be interesting to define the notion of a kink with respect to a triangulation not necessarily having signature zero, and proving the corresponding invariance under flips.

Remark 5.6 By the work of Hansper [12], the two possible notions of an *admissible* string or band on a clannish algebra are equivalent; see also [10]. Now, it is extremely easy to see that if τ is a triangulation of

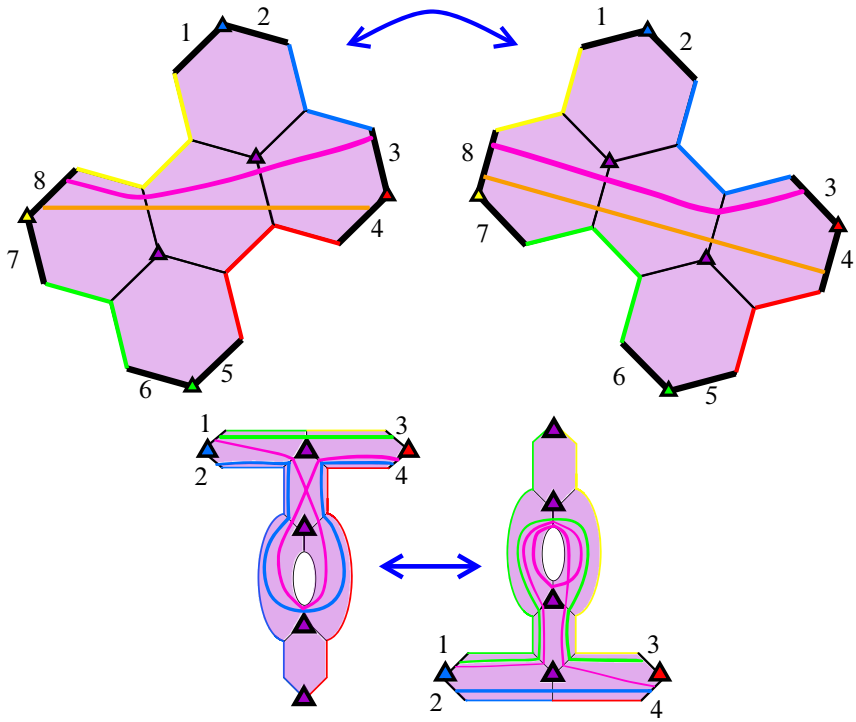


Figure 18

signature zero, then the Jacobian algebra $\Lambda(\tau)$, associated to τ by the second author in [16], is a clannish algebra. It turns out that the string or band on $\Lambda(\tau)$ determined by a curve on $(\Sigma, \mathbb{M}, \mathbb{P})$ is admissible if and only if the curve itself has no kinks.

6 Equivalence of orbifold fundamental groupoids

Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a surface with nonempty boundary. We define a subgroupoid of the fundamental groupoid $\pi_1(\Sigma \setminus \mathbb{P})$ as follows. Form a set E by picking exactly one point from each connected component of $(\partial\Sigma) \setminus \mathbb{M}$. Define $\pi_1(\Sigma \setminus \mathbb{P}, E)$ to be the full subcategory of $\pi_1(\Sigma \setminus \mathbb{P})$ having E as its set of objects.

For each point $u \in E$, each puncture $p \in \mathbb{P}$, and each continuous function $c\colon [0,1] \to \Sigma$ connecting u to p such that $(\partial\Sigma \cup \mathbb{P}) \cap c(0,1) = \varnothing$, take a continuous function $\gamma_{u,p,c}\colon [0,1] \to \Sigma$ with $\gamma_{u,p,c}(0) = u = \gamma_{u,p,c}(1)$ and $(\partial\Sigma \cup \mathbb{P}) \cap c(0,1) = \varnothing$, closely following c but surrounding p clockwise instead of going into it. Thus $\gamma_{u,p,c}$ cuts out a once-punctured monogon based at u, the puncture being p.

For each $u \in E$ form the set

$$S(u) := \{[\gamma_{u,p,c}]^2 \mid p \in \mathbb{P} \text{ and } c \text{ is as above}\} \subseteq \pi_1(\Sigma \setminus \mathbb{P}, E)(u,u) = \pi_1(\Sigma \setminus \mathbb{P})(u,u).$$

Although $S(u)$ fails to be a subgroup of $\pi_1(\Sigma \setminus \mathbb{P}, E)(u,u)$, the collection $S := (S(u))_{u\in E}$ has the property that for every $u_0, v_0 \in E$, every $h \in S(u_0)$, and every morphism $g \in \pi_1(\Sigma \setminus \mathbb{P}, E)(x,y)$, we have $ghg^{-1} \in S(v_0)$. Thus, if we let $H(u)$ be the subgroup of $\pi_1(\Sigma \setminus \mathbb{P}, E)(u,u)$ generated by $S(u)$, then $H := (H(u))_{u\in E}$ is a normal multilocular subgroup of $\pi_1(\Sigma \setminus \mathbb{P}, E)$.

Definition 6.1 [5; 7] The quotient groupoid

$$\pi^{\mathrm{orb}}_{1,\mathbb{P}}(\Sigma, E) := \pi_1(\Sigma \setminus \mathbb{P}, E)/H$$

will be called the 2-*orbifold fundamental groupoid* of $(\Sigma, \mathbb{M}, \mathbb{P})$.

Let τ be a triangulation of $(\Sigma, \mathbb{M}, \mathbb{P})$ having signature zero. Each point u in E belongs to exactly one boundary segment b_u of Σ. This segment b_u is a vertex of both of the graphs $G(\tau)$ and $G^{\diamond}(\tau)$. We will identify each u as the vertex b_u whenever we consider either of these graphs. Under this identification, we define $\pi_1(G^{\diamond}(\tau), E)$ to be the full subcategory of $\pi_1(G^{\diamond}(\tau))$ having E as set of objects.

For each vertex $u \in E$, each self-folded triangle v of τ, and each morphism $c \in G^{\diamond}(\tau)(u,v)$, set $\gamma_{u,v,c} := c * (\eta_{v,1}, \eta_{v_2}) * c^{-1} = c^{-1} \circ (\eta_{v,1}, \eta_{v_2}) \circ c$. For each $u \in E$ form the set

$$R(u) := \{\gamma^2_{u,v,c} \mid v \text{ is as above and } c \in G^{\diamond}(\tau)(u,v)\} \subseteq \pi_1(G^{\diamond}(\tau), E)(u,u) = \pi_1(G^{\diamond}(\tau))(u,u).$$

Although $R(u)$ fails to be a subgroup of $\pi_1(G^{\diamond}(\tau), E)(u,u)$, the collection $R := (R(u))_{u\in E}$ has the property that for every $u_0, v_0 \in E$, every $h \in R(u_0)$, and every morphism $g \in \pi_1(G^{\diamond}(\tau), E)(x,y)$, we have $ghg^{-1} \in R(v_0)$. Thus if we let $K(u)$ be the subgroup of $\pi_1(\Sigma \setminus \mathbb{P}, E)(u,u)$ generated by $R(u)$, then $K := (K(u))_{u\in E}$ is a normal multilocular subgroup of $\pi_1(G^{\diamond}(\tau), E)$.

Definition 6.2 The quotient groupoid

$$\pi^{\mathrm{orb}}_{1,\mathbb{P}}(G^{\diamond}(\tau), E) := \pi_1(G^{\diamond}(\tau), E)/K$$

will be called the 2-*orbifold fundamental groupoid* of the graph $G^{\diamond}(\tau)$.

The proof of the next result follows from the definitions.

Theorem 6.3 *Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a surface with nonempty boundary, and let τ be an ideal triangulation of $(\Sigma, \mathbb{M}, \mathbb{P})$ of signature zero. The strong deformation retraction $\rho\colon \Sigma \setminus \mathbb{P} \to G^{\diamond}(\tau)$ from Theorem 3.6 induces a commutative diagram of groupoids*

$$\begin{array}{ccc} \pi_1(\Sigma \setminus \mathbb{P}, E) & \xrightarrow[\cong]{\rho_\#} & \pi_1(G^{\diamond}(\tau), E) \\ \Big\downarrow \mathfrak{p} & & \Big\downarrow \mathfrak{p} \\ \pi^{\mathrm{orb}}_{1,\mathbb{P}}(\Sigma, E) & \xrightarrow[\bar{\rho}_\#]{\cong} & \pi^{\mathrm{orb}}_{1,\mathbb{P}}(G^{\diamond}(\tau), E) \end{array}$$

whose horizontal arrows are isomorphisms.

7 Main result: uniqueness of kink-free representative curves

Let $(\Sigma, \mathbb{M}, \mathbb{P})$ and E be as in Section 6. For each $u \in E$, define $L(u)$ to be the image, under the projection $\pi_1(\Sigma \setminus \mathbb{P})(u,u) \to \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u,u)$, of the set of homotopy classes of the curves $\gamma_{u,p,c}$ described in the second paragraph of Section 6. Since $\pi_1(\Sigma \setminus \mathbb{P})(u,u)$ is isomorphic to a torsion-free Fuchsian group having a finite Dirichlet polygon of finite hyperbolic area, we can deduce from eg [15, Theorem 3.5.2 and Corollary 4.2.6; 22, Proposition VI.1.4] that $L(u) = \{f \in \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u,u) \mid f \neq \mathbb{1} \text{ and has finite order}\}$, and thus $L(u) = \{f \in \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u,u) \mid \mathbb{1} \neq f = f^{-1}\}$.

Theorem 7.1 *Let $(\Sigma, \mathbb{M}, \mathbb{P})$ be a surface with nonempty boundary.*

- *Given $u_0, v_0 \in E$, there exists exactly one function*

$$\iota\colon \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u_0,v_0) \setminus L(u_0) \to \pi_1(\Sigma \setminus \mathbb{P}, E)(u_0,v_0)$$

  such that

  (1) $\mathfrak{p} \circ \iota = \mathbb{1}$,

  (2) *for every $f \in \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u_0,v_0) \setminus L(u_0)$ there exists a representative curve $\gamma \in \iota(f)$ that has no kinks.*

- *There is exactly one function*

$$\iota\colon \pi_{1,\mathbb{P}}^{\mathrm{orb,free}}(\Sigma, E)/\!\sim \to \pi_1^{\mathrm{free}}(\Sigma \setminus \mathbb{P}, E)/\!\sim$$

  such that

  (1) $\mathfrak{p} \circ \iota = \mathbb{1}$,

  (2) *for every $f \in \pi_{1,\mathbb{P}}^{\mathrm{orb,free}}(\Sigma, E)/\!\sim$ there exists a representative curve $\gamma \in \iota(f)$ that has no kinks,*

  where $\sim$ is the equivalence relation that identifies each closed curve with its opposite orientation.

Proof We prove the first statement; the proof of the second statement is similar. By Proposition 5.4 and Theorem 6.3, it is enough to take a triangulation τ of signature zero, and prove the statements for $\pi_{1,\mathbb{P}}^{\mathrm{orb}}(G^\circlearrowright(\tau), E)$ and $\pi_1(G^\circlearrowright(\tau), E)$ instead of $\pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)$ and $\pi_1(\Sigma \setminus \mathbb{P}, E)$.

Take $f \in \pi_1(G^\circlearrowright(\tau), E)(u_0,v_0) \setminus \bar{\rho}_\#(L(u_0))$, where $\bar{\rho}_\#$ is the morphism of groupoids in Theorem 6.3. Assume f is written in standard form. It is clear that applying resolutions of kinks to f does not affect the class of f in $\pi_{1,\mathbb{P}}^{\mathrm{orb}}(G^\circlearrowright(\tau), E)(u_0,v_0)$. Furthermore, by Theorem 4.11, the element $\iota(f) \in \pi_1(G^\circlearrowright(\tau), E)(u_0,v_0)$ obtained after fully resolving all kinks from f is independent of the order in which the kinks are resolved. This produces one function

$$\iota\colon \pi_{1,\mathbb{P}}^{\mathrm{orb}}(\Sigma, E)(u_0,v_0) \setminus L(u_0) \to \pi_1(\Sigma \setminus \mathbb{P}, E)(u_0,v_0)$$

with the two stated properties.

Uniqueness follows also from Theorem 4.11 after realizing that, by Definition 6.2, the paragraph preceding it, and Lemma 4.8, if $g \in \pi_1(G^\circlearrowright(\tau), E)(u_0,v_0)$ is a morphism without kinks and such that $\mathfrak{p}(g) = \mathfrak{p}(f)$, then it is possible to obtain g from f by applying a finite sequence of resolutions of kinks. □

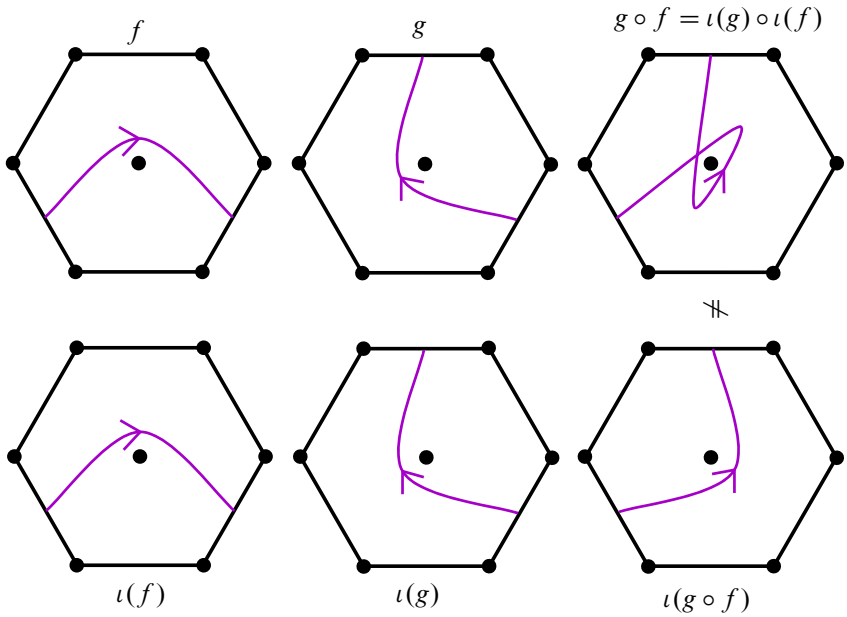


Figure 19

Remark 7.2 The functions ι in the first assertion of Theorem 7.1 may fail to constitute a morphism of groupoids; see Figure 19.

References

[1] **C Amiot**, *The derived category of surface algebras: the case of the torus with one boundary component*, Algebr. Represent. Theory 19 (2016) 1059–1080 MR Zbl

[2] **C Amiot**, **T Brüstle**, *Derived equivalences between skew-gentle algebras using orbifolds*, Doc. Math. 27 (2022) 933–982 MR Zbl

[3] **C Amiot**, **Y Grimeland**, *Derived invariants for surface algebras*, J. Pure Appl. Algebra 220 (2016) 3133–3155 MR Zbl

[4] **C Amiot**, **D Labardini-Fragoso**, **P-G Plamondon**, *Derived invariants for surface cut algebras of global dimension 2, II: The punctured case*, Comm. Algebra 49 (2021) 114–150 MR Zbl

[5] **C Amiot**, **P-G Plamondon**, *The cluster category of a surface with punctures via group actions*, Adv. Math. 389 (2021) art. id. 107884 MR Zbl

[6] **I Assem**, **T Brüstle**, **G Charbonneau-Jodoin**, **P-G Plamondon**, *Gentle algebras arising from surface triangulations*, Algebra Number Theory 4 (2010) 201–229 MR Zbl

[7] **M Chas**, **S Gadgil**, *The extended Goldman bracket determines intersection numbers for surfaces and orbifolds*, Algebr. Geom. Topol. 16 (2016) 2813–2838 MR Zbl

[8] **V V Fock**, **A B Goncharov**, *Dual Teichmüller and lamination spaces*, from "Handbook of Teichmüller theory, I", IRMA Lect. Math. Theor. Phys. 11, Eur. Math. Soc., Zürich (2007) 647–684 MR Zbl

[9] **S Fomin**, **M Shapiro**, **D Thurston**, *Cluster algebras and triangulated surfaces, I: Cluster complexes*, Acta Math. 201 (2008) 83–146 MR Zbl

[10] **C Geiss**, *On homomorphisms and generically τ-reduced components for skewed-gentle algebras*, preprint (2023) arXiv 2307.10306

[11] **F Haiden**, **L Katzarkov**, **M Kontsevich**, *Flat surfaces and stability structures*, Publ. Math. Inst. Hautes Études Sci. 126 (2017) 247–318 MR Zbl

[12] **U Hansper**, *Classification of the indecomposable finite dimensional modules of clannish algebras*, PhD thesis, Universität Bielefeld (2022) Available at https://doi.org/10.4119/unibi/2962952

[13] **J Hass**, **P Scott**, *Intersections of curves on surfaces*, Israel J. Math. 51 (1985) 90–120 MR Zbl

[14] **G Huet**, *Confluent reductions: abstract properties and applications to term rewriting systems*, J. Assoc. Comput. Mach. 27 (1980) 797–821 MR Zbl

[15] **S Katok**, *Fuchsian groups*, Univ. Chicago Press (1992) MR Zbl

[16] **D Labardini-Fragoso**, *Quivers with potentials associated to triangulated surfaces*, Proc. Lond. Math. Soc. 98 (2009) 797–839 MR Zbl

[17] **D Labardini-Fragoso**, **S Schroll**, **Y Valdivieso**, *Derived categories of skew-gentle algebras and orbifolds*, Glasg. Math. J. 64 (2022) 649–674 MR Zbl

[18] **Y Lekili**, **A Polishchuk**, *Derived equivalences of gentle algebras via Fukaya categories*, Math. Ann. 376 (2020) 187–225 MR Zbl

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

[20] **M H A Newman**, *On theories with a combinatorial definition of "equivalence"*, Ann. of Math. 43 (1942) 223–243 MR Zbl

[21] **S Opper**, **P-G Plamondon**, **S Schroll**, *A geometric model for the derived category of gentle algebras*, preprint (2018) arXiv 1801.09659

[22] **H P de Saint-Gervais**, *Uniformization of Riemann surfaces: revisiting a hundred-year-old theorem*, Eur. Math. Soc., Zürich (2016) MR Zbl

[23] **M V Sapir**, *Combinatorial algebra: syntax and semantics*, Springer (2014) MR Zbl

CG, D L-F: *Instituto de Matemáticas, Universidad Nacional Autónoma de México*
Mexico City, Mexico

Current address for D L-F: *Dipartimento di Matematica "Tullio Levi-Civita", Università degli Studi di Padova*
Padova, Italy

christof.geiss@im.unam.mx, daniel.labardini_fragoso@math.unipd.it

https://www.matem.unam.mx/~christof, https://www.math.unipd.it/~labardin

Received: 12 September 2023 Revised: 9 April 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3707–3736
DOI: 10.2140/agt.2025.25.3707
Published: 1 October 2025

Weinstein presentations for high-dimensional antisurgery

IPSITA DATTA
OLEG LAZAREV
CHINDU MOHANAKUMAR
ANGELA WU

We give an algorithm for describing the Weinstein presentation of Weinstein subdomains obtained by carving out regular Lagrangians. Our work generalizes previous work in dimension 3 and requires a novel Legendrian isotopy move (the "boat move") that changes the local index of Reeb chords in a front projection. As applications, we describe presentations for certain exotic Weinstein subdomains and give explicit descriptions of P-loose Legendrians.

53D05, 57R17

1 Introduction

Weinstein domains [26] are exact symplectic manifolds equipped with symplectic handlebody decompositions, analogous to CW complexes in topology. These domains are relatively easy to construct by consecutively attaching handles along isotropic spheres in contact manifolds. Weinstein presentations or diagrams keep track of these isotropic spheres and their interactions with each other, and make computation of invariants, like the wrapped Fukaya category, tractable via surgery formulas and gluing formulas; see Bourgeois, Ekholm, and Eliashberg [4] and Ganatra, Pardon, and Shende [14].

A wealth of symplectically exotic Weinstein domains can be constructed as *subdomains* of more standard Weinstein domains, obtained by *carving out* Lagrangian disks. For example, Sylvan and the second author [20] showed that if $n \geq 5$, the standard cotangent bundle T^*S^n has infinitely many Weinstein subdomains that are diffeomorphic to T^*S^n but pairwise nonsymplectomorphic. They also constructed P-loose Legendrians as subdomains of the sector T^*D^n and showed that these P-loose Legendrians are smoothly isotopic but not Legendrian isotopic. The contact analog of carving out Lagrangian disks — *contact antisurgery* — is important for the construction of contact structures; for example, any contact structure on S^{2n-1} is obtained by doing a single contact surgery and antisurgery on the standard contact structure $(S^{2n-1}, \xi_{\text{std}})$; see Lazarev [18]. Weinstein subdomains are also attractive from the point of view of categorical invariants; their wrapped Fukaya categories are localizations of the Fukaya category of the ambient domain by the localization formula in [14]. Finally, any Weinstein domain deformation retracts to its singular Lagrangian skeleton; therefore the question of studying Weinstein subdomains of a fixed domain X is precisely the question of finding singular Lagrangian skeleta in X.

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Carving out Lagrangian disks and forming Weinstein subdomains appear naturally in two other geometric problems. First, Weinstein subdomains arise naturally when relating complements of toric divisors and their (partial) smoothings. That is, $X \setminus D$ is a Weinstein subdomain of $X \setminus \widetilde{D}$ for a Weinstein domain X, divisor $D \subset X$, and smoothing $\widetilde{D}$ of D. In 4 dimensions, the Weinstein presentations of such manifolds have been related in this context by Acu, Capovilla-Searle, Gadbled, Marinković, Starkston, and the fourth author [2]. They define a necessary condition on a Delzant polytope of the toric manifold, which ensures that the complement of a corresponding partial smoothing of the toric divisor supports a Weinstein structure. They give an algorithm to construct an explicit Weinstein presentation for the complement of such a partially smoothed toric divisor.

Second, carving out Lagrangian disks plays a role in constructing convex hypersurfaces in contact manifolds. See Honda and Huang [17] and Sackel [24] for background. Briefly, a Weinstein convex hypersurface $H^{2n} \subset Y^{2n+1}$ is the union of two Weinstein hypersurfaces H_1^{2n} and H_2^{2n} glued along a contactomorphism of the contact boundaries ∂H_1^{2n} and ∂H_2^{2n}. One can modify Weinstein hypersurfaces via contact handle attachment. Index $k \leq n$ handles correspond to usual Weinstein handle attachment to both H_1 and H_2 along some framed isotropic sphere in $\partial H_1 \cong \partial H_2$. However, index-$(n+1)$ contact handles, which play a crucial role in *bypasses*, are attached along a sphere S^n in H^{2n} that is the union of a Lagrangian disk $D_1 \subset H_1^{2n}$ and a Lagrangian disk $D_2 \subset H_2^{2n}$ that have the same Legendrian boundary in $\partial H_1 \cong \partial H_2$. After the index-$(n+1)$ handle is attached, the resulting Weinstein hypersurface is the union of the *subdomains* $H_1 \setminus D_1$ and $H_2 \setminus D_2$, the result of carving out D_1 and D_2 from H_1 and H_2, respectively. Hence, the problem of explicitly describing convex hypersurfaces as unions of Weinstein hypersurfaces is closely related to the problem of explicitly describing Weinstein subdomains.

On the other hand, Weinstein presentations have not been described yet for the constructions in [20], nor does there exist a general procedure for describing explicit Weinstein presentations for general Weinstein subdomains. For instance, the construction of P-flexible Weinstein manifolds was relatively inexplicit due to the fact that it was not clear how the carving out/antisurgery modified the front of the original Legendrian. In particular, the front projection of these P-loose Legendrians was not known.

Our goal here is to remedy this situation. We focus on the problem of constructing explicit Weinstein presentations of exotic Weinstein subdomains constructed by carving out Lagrangian disks. We want the presentation of such a subdomain to be in terms of a Weinstein presentation of the original Weinstein domain, which is compatible with the Lagrangian disk. In the process, we introduce a new Legendrian isotopy move, called the boat move.

As a concrete application, we give an explicit front projection for the P-loose Legendrians constructed indirectly in [20]. For any collection of integers P, a Legendrian Λ is said to be *P-loose* if it is isotopic to $\Lambda \# \Lambda_P$, the connected sum of Λ and Λ_P, where Λ_P is a P-loose Legendrian unknot, defined in $\mathbb{R}^{2n+1}$ for $n \geq 4$. This operation of taking connect sum with Λ_P can be used to make P-loose Legendrian representatives of any smooth n-dimensional knot type. A Weinstein manifold constructed via handle attachments along P-loose Legendrians is called *P-flexible*. In [20], it was shown that P-loose

Legendrians have properties that generalize those of loose Legendrians, which were introduced by Murphy in [21]. If $0 \in P$, then Λ_P is a loose Legendrian unknot. In general, Λ_P is not necessarily loose but has Legendrian dga (with loop space coefficients) equal to $\mathbb{Z}[1/P]$; see Section 4. Furthermore, for any Legendrian $\Lambda \subset Y$, the Chekanov–Eliashberg dga satisfies

$$\mathrm{CE}(\Lambda \# \Lambda_P) \cong \mathrm{CE}(\Lambda)[P^{-1}],$$

and hence vanishes with $\mathbb{Z}/P\mathbb{Z}$ coefficients.

Main results

For any Weinstein subdomain X_0 of a domain X, the complement $X \setminus X_0$ admits the structure of a Weinstein cobordism C. This cobordism has a subcritical part C_{sub} that does not change invariants like the Fukaya category, and some critical handles H_i^n for $i = 1, \dots, l$, with Lagrangian cocore disks $L_i^n \subset X$. Hence X_0 can be described, up to subcritical cobordism, as $X \setminus \left(\bigcup_{i=1}^{l} L_i\right)$.

Conversely, given any Lagrangian disk $L \subset X$, $X \setminus L$ is an exact subdomain of X. The contact boundary $\partial(X \setminus L)$ of $X \setminus L$ is obtained from the contact boundary ∂X of X by doing *antisurgery*, or $(+1)$-contact surgery, along the Legendrian sphere $\partial L \subset \partial X$. To ensure that $X \setminus L$ is Weinstein, we assume that L is *regular* (see Eliashberg, Ganatra, and Lazarev [13]), which implies that our starting Weinstein presentation for X is compatible with L, as we describe later.

Our main goal is to give explicit constructions (Weinstein presentations and handlebody decompositions) for Weinstein manifolds obtained via antisurgery. To do this we introduce a new family of n-dimensional Legendrian moves in the front projection. We construct Legendrian isotopies called D^k*-suspensions* that consist of a Legendrian isotopy ψ of an $(n-k)$-dimensional slice suspended over a k-dimensional disk.

Proposition 1.1 *Given a Legendrian isotopy $\psi : D^{n-k} \times [0,1]_t \to \mathbb{R}^{2(n-k)+1}$ which is the identity near ∂D^{n-k} and t-independent near $\partial[0,1]$, its D^k-suspension $\Sigma_{D^k}\{\psi\}$ is a Legendrian in $\mathbb{R}^{2n+1}$, and is Legendrian isotopic to $D^k \times D^{n-k}$ relative to the boundary.*

We call the D^k-suspension of an $(n-k)$-dimensional Reidemeister 1 move an (n,k)-*boat move*; see for instance Figure 9. Using these boat moves, we can construct our desired presentations for Weinstein domains obtained via antisurgery.

In dimension 3, there are several existing results, for example in work of Ding and Geiges [10], explaining how to do antisurgery along Legendrian circles that admit Lagrangian disk fillings. Our main contribution is that we are able to replicate such explicit constructions in higher dimensions by first using (n,k)-boat moves. We do this assuming we know a Weinstein presentation for the original manifold.

Next we state our main result. Let $L \subset X$ be a regular Lagrangian disk in a Weinstein domain X^{2n}. By [13], any regular Lagrangian disk $L \subset X$ in a Weinstein domain X can be presented as

$$D^n \subset T^*D^n \cup \left(\bigcup_i H_i\right),$$

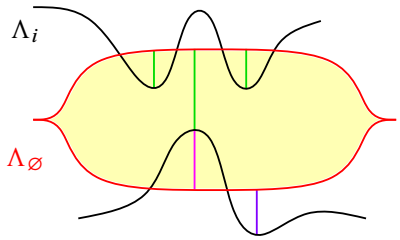


Figure 1: The green Reeb chords are bounded by the flying saucer and go from the black Λ_i to the red $\Lambda_\varnothing$; these have critical points on Λ_i with local indices 1, 0, and 1 for the height difference Morse function from Λ_i to $\Lambda_\varnothing$. We will apply $(1,0)$, $(1,1)$, and $(1,0)$ boat moves at these critical points, respectively, which correspond to doing a Reidemeister 1 move at the two index-1 critical points and nothing at the middle index-0 critical point. The purple Reeb chord is not bounded by the flying saucer. The pink Reeb chord is bounded by the flying saucer but does not go from Λ_i to $\Lambda_\varnothing$.

where H_i are Weinstein handles attached to T^*D^n in the complement of $\partial D^n \subset T^*D^n$. For simplicity, we assume here that all of these handles have index n, although this assumption can be removed. Let

$$\Lambda_\varnothing \cup \left(\bigcup_i \Lambda_i\right) \subset \mathbb{R}^{2n-1}$$

be the Legendrian link formed by the following Legendrians. First, $\Lambda_\varnothing = \partial D^n$ is the standard Legendrian unknot with front projection in $\mathbb{R}^n$ given by the "flying saucer" (with a S^{n-2}-family of cusps and no other singularities). Second, Λ_i are the attaching Legendrians of H_i. Let C_i denote the set of Reeb chords from Λ_i to $\Lambda_\varnothing$ with front projection contained in the subset of $\mathbb{R}^n$ bounded by the flying saucer front projection $\pi(\Lambda_\varnothing) \subset \mathbb{R}^n$, and let $C = \bigcup_i C_i$; see Figure 1. We assume all such chords are nondegenerate, and furthermore correspond to critical points of a local Morse function whose indices we call the local index of the Reeb chord; in particular, C is finite.

Theorem 1.2 *There is a Weinstein presentation for the Weinstein subdomain $X \setminus L \subset X$ with the following properties*:

- *The Weinstein presentation of $X \setminus L$ has one more $(n-1)$-handle than the Weinstein presentation for X.*
- *The n-handles for $X \setminus L$ are in one-to-one correspondence with the n-handles of X.*
- *The attaching sphere Λ_i' of the n-handle H_i' of $X \setminus L$ is obtained from the attaching sphere Λ_i of the corresponding handle H_i of X in the following way: for each Reeb chord γ in C_i with local index k, we apply an $(n, n-k)$-boat move to Λ_i and do a cusp connected sum with a Legendrian that goes through the new $(n-1)$-handle one time.*

Remark 1.3 The assumptions for Theorem 1.2 can be weakened. For example, as explained in Lemma 4.1, any Legendrian link in $\mathbb{R}^{2n+1}$ can be perturbed by a C^∞-small isotopy so that all Reeb chords in C are nondegenerate and correspond to critical points of a Morse function; however, perturbing a degenerate Reeb chord may result in a larger (finite) number of nondegenerate Reeb chords.

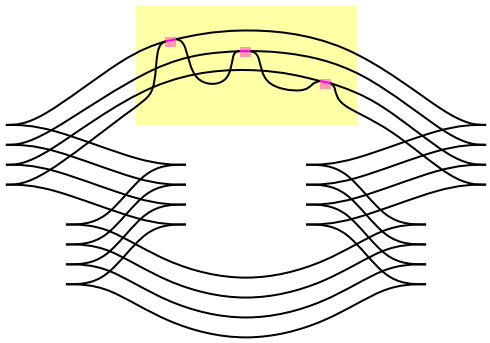


Figure 2: A low-dimensional slice of a P-loose Legendrian which cuts through the three boat move cusp connect sums, represented by the square regions. Away from a neighborhood of this slice, the black Legendrians may be linked in the shaded region but are parallel outside of it.

Also, the assumption that X takes the form $T^*D^n \cup \left(\bigcup_i H_i^n\right)$ can be generalized to allow subcritical handles. Generically, there will be no Reeb chords between $\Lambda_\varnothing$ and the attaching spheres of these subcritical handles, but the attaching spheres of the critical handles may interact with these subcritical handles. Hence, only a portion of these attaching spheres map to the contact boundary $\partial_\infty T^*D^n$, resulting in a Legendrian subset $\Lambda_i \subset \partial_\infty T^*D^n$ (which are not necessarily spheres but manifolds with boundary). In that case, a generalization of Theorem 1.2 holds by considering Reeb chords between $\Lambda_\varnothing$ and the subset Λ_i.

Additionally, as a main application of Theorem 1.2, we construct explicitly the P-loose Legendrians in the Weinstein presentations of P-flexible Weinstein domains.

Corollary 1.4 *For integers $p \geq 0$ and $n \geq 2$, the P-loose Legendrian unknot, $\Lambda_P \subset \mathbb{R}^{2n+1}$ for $P = \{p\}$, is a Legendrian submanifold consisting of four loose Legendrian unknots which are completely parallel away from a bounded region disjoint from the loose charts. Within this bounded region they may be linked (in a way that depends on p) and are connected via three boat moves and cusp connected sum gluings* (see Figure 2); *the front projection in this bounded region has no singularities except for the singularities in the local boat moves.*

Structure

In Section 2, we give some background on antisurgery and P-loose Legendrians. In Section 3, we discuss the boat move and prove Proposition 1.1. In Section 4, we present our construction for producing Weinstein presentations for subdomains and prove Theorem 1.2 and Corollary 1.4. We then use our construction to give explicit examples of Weinstein antisurgery manifolds and conclude with some open questions.

Acknowledgements

The authors would like to thank the organizers (Orsola Capovilla-Searle, Roger Casals, and Caitlin Leverson) of the SYNC Early Career Workshop at the University of California, Davis, where this project

was started in August, 2022. The authors would also like to thank Hiro Lee Tanaka, Lea Kenigsberg, Tonie Scroggin, Georgios Dimitroglou Rizell, Jonathan Michala, Josh Sabloff, and Daniel Álvarez-Gavela for helpful discussions. Datta was supported by NSF grant DMS-1926686. Lazarev was supported by NSF grant DMS-2305392. Mohanakumar was supported by NSF grant DMS-2003404. Wu was supported by NSF grant DMS-2238131. During the research in this paper, Datta was at the IAS, Princeton.

2 Background

2.1 Contact surgery and Weinstein manifolds

Let M be an n-dimensional manifold. In the topological setting, we perform k-surgery on M, for $k \leq n$, by first removing a tubular neighborhood $N(S^k) \cong S^k \times D^{n-k}$ of a k-sphere, and then gluing back a copy of $D^{k+1} \times S^{n-k-1}$ along boundary $S^{n-k} \times S^{n-k-1}$. A choice of framing, that is, the particular identification $\varphi\colon N(S^k) \cong S^k \times D^{n-k}$, specifies our gluing, which then determines our surgered manifold up to diffeomorphism.

When the k-sphere is a Legendrian sphere, Λ, inside M, a contact manifold, there is a canonical framing, providing an identification $\Lambda \cong S^k \subset \mathbb{R}^{k+1}$. To see this, we recall some facts and notation from standard contact geometry. The contact hyperplane bundle supports a symplectic form given by $d\alpha|_\xi$. We may additionally endow this bundle with an almost-complex structure J that is compatible with $d\alpha$. Then, along λ, the tangent bundle TM can be decomposed as $TM|_\Lambda = T\Lambda \oplus \langle R_\alpha \rangle \oplus J(T\Lambda)$, and so $N(\Lambda) \cong \langle R_\alpha \rangle \oplus J(T\Lambda)$ for a contact form α and corresponding Reeb vector field R_α. The bundle $\langle R_\alpha \rangle \oplus J(T\Lambda)$ can be identified with the stabilized tangent bundle of Λ, which carries a canonical trivialization after an identification of Λ with $S^k \subset \mathbb{R}^{k+1}$ — this trivialization is the canonical framing. At this point, we drop the dimension k of our sphere from the notation, and refer to such a surgery as $(\pm m/k)$-surgery, where $\pm m/k$ is a fraction relating our choice of framing to the canonical framing. For example, attaching a critical Weinstein handle affects the contact boundary by a (-1)-surgery.

2.2 Legendrian moves

Legendrian moves refer to the replacement of a Legendrian by something Legendrian isotopic that differs from it only within a Darboux neighborhood. These are often depicted via front diagrams.

Legendrian Reidemeister moves are analogous to knot diagram moves which preserve the topological knot type. They usually refer to replacements in the front diagrams of Legendrian knots in a contact 3-manifold depicted in Figure 3. These fully characterize Legendrian isotopies for 1-dimensional Legendrians in contact 3-manifolds.

Theorem 2.1 [25] *Two front diagrams represent Legendrian isotopic Legendrian knots if and only if they are related by regular homotopy and a sequence of moves shown in Figure 3.*

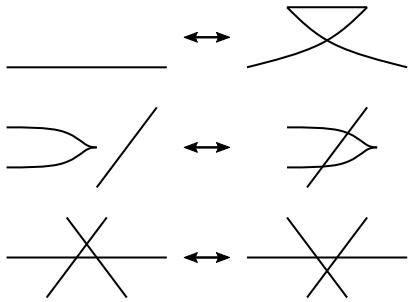


Figure 3: The Legendrian Reidemeister moves.

We will also be using higher-dimensional first Reidemeister moves. For any Legendrian submanifold $\Lambda \subset (Y, \xi)$, the Legendrian submanifold Λ' obtained by replacing a graphical portion of the front of Λ (with respect to a particular Darboux chart) by the rightmost front depicted in Figure 4 is Legendrian isotopic to Λ. To see this, note that the first arrow is the same as one of the higher-dimensional first Reidemeister moves described in [5]. The subsequent two arrows of Figure 4 are front diagrams of Legendrian isotopies where the domed part of the front is pushed inwards via an isotopy of the $(x_1, \dots, x_n, z)$-plane that does not have vertical tangencies at all times. We refer to this replacement within a Darboux chart as a *k-dimensional first Reidemeister move*.[1] Let

$$R1_k : [-1, 1] \times D^k \to \mathbb{R}^{2k+1}$$

denote the associated Legendrian isotopy. Note that the fronts depicted in Figure 4 have S^{k-2}-symmetry about the z-axis passing through the point of singularity.

Remark 2.2 The fronts depicted in Figure 4, and therefore fronts obtained whenever we apply the n-dimensional R1, are not generic [8] as there is a singularity at the cone point. This does not matter for our purposes, but is important to keep in mind especially if computing Legendrian contact homology.

Remark 2.3 Higher-dimensional versions of the second and third Legendrian Reidemeister moves also exist; see for instance their depictions in [6, Section 2.4]. In higher dimensions, these three moves in the front do not fully characterize all Legendrian isotopies. Generally, we can isotope Legendrian fronts past each other by the Reeb flow, as long as there are no Reeb chords between them.

In addition to Reidemeister moves, we will use Legendrian handleslides and the addition or removal of a canceling pair of Weinstein handles in our diagrammatic calculus. These are moves that allow us to pass between different Weinstein handle diagrams of equivalent Weinstein domains. In other words, before and after these moves, the Legendrians depicted are isotopic in the surgered contact boundaries (but maybe not isotopic in the original contact boundaries).

[1] Replacing the figures in Figure 4 by their reflections about the $(x_1, \dots, x_n)$-plane gives analogous Legendrian moves. We omit these in the discussion for simplicity.

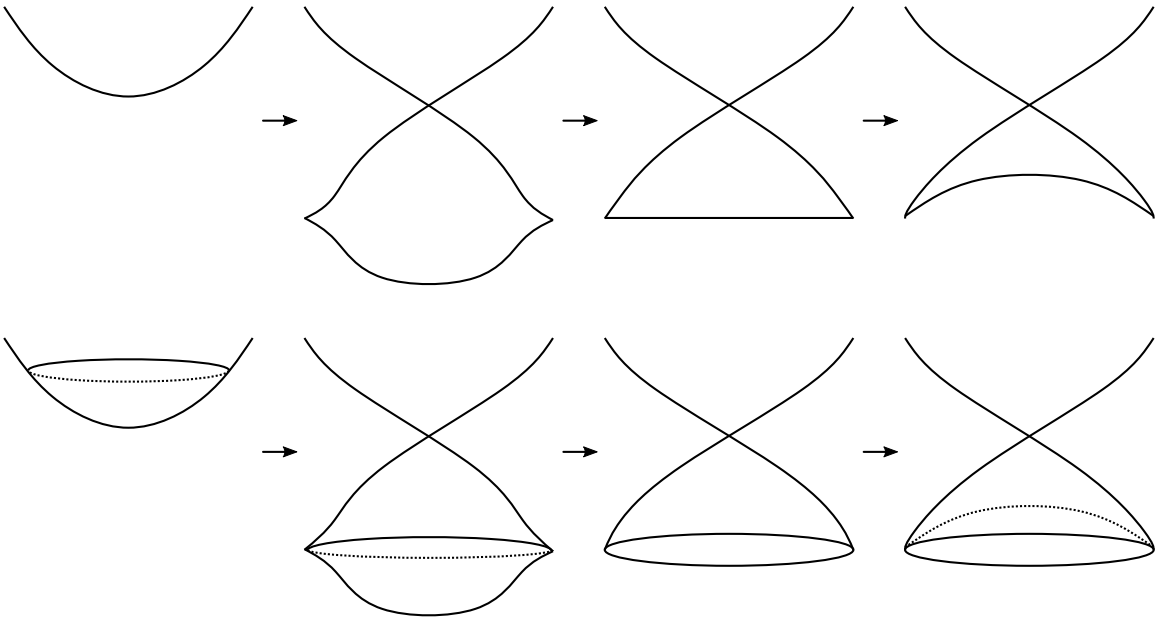


Figure 4: A Legendrian Reidemeister 1 and isotopy for a knot (top) and a surface (bottom).

Following Casals and Murphy [6], we depict the effect of Legendrian handleslides in Figure 5. A handleslide over a $(+1)$-surgery curve produces a cone singularity, while sliding over a (-1)-surgery produces a circle of cusp singularities. Meanwhile, for a Weinstein manifold of dimension $2n$, the addition or removal of a canceling pair refers to adding in an n-handle and an $(n-1)$-handle such that the attaching sphere of the n-handle and the belt sphere of the $(n-1)$-handle intersect exactly once.

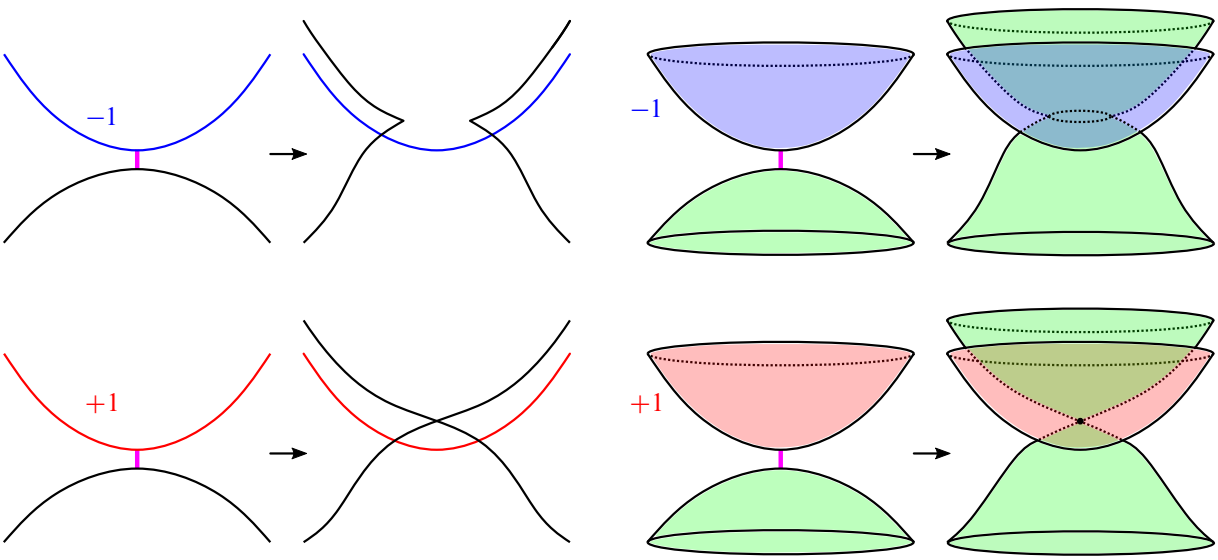


Figure 5: Handleslides over (-1) and $(+1)$ Legendrians (in blue and red, respectively) in 3 and 5 dimensions.

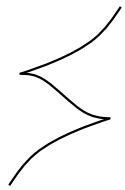


Figure 6: A 1-dimensional stabilized Legendrian arc.

2.3 Loose and P-loose Legendrians, flexible and P-flexible domains

In [21], Murphy introduced a class of Legendrians called loose Legendrians. These Legendrians are characterized by an explicit local model.

Definition 2.4 A loose unknot $\Lambda_l \subset (\mathbb{R}^{2n+1}, \xi_{\text{std}})$ for $n \geq 2$ is a Legendrian that is formally isotopic to the Legendrian unknot, and where there is an $\mathbb{R}^3$ slice that intersects Λ_l transversely and the front projection of $\Lambda_l \cap \mathbb{R}^3$ is the 1-dimensional stabilized arc (see Figure 6). A Legendrian $\Lambda \subset (Y, \xi)$ is loose if it is Legendrian isotopic to $\Lambda \# \Lambda_l \subset Y \# \mathbb{R}^{2n+1} \cong Y$.

Remark 2.5 Since we prefer to work with closed Legendrians, the above definition involves loose Legendrian unknots (which are spheres) instead of the original definition, which involves loose Legendrian *charts*, which are Legendrian disks; these definitions are equivalent.

We emphasize that there is not a canonical model for the loose Legendrian unknot (except in dimension 1, which must be excluded for reasons explained below). For example, we can take any codimension-zero subdomain U near the cusp of the Legendrian unknot $\Lambda_\varnothing$ and push through U past the cusp to create Λ_U (see Figure 7). Then Λ_U is always loose and if U has Euler characteristic zero, then Λ_U is Legendrian isotopic to the loose Legendrian unknot. Alternatively, one can take any *closed* codimension-1 submanifold V of $\Lambda_\varnothing$ and add a 1-dimensional zigzag spun by V to create Λ_V. Again if V has Euler characteristic zero, Λ_V is a Legendrian unknot. Another way to see the nonexistence of a canonical model is to observe that there is no canonical way to extend the 1-dimensional stabilized arc to a higher-dimensional Legendrian disk so that the extension is standard near the boundary.

This indicates a proliferation of loose Legendrians unknots, making it impossible to speak of *the* loose Legendrian unknot. However, the main result proven in [21] about loose Legendrians is that they satisfy an h-principle, which essentially says that, if loose Legendrians Λ_1 and Λ_2 in (Y, ξ) are smoothly isotopic (and satisfy additional tangential data), then they are Legendrian isotopic. This means that the

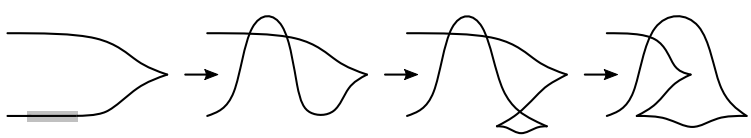


Figure 7: Near the cusp of the Legendrian, we push through the highlighted subdomain, which perturbs the knot into becoming loose.

symplectic geometry of loose Legendrians reduces just to classical differential topology. In particular, all loose Legendrian unknots are Legendrian isotopic. For $n = 1$, the local model still makes sense but the h-principle does not hold, and therefore we speak of loose Legendrians only for $n \geq 2$.

Using loose Legendrians, Cieliebak and Eliashberg defined the class of flexible Weinstein domains [7]:

Definition 2.6 A Weinstein domain is *flexible* if the attaching spheres for its half-dimensional handles are loose Legendrians.

More generally, we can consider flexible Weinstein sectors. A sector is equivalently a Weinstein domain with the extra data of a Weinstein hypersurface in its contact boundary. A flexible sector is one for which the ambient domain is flexible and the Weinstein hypersurface is loose (in the sense that all core disks of its critical handles are loose Legendrians). We point out that, in work of Murphy and Siegel [22], these domains are called *explicitly* flexible and Weinstein domains are called flexible if they admit a Weinstein homotopy to an explicitly flexible one; this notion of flexibility is tautologically preserved under Weinstein homotopy.

2.4 Antisurgery construction of P-loose Legendrian unknots

In this section we review the construction of P-loose Legendrian unknots via antisurgery in [20]. There, the authors considered Lagrangians disks $D_P \subset T^*D^n$ (described below) with Legendrian boundary $\partial D_P \subset \partial_\infty T^*D^n$ disjoint from the boundary of the zero-section ∂D^n. Then they carved out these disks to obtain subdomains $T^*D^n \setminus D_P$ of T^*D^n.

To construct the P-loose Legendrian unknot Λ_P, we view $T^*D^n \setminus D_P$ as a Weinstein sector, or equivalently as a Weinstein domain $B^{2n} \setminus D_P$ with a Legendrian stop in its contact boundary. This can be done as follows. As a Weinstein domain, $B^{2n} \setminus D_P$ is just $B^{2n} \cup H^{n-1}$. Since ∂D^n and ∂D_P are disjoint, ∂D^n can be viewed as a Legendrian in the carved-out domain $B^{2n} \cup H^{n-1} = B^{2n} \setminus D_P$. Next, we recover B^{2n} by attaching a flexible handle along any Legendrian that is loose in the complement of $\partial D^n \subset \partial(B^{2n} \cup H^{n-1})$. The image of ∂D^n in this new B^{2n} is denoted by Λ_P, the P-loose unknot.

To complete the construction of Λ_P, we need to describe the construction of the Lagrangian disks $D_P \subset T^*D^n$. These disks were originally introduced by Abouzaid and Seidel in [1, Section 3b]. Let $U \subset S^{n-1}$ be a compact codimension-zero submanifold with smooth boundary. Let

$$f : S^{n-1} \to \mathbb{R}$$

be a C^1-small function with zero as a regular value, so that f is strictly negative in the interior of U, zero on ∂U, and strictly positive on $S^{n-1} \setminus \overline{U}$. Next, we consider S^{n-1} as the $\frac{1}{2}$-radius sphere $S^{n-1}_{1/2} \subset D^n$ and extend f to a smooth Morse function (again called)

$$f : D^n \to \mathbb{R},$$

so that f is C^0-small in the $\frac{1}{2}$-radius disk and satisfies

$$f(tq) = |t|^2 f(q) \quad \text{for } q \in S^{n-1}_{1/2} \text{ and } 1 \le t \le 2.$$

Let $\Gamma(df)$ be the graph of df in T^*D^n and let

$$D_U = \Gamma(df) \cap B^{2n}.$$

Since f is homogeneous for $|q| \ge \frac{1}{2}$ and 0 is a regular value of g, D_U has Legendrian boundary (with respect to the standard radial Liouville vector field on B^{2n}) that is disjoint from ∂D^n. Furthermore, there is a Lagrangian isotopy $\Gamma(d(sf))_{s\in[0,1]}$ from the zero-section $D^n \subset T^*D^n$ to D_U (that intersects the stop ∂D^n precisely when $s = 0$).

Now let U be a P-Moore space: a CW complex whose reduced singular cohomology is isomorphic to $\mathbb{Z}/P\mathbb{Z}$ in some degree. For example, one can take the mapping cone of the degree-P map

$$\varphi_P \colon S^k \to S^k.$$

Then the resulting Lagrangian disk D_U is called D_P, and is the Lagrangian disk D_P used earlier to construct Λ_P.

Remark 2.7 The above construction of $D_U \subset T^*D^n$ requires a smooth embedding $U \subset S^{n-1}$. If U is a P-Moore space, this is possible only if $n \ge 5$, and hence there exist P-loose Legendrians of dimension at least 4.

The sector $T^*D^n \setminus D_U$ is obtained by doing antisurgery on the boundary of D_U, while keeping track of the original stop ∂D^n of T^*D^n. In particular, the Legendrian boundaries

$$\Lambda_U := \partial D_U \quad \text{and} \quad \Lambda^{n-1} := \partial D^n$$

are linked in a way we now describe. Note that we can view $\Lambda^{n-1} \subset \partial T^*D^n$ as the Legendrian unknot in $\mathbb{R}^{2n-1}$.

Proposition 2.8 *Λ_U is contained in a small neighborhood of ∂D^n, it is given by the 1-jet of $f|_{\partial D^n}$, and its front projection in $\partial D^n \times \mathbb{R}$ is given by the graph of $f|_{\partial D^n}$. In particular, Λ_U can be isotoped so that its front is obtained from ∂D^n by a negative Reeb pushoff on U', a smaller open set $U' \subset U$, and a positive Reeb pushoff in $\partial D^n \setminus U''$ for some larger open set $U'' \supset U$. See Figure 8.*

Proof As stated in the construction of $D_U = \Gamma(df)$ above, to ensure that D_U has Legendrian boundary, we consider T^*D^n as the Liouville domain

$$\left(B^{2n}, \frac{1}{2}\left(\sum_{i=1}^{n} x_i\, dy_i - y_i\, dx_i\right)\right)$$

equipped with a stop ∂D^n, which we identify with the Legendrian unknot. The Liouville vector field associated to the Liouville form $\frac{1}{2}\left(\sum_{i=1}^n x_i\, dy_i - y_i\, dx_i\right)$ is the radial vector field $\frac{1}{2}\sum_{i=1}^n (x_i\partial_{x_i} + y_i\partial_{y_i})$.

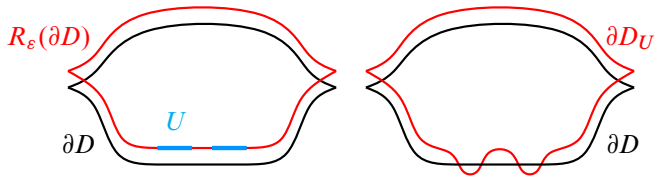


Figure 8: Left: Two parallel Legendrian unknots. The blue region in the top Legendrian is the smooth subdomain U. Right: The Legendrian link $\partial D_U \cup \partial D$, with ∂D_U in red and ∂D in black. This figure is obtained by pushing through U in the top Legendrian down past the bottom Legendrian.

So the fact that f has the form $t^2 f(\theta)$ near $\partial D^n = S^{n-1}$, with coordinate θ, implies that this Liouville vector field is tangent to $\Gamma(df)$ and hence $\Gamma(df)$ has Legendrian boundary. Near ∂D^n,

$$\left(B^{2n}, \frac{1}{2}\left(\sum_{i=1}^{n} x_i\, dy_i - y_i\, dx_i\right)\right)$$

equals

$$(T^*S^{n-1}, \lambda_{T^*S^{n-1}}) \times \left(T^*[1-\varepsilon, 1], \tfrac{1}{2}(t\, dp - p\, dt)\right) = \left(T^*S^{n-1} \times T^*[1-\varepsilon, 1], \lambda_{T^*S^{n-1}} + \tfrac{1}{2}(t\, dp - p\, dt)\right).$$

To see this, note that both of these Liouville structures have the radial vector field as their Liouville vector field, which determines the Liouville form. The (convex) contact boundary of this domain is $T^*S^{n-1} \times T_1^*[1-\varepsilon, 1]$, the restriction to the cotangent fiber at 1. The induced contact form is

$$\lambda_{T^*S^{n-1}} + \tfrac{1}{2}\, dp,$$

given by restricting the Liouville form $\lambda_{T^*S^{n-1}} + \frac{1}{2}(t\, dp - p\, dt)$ to the cotangent fiber at 1. Furthermore, this contact boundary is contactomorphic to

$$(J^1(S^{n-1}) = T^*S^{n-1} \times \mathbb{R}_z, \lambda_{T^*S^{n-1}} + dz)$$

via the map $z = \frac{1}{2}p$.

Near its boundary, our Lagrangian is given by

$$d(t^2 f(\theta)) = t^2 d_\theta f + 2t f(\theta)\, dt,$$

and hence its Legendrian restriction to the contact boundary $T^*S^{n-1} \times T_1^*[1-\varepsilon, 1]$ is $(d_\theta f, 2f(\theta))$. Under the contactomorphism $z = \frac{1}{2}p$, this Legendrian maps to $(d_\theta f, f(\theta))$, the 1-jet of the function $f(\theta)$. Furthermore, we can isotopy f through functions vanishing precisely on ∂U so that f is equal to -1 on U' for a smaller open set $U' \subset U$ and equal to $+1$ on $S^{n-1} \setminus U''$ for a larger open neighborhood $U'' \supset U$, giving us the claimed result. □

One of our goals is to is to do antisurgery on ∂D_U and find a presentation for ∂D^n in the resulting contact manifold.

3 Constructions of higher-dimensional Legendrian isotopy moves

In this section we describe constructions of some higher-dimensional Legendrian isotopies. These isotopies are compactly supported, so we may view these as higher-dimensional Legendrian moves. In the subsequent section we will use these Legendrian moves in the construction of handle diagrams for Weinstein manifolds obtained via antisurgery.

3.1 Suspensions of Legendrian isotopies

Consider a Darboux chart with coordinates $x_1,\dots,x_n,y_1,\dots,y_n,z$ and contact form $\alpha = dz - \sum_{i=1}^n y_i\,dx_i$. Let us use $\mathbb{R}^{2k}$ to refer to the span of $x_1,\dots,x_k,y_1,\dots,y_k$ and $\mathbb{R}^{2(n-k)+1}$ to refer to the span of $x_{k+1},\dots,x_n,y_{k+1},\dots,y_n,z$. Then, with the contact form

$$\alpha_{n-k} := dz - \sum_{i=k+1}^{n} y_i\,dx_i,$$

$(\mathbb{R}^{2(n-k)+1},\alpha_{n-k})$ is a contact manifold. With this notation in mind, we may view

$$\mathbb{R}^{2n+1} = \mathbb{R}^{2k}\times\mathbb{R}^{2(n-k)+1} \cong T^*\mathbb{R}^n\times\mathbb{R}^{2(n-k)+1}.$$

Consider the Legendrian (with boundary) in this Darboux chart given by

$$D^k\times\{0\}\times D^{n-k}\times\{0\}\subset \mathbb{R}^k_{x_1,\dots,x_k}\times\mathbb{R}^k_{y_1,\dots,y_k}\times\mathbb{R}^{n-k}_{x_{k+1},\dots,x_n}\times\mathbb{R}^{n-k+1}_{y_{k+1},\dots,y_n,z}.$$

For ease of notation we write $D^k\times D^{n-k} := D^k\times\{0\}\times D^{n-k}\times\{0\}$. We view $D^k\times D^{n-k}$ as a D^k-parameter family of Legendrian D^{n-k}'s. Let $s\in D^k$ and $\theta\in D^{n-k}$ denote arbitrary elements.

Consider a Legendrian isotopy

$$\psi\colon D^{n-k}\times[0,1]_t\to\mathbb{R}^{2(n-k)+1}.$$

We use the same notation $\psi\colon\mathbb{R}^{2(n-k)+1}\times[0,1]\to\mathbb{R}^{2(n-k)+1}$ to denote a contact isotopy that extends this Legendrian isotopy. Assume that ψ is the identity near ∂D^{n-k} and is t-independent near $\partial[0,1]$, ie

$$\frac{\partial\psi}{\partial t}(\theta,0) = \frac{\partial\psi}{\partial t}(\theta,1) = 0 \quad \text{for all } \theta\in D^{n-k}.$$

Fix a smooth "bump function" $\beta_k\colon D^k\to[0,1]$ on the parameter space such that

- β_k has a unique critical point which is a maximum at 0 with $\beta_k(0)=1$,
- β_k has radial symmetry, that is, $\beta_k(t)=\beta_k(t')$ whenever $|t|=|t'|$, and
- $\beta_k|_{\partial D^k}\equiv 0$.

Definition 3.1 The *D^k-suspension of a Legendrian isotopy*, $\Sigma_{D^k}\{\psi\}$, is the unique Legendrian lift of

$$\{(s,\psi(\theta,\beta_k(s)))\mid s\in D^k,\,\theta\in D^{n-k}\}$$

under the projection

$$\Pi\colon T^*\mathbb{R}^k\times\mathbb{R}^{2(n-k)+1}\to\mathbb{R}^k\times\mathbb{R}^{2(n-k)+1},$$
$$(x_1,\dots,x_k,y_1,\dots,y_k,x_{k+1},\dots,y_n,z)\mapsto(x_1,\dots,x_k,x_{k+1},\dots,y_n,z).$$

The Legendrian condition implies that the momentum coordinates $y_1, \dots, y_k$ of $T^*\mathbb{R}^k$, and thus the Legendrian submanifold itself, can be uniquely recovered from its projection $\Pi(\Sigma_{D^k}\{\psi\})$. The boundary of $D^k \times D^{n-k}$ is equal to $\partial D^k \times D^{n-k} \cup D^k \times \partial D^{n-k}$. By our assumptions on ψ and β_k, and uniqueness of the Legendrian lift,

$$\partial \Sigma_{D^k}\{\psi\} = \partial(D^k \times D^{n-k}).$$

We now prove Proposition 1.1 from the introduction, which states that $\Sigma_{D^k}\{\psi\}$ is Legendrian isotopic to $D^k \times D^{n-k}$ relative to boundary.

Proof of Proposition 1.1 We can define the isotopy by setting, for every $\tau \in [0,1]$, $\varphi_\tau(D^k \times D^{n-k})$ to be the unique Legendrian lift of

$$\{(s, \psi(\theta, \tau\beta_k(s))) \mid s \in D^k, \theta \in D^{n-k}\}$$

under the canonical projection Π. Therefore φ_τ is a Legendrian embedding and φ is a Legendrian isotopy from $D^n \times D^{n-k}$ to $\Sigma_{D^k}(\psi)$.

As $\partial \Sigma_{D^k}\{\psi\} = \partial(D^k \times D^{n-k})$, for every $\tau \in [0,1]$, φ_τ is the identity on the boundary $\partial(D^k \times D^{n-k})$. □

Remark 3.2 The word "suspension" in the context of Legendrians has been used to denote suspensions when the parameter space is S^k. In [9] the suspension $\Sigma_{S^k}\{\Lambda_\theta\}$ is defined similarly to Definition 3.1 for S^k-parametrized families of (closed) Legendrian embeddings. If the S^k-family is taken to be a constant family, then one recovers the S^k-spun of Legendrians, which first appeared in [11] for $k = 1$ and then was extended to general k in [15]. This construction was also studied by Sabloff and Sullivan, who called it "front spinning" in [23].

In this paper, our construction is "local". We obtain a Legendrian with boundary within a Darboux chart. This is in contrast to the previous constructions where the obtained Legendrian is closed, that is, has no boundary.

Another small point of difference from earlier constructions is that instead of beginning with a D^k-parametrized family of Legendrian embeddings, we first create such a family from a 1-parameter Legendrian isotopy.

We may generalize the suspension construction to the case when the initial Legendrian has a graphical front for a nonzero function. That is, instead of $D^k \times D^{n-k}$, consider Legendrian N contained in a Darboux chart so that the front of N in $D^n \times D^{n-k} \times \mathbb{R}$ is given by

$$\pi(N) = \Gamma(f_1 + f_2) = \{(x_1, \dots, x_k, x_{k+1}, \dots, x_n, f_1(x_1, \dots, x_k) + f_2(x_{k+1}, \dots, x_n)\}$$

for two smooth functions

$$f_1 : D^k \to \mathbb{R} \quad \text{and} \quad f_2 : D^{n-k} \to \mathbb{R}.$$

Let Λ_{f_2} be the unique Legendrian lift in $\mathbb{R}^{2(n-k)+1}$ of the front $\Gamma(f_2)$. Again consider a Legendrian isotopy

$$\psi : \Lambda_{f_2} \times [0,1] \to \mathbb{R}^{2(n-k)+1}$$

which is the identity near the boundary of Λ_{f_2} and satisfies

$$\frac{\partial\psi}{\partial t}(\theta,0)=\frac{\partial\psi}{\partial t}(\theta,1)=0 \quad \text{for all } \theta\in\Lambda_{f_2}.$$

Definition 3.3 The $\Gamma(f_1)$-*suspension of a Legendrian isotopy*, $\Sigma_{\Gamma(f_1)}\{\psi\}$, is the unique Legendrian lift of

$$\Pi(\Sigma_{\Gamma(f)}\{\psi\})=\{(s,(0,0,f_1(s))+\psi_{\beta_k(s)}(\theta))\in D^k\times\mathbb{R}^{2(n-k)+1}\mid s\in D^k,\,\theta\in\Lambda_{f_2}\}.$$

Here $s\in D^k$ is in the first k position coordinates, $(0,0,f_1(s))\in\mathbb{R}^{n-k}\times\mathbb{R}^{n-k}\times\mathbb{R}$, and $\psi_{\beta_k(s)}(\theta)\in\mathbb{R}^{2(n-k)+1}$. Analogous to Proposition 1.1, we have the following proposition. We do not include a detailed proof as it would merely be a slight tweak of the proof of Proposition 1.1.

Proposition 3.4 *The $\Gamma(f)$-suspension, $\Sigma_{\Gamma(f)}\{\psi\}$, is Legendrian isotopic to N relative to boundary.*

3.2 The boat move

In this section we introduce a new Legendrian move that can be viewed as a generalization of the Reidemeister 1 move. Introducing a boat is a way of swapping nonmaximum critical points in a graphical front with maxima and some cusp singularities.

Consider a Legendrian $\Lambda\subset(Y,\xi)$ with front equal to the graph of a Morse function, ie the front in some Darboux chart is $\Gamma(f)$ for $f\colon D^n\to\mathbb{R}$ a Morse function. Suppose there exists a minimum. Then the n-dimensional Reidemeister 1 move swaps in a maximum for a minimum, as seen in Figure 4. We want to convert other index critical points to maxima as well.

Definition 3.5 Suppose the front of an open subset $\Lambda_0\subset\Lambda$ locally is given as the graph of a Morse function $f\colon D^n\to\mathbb{R}$ with an index-k critical point at $\mathbf{0}$. By the Morse lemma, there exist local coordinates $x_1,\dots,x_n$ such that

$$f(x_1,\dots,x_n)=-x_1^2-\cdots-x_k^2+x_{k+1}^2+\cdots+x_n^2.$$

Locally we can extend $(x_1,\dots,x_n)$ to Darboux coordinates $(x_1,\dots,x_n,y_1,\dots,y_n,z)$, as any diffeomorphism of $\mathbb{R}^n$ can be extended to a contactomorphism of the 1-jet space, $J^1(\mathbb{R}^n)$. If we set

$$f_1\colon\mathbb{R}^k\to\mathbb{R},\qquad f_1(x_1,\dots,x_k)=-x_1^2-\cdots-x_k^2,$$
$$f_2\colon\mathbb{R}^{n-k}\to\mathbb{R},\quad f_2(x_{k+1},\dots,x_n)=x_{k+1}^2+\cdots+x_n^2,$$

we are in the setup of Definition 3.3. We define the (n,k)-*boat move* to be the replacement of the graphical open subset Λ_0 by the $\Gamma(f_1)$-suspension of $R1_{n-k}$, $\Sigma_{\Gamma(f_1)}(R1_{n-k})$. Let $B_{n,k}$ denote the resulting Legendrian (locally). We refer to $B_{n,k}$ as the (n,k)-*boat*.

Remark 3.6 If $k=n$, the boat move does not change anything. If $k=0$, the boat move is equal to the n-dimensional first Reidemeister move.

Proposition 3.7 *The (n,k)-boat $B_{n,k}$ is Legendrian isotopic to Λ_0 relative to boundary.*

Proof As the boat move is a special case of the suspension defined in Definition 3.3, $B_{n,k}$ is Legendrian isotopic to Λ_0 relative to boundary, directly from Proposition 1.1. □

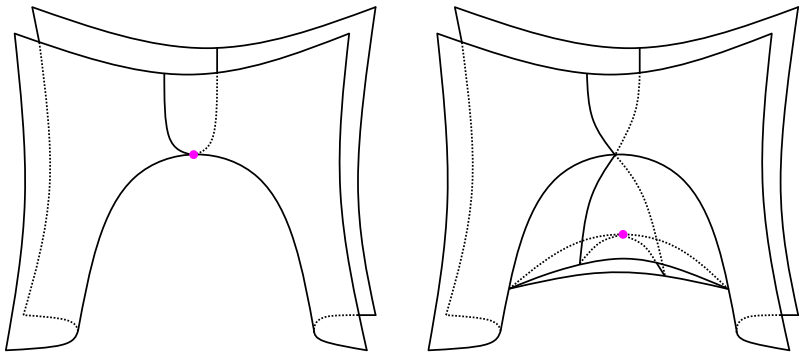


Figure 9: A (2,1)-boat move which changes a saddle point into a maximum.

Remark 3.8 The boat move gets its name because the $(2,1)$-boat looks like a overturned boat or canoe, as seen in Figure 9. The $(2,1)$-boat is very similar to the unigerm $A_3^{e,\pm}$, birth of cuspidal lips; see [3] or [16]. The main difference is the parametrization of the Legendrian front before the move.

Our goal for introducing the boat was that we wanted all the Reeb chords to correspond to maxima of the front. To do so, we will use the following proposition:

Proposition 3.9 *Decompose the front of $B_{n,k}$ into a disjoint union of a finite number of graphical components away from nonsmooth points. All but one of these graphical components are graphs of functions with no critical points. Further, the unique component with critical points has a unique critical point that is a maximum. Moreover, the tangent planes at the cusps are not parallel to the $(x_1,\dots,x_n)$-plane.*

Proof The front of $B_{n,k}$ has a critical point only at points where both the front of $R1_{n-k}$ and f_1 have critical points. This happens at only one point, namely $x_1 = \cdots = x_n = 0$. By construction of f_1, $x_1 = \cdots = x_n = 0$ is assumed to be a maximum in the first k coordinates. The Reidemeister move then converts the critical point to a maximum in the last $n-k$ coordinates, thus proving the proposition.

We get that the cusps are not parallel to the $(x_1,\dots,x_n)$-plane as the $(n-k)$-dimensional Reidemeister front does not have cusps parallel to the $(x_{k+1},\dots,x_n)$-plane. □

Remark 3.10 Consider a Reeb chord with one end on the pink dot in the left figure of Figure 9. This Reeb chord survives the boat move. One may be confused about how the grading of this Reeb chord as an element of the Legendrian dga changes, since the local Morse index is changing. However, note that the relationship between the grading and local Morse index is not the typical one [12], as the cusps we use do not rotate the Lagrangian planes the way typical cusps in the literature do. As such, one can check that the associated Maslov indices of loops of tangent vectors from the top to the bottom of the Reeb chord actually stay the same under our boat moves, preserving the grading.

4 Weinstein presentations of subdomains

In this section we prove Theorem 1.2 and Corollary 1.4. We then apply this theorem to construct several explicit examples of handle decompositions for Weinstein manifolds obtained via antisurgery. We conclude with some open questions.

4.1 Proof of Theorem 1.2

Recall that Theorem 1.2 required that certain Reeb chords be nondegenerate and correspond to critical points of Morse functions. The following result shows that locally, Reeb chords can always be assumed to correspond to critical points of a Morse function. Given a parametrized Reeb chord $c\colon[0,T]\to Y$, we will let $\partial_- c$ and $\partial_+ c$ denote the negative and positive endpoints $c(0)$ and $c(T)$, respectively, of c.

Lemma 4.1 *Let $\Lambda_2\subset J^1(\Lambda_1)$ be a compact Legendrian submanifold of the 1-jet space of a smooth manifold Λ_1, equipped with the standard contact structure. Then we may perturb Λ_2 by a C^∞-small isotopy so that there are finitely many Reeb chords c_i from the zero-section Λ_1 of $J^1(\Lambda_1)$ to Λ_2 and so that, in a neighborhood of the positive endpoint $\partial_+ c_i\in\Lambda_2$, the Legendrian Λ_2 looks like (x, df, f) for f a Morse function on a neighborhood (in Λ_1) of the negative endpoint $\partial_- c_i\in\Lambda_1$.*

Proof Let $\pi\colon J^1(\Lambda_1)\cong T^*(\Lambda_1)\times\mathbb{R}\to T^*(\Lambda_1)$ be the Lagrangian projection onto the cotangent bundle. We will use the notation Λ_1 and Λ_2 for both the original Legendrians as well as their images under the projection. By the Thom transversality theorem, we can perturb Λ_2 by a C^∞-small Hamiltonian isotopy in $T^*\Lambda_1$ so that Λ_1 and Λ_2 intersect transversely (this then corresponds to a C^∞-small Legendrian isotopy in the lift). Let $q\in\Lambda_1$ be one of these intersection points. Since $T\Lambda_2$ intersects $T\Lambda_1$ transversely at q, we have that $T\Lambda_2$ is graphical over the cotangent fiber $T(T^*_q\Lambda_1)$, and hence given by dQ, where $Q\colon T^*_q\Lambda_1\to\mathbb{R}$ is a quadratic form. Now we may apply a C^∞-small isotopy in the space of quadratic forms from Q to a nondegenerate quadratic form $Q'\colon T^*_q\Lambda_1\to\mathbb{R}$. This isotopy induces an isotopy of Lagrangian planes from dQ to dQ' in $T^*(T^*_q\Lambda_1)$. The Lagrangian plane dQ' is transverse to $T_q\Lambda_1$ since it is still graphical over $T^*_q\Lambda_1$ and now it is transverse to $T^*_q\Lambda_1$ since the quadratic form is nondegenerate. This (linear) isotopy of Lagrangian planes can be lifted to C^∞-small linear Hamiltonian isotopy. By taking a bump function, we get a compactly supported C^∞-small Hamiltonian function that takes Λ_2 to a Lagrangian Λ_2' that intersects Λ_1 at transverse intersection points $q\in\Lambda_1$ where Λ_2' also intersects the fiber $T^*_q\Lambda_1$ transversely. Since the isotopy is C^∞-small, no new intersection points are introduced, which means no new Reeb chords are introduced. Finally, we observe, because of these transversality conditions, $\pi(\Lambda_2')$ locally looks like df, where f is a Morse function locally defined near q. □

Remark 4.2 In Lemma 4.3, the Reeb chords we investigate may be interpreted as lying between Λ and Λ_- or between Λ and Λ_+. This is relevant when we refer to the Reeb chords as being represented as a maximum or a minimum of a Morse function in a suitable neighborhood; what is maximum with regard to a function on Λ_- will be a minimum with regard to a function on Λ_+. Here it should be understood

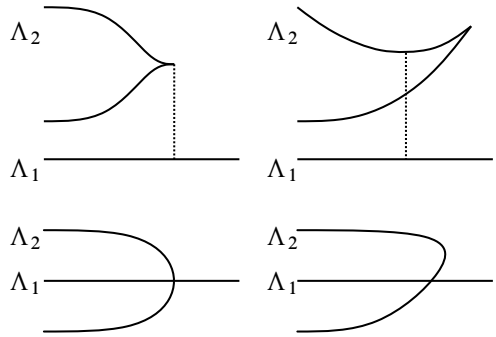


Figure 10: Front and Lagrangian projections of two Legendrians Λ_1 (zero-section) and Λ_2. Top left: there is a Reeb chord (depicted by the dotted line) between the Legendrians Λ_2 and Λ_1 with endpoint on the cusp of Λ_2. Bottom left: the corresponding Lagrangian projection. Top right: after Legendrian isotopy of Λ_2, the Reeb chord between Λ_1 and Λ_2 has endpoint on a smooth branch of Λ_2 which is locally described by the graph of a Morse function on Λ_1 (with minimum corresponding to the Reeb chord endpoint). Bottom right: the corresponding Lagrangian projection.

that we can only handleslide Λ past Λ_- at Reeb chords represented by maxima with respect to a locally defined Morse function on Λ_+ (resp. minima with respect to a locally defined Morse function on Λ_-). The following lemma uses suitable boat moves to turn all critical points into such maxima (resp. minima).

Lemma 4.3 *Suppose Λ is a Legendrian sphere in a contact manifold (M_0,ξ_0), and that (-1)-surgery on Λ produces the contact manifold (M',ξ'). Suppose Λ_+ and Λ_- are a pair of n-dimensional Legendrian spheres which are completely parallel (that is, Λ_- is the ε Reeb pushoff of Λ_+ for $\varepsilon>0$) in (M_0,ξ_0) but not necessarily parallel in (M',ξ'), ie they may be distinctly linked with Λ.*

Let (M,ξ) be the $(2n+1)$-dimensional contact manifold obtained from (M_0,ξ_0) by (-1)-surgery along Λ and Λ_-, and $(+1)$-surgery along Λ_+. Then there exists a Legendrian submanifold $\Lambda_f\subset(M_0,\xi_0)$ such that (M,ξ) can be obtained by only a (-1)-surgery along the components of Λ_f.

Proof To obtain Λ_f, our goal will be to cancel Λ_+ with Λ_- in a surgery diagram of (M,ξ). In order to do so, we need Λ_+ and Λ_- to be completely parallel, ie they must be identically linked with Λ in (M_0,ξ_0). We will parallelize Λ_+ and Λ_- by performing a sequence of Legendrian isotopies and handleslides that preserve the resulting surgered contact manifold (M,ξ); see the example of Figure 11.

Consider a neighborhood $U\cong J^1(\Lambda_+)$ of Λ_+ in M_0 that contains Λ_- as the graph of a constant positive function ε and the subset $\Lambda\cap U$, which we will also call Λ. By Lemma 4.1, we may perturb Λ by a Legendrian isotopy so that we have finitely many Reeb chords, and in a neighborhood of each such Reeb chord the height difference is a Morse function. In general, there may be many other large-period Reeb chords in M_0 from Λ to Λ_- that leave U, but these will not affect our construction, which is local to Λ_-.

If we attempt to isotope Λ_- towards Λ_+ by pushing Λ_- in the z-coordinate in the front, we are obstructed whenever f has a Morse critical point on a part of Λ between Λ_- and Λ_+. These critical points

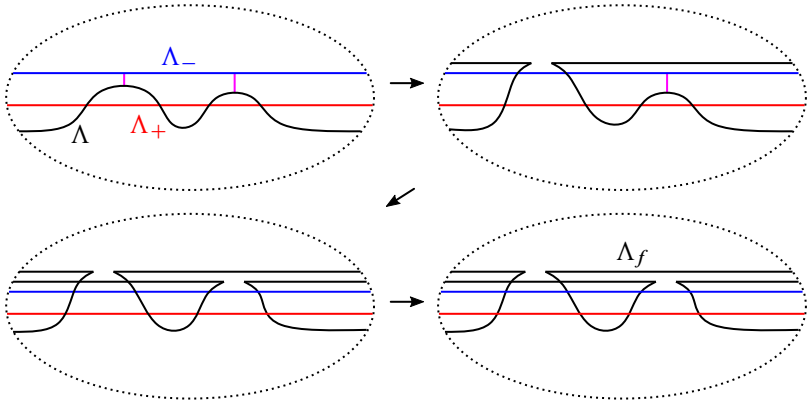


Figure 11: Assuming they are parallel outside of the dotted circle, the red Legendrian Λ_+ and blue Legendrian Λ_-, where we perform a $(+1)$ and (-1) Legendrian surgery, respectively, are made parallel by handlesliding the black Legendrian Λ over Λ_- at its maximum. The Λ_+ and Λ_- are then canceled, leaving only the new black Legendrian Λ_f.

correspond to nondegenerate Reeb chords between Λ_- and Λ and must be removed. We do so one at a time: we always work with the smallest remaining critical value, that is, the shortest remaining Reeb chord that obstructs our isotopy.

Consider the critical point, say q, with least critical value. If q has Morse index 0, then in the front projection we see a local maximum with respect to Λ_+ (see Remark 4.2). We remove it by performing a handleslide of Λ over Λ_- along the Reeb chord c between Λ_- and Λ corresponding to q. We then isotope Λ_- towards Λ_+. See Figure 11. Note that after the handleslide, there is exactly one fewer Reeb chord from Λ to Λ_- contained in U. First, the handleslide along c removes c. Furthermore, the rest of the Reeb chords remain and no other Reeb chords from Λ to Λ_- are produced in U. To see this, we note that the handleslide modifies Λ by first pushing the maximum through Λ_- and doing a cusp connected sum with a positive Reeb pushoff $R_\varepsilon(\Lambda_-)$ of Λ_-. The first step removes the Reeb chord c but does not create new Reeb chords since the pushing through the maximum can be obtained by increasing the function, without creating new critical points (this is not true if we push through an index $k > 0$ critical point as we discuss in the next paragraph). The second step of cusp connect summing with $R_\varepsilon(\Lambda_-)$ creates a family of Reeb chords from Λ_- to Λ but no Reeb chords from Λ to Λ_- since there are no Reeb chords from the positive pushoff $R_\varepsilon(\Lambda_-)$ to Λ_- contained in U. As before, there may be large-period Reeb chords from the handleslid Λ to Λ_- (for example from $R_\varepsilon(\Lambda_-)$ to Λ_-) that leave the neighborhood U. However, these large Reeb chords do not affect our construction, whose goal is to remove local Reeb chords that obstruct the isotopy between Λ_- and Λ_+ in U.

If q has Morse index $n-k$, for $0 \le k < n$, we first perform a (n,k)-boat move to Λ, as defined in Definition 3.5. (If the critical point has Morse index n, then it is a minimum in the front projection, so we perform an $(n,0)$-boat move, which is just an n-dimensional first Reidemeister move). By Proposition 3.9,

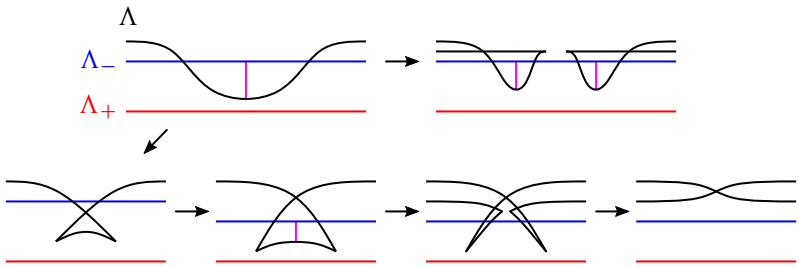


Figure 12: Top row: doing a -1 handleslide of Λ over Λ_- at an index-1 critical point (of the height difference function) creates two more Reeb chords. Bottom row: first doing a boat move at the index-1 critical point (a first Reidemeister move) and then a -1 handleslide over Λ_- removes all Reeb chords. We observe that the resulting Legendrian is a crossing connect sum of Λ and Λ_- which appears when doing a $+1$ handleslide; see Figure 5.

the (n,k)-boat move is a Legendrian isotopy that converts the index-$(n-k)$ critical point q to a maximum, say q', in the front projection. We now have a Reeb chord, say $\gamma_{q'}$, between Λ and Λ_- corresponding to q'. Proposition 3.9 implies that within a small neighborhood no other Reeb chords are created in this process. To see this note that such Reeb chords only exist when Λ and Λ_- are parallel in the front projection. Within the Darboux neighborhood where we perform the boat move, the front of Λ_- is given by the $(x_1,\dots,x_n)$-plane. As Proposition 3.9 says no other smooth critical points are created and the tangent spaces at the newly created cusp points are not parallel to the $(x_1,\dots,x_n)$-plane, the boat move does not create any additional Reeb chords within this neighborhood. Outside the neighborhood, we do not make any changes to both Λ and Λ_-. Perhaps we make changes to some Reeb chords but these are not within our neighborhood.

We isotope Λ_- towards q' until $\gamma_{q'}$ is uninterrupted. Next, we handleslide Λ over Λ_- along the Reeb chord $\gamma_{q'}$ at q'. We can then isotope Λ_- further towards Λ_+. This process is fully illustrated in Figures 12 and 13.

We repeat this process until no obstructing nondegenerate Reeb chords remain and Λ_+ and Λ_- are in canceling position. Once canceled, we are left with a single Legendrian Λ_f in the surgery diagram of (M,ξ). In summary, Λ_f corresponds to Λ in the following way: for each Reeb chord γ from Λ to Λ_- with local index $n-k$, we applied an (n,k)-boat move to Λ and did a cusp connected sum with Λ_-. □

Remark 4.4 If we handleslide over a Reeb chord corresponding to an index-k critical point without first doing a boat move, then new Reeb chords are created; see the top row of Figure 12. By first doing the boat move, which does not create any new Reeb chords but only changes the local index of the existing Reeb chord, we ensure that handlesliding removes that Reeb chord without creating any new chords.

Remark 4.5 Although Lemma 4.3 is stated in the language of contact manifolds and contact (±1)-surgeries, the result also holds with Weinstein homotopies. We recall that (-1)-surgeries correspond to Weinstein handle attachment while $(+1)$-surgeries correspond to handle removal, and in general do not

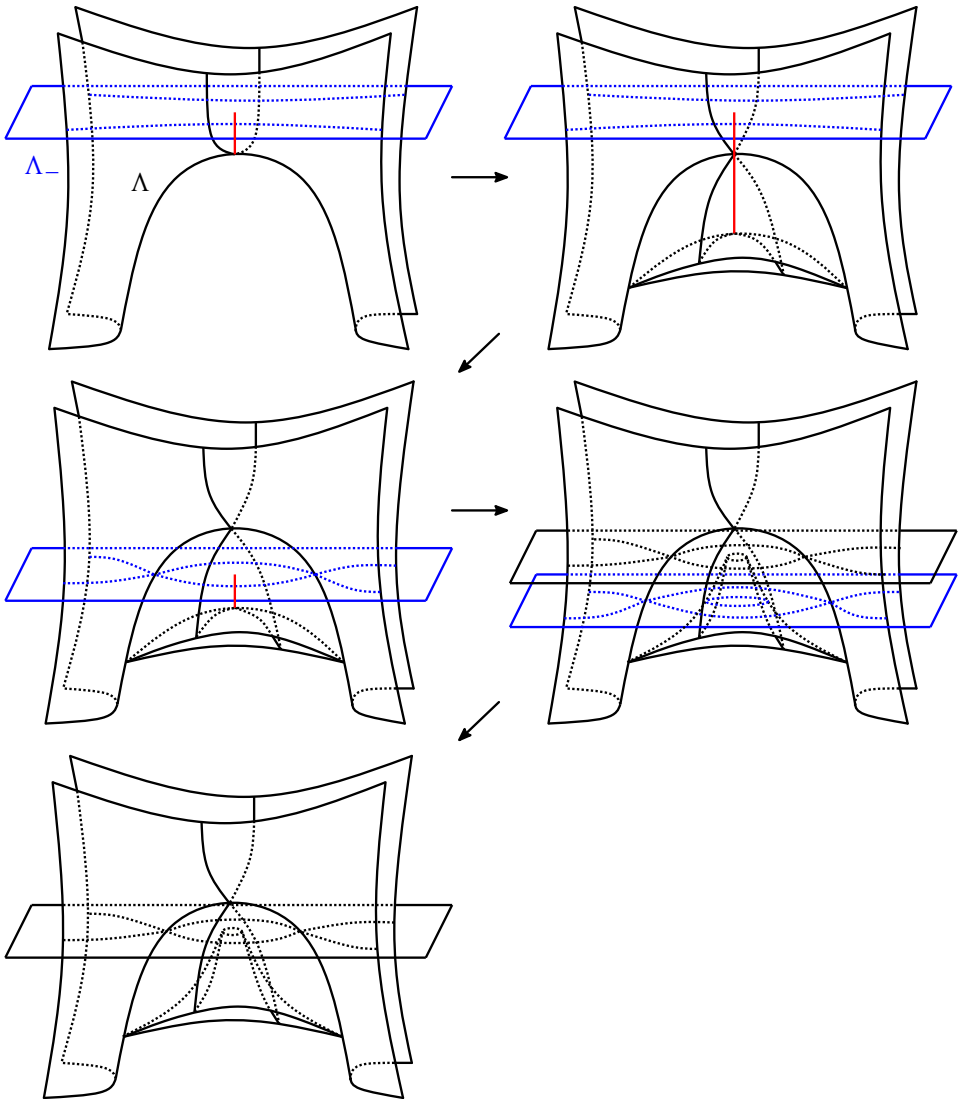


Figure 13: The process of performing a boat move locally for an index-1 critical point on a surface Legendrian to enable a handleslide: (1) The black Legendrian Λ has an index-1 critical point at which there is a red Reeb chord to the blue Legendrian Λ_-. (2) Perform a boat move at the critical point; the Reeb chord now meets the Λ at a maximum in the front diagram. (3) Isotope Λ_- downwards; the Reeb chord is now uninterrupted in the front diagram. (4) Perform a handleslide along the Reeb chord, producing a small circle of cusps in the middle of the boat. (5) Isotope the Λ_- past the boat.

produce fillable contact manifolds. However, in our case, all surgery moves are handleslides are over the (-1)-Legendrian, and hence we can view the $(+1)$-Legendrian as a placeholder for the boundary of the Lagrangian disk, which will be removed at the last step. Hence all our contact moves are really Weinstein homotopies.

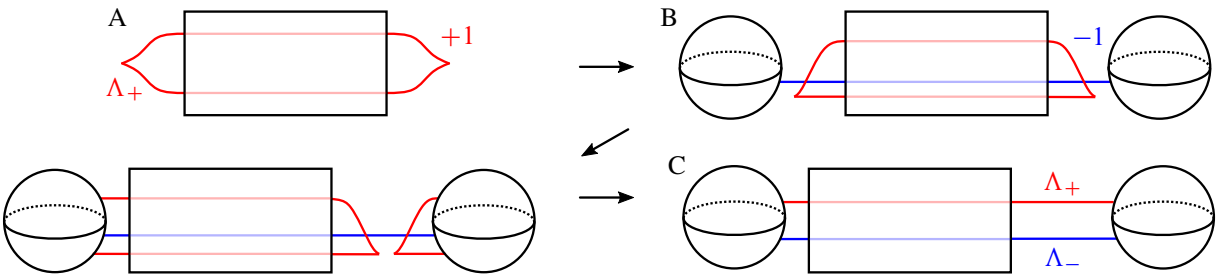


Figure 14: The diagrams A, B, and C. In the top left, we see a red Legendrian unknot in the boundary of a Weinstein domain. This unknot may be linked with attaching handles of the domain away from its cusps, represented by the black rectangle. This diagram is related to the figure on the bottom right by the addition of a canceling pair and a single handleslide.

Remark 4.6 In the next two proofs, we abuse terminology and refer interchangeably to Weinstein diagrams and Weinstein domains, so we sometimes say "attach a handle to the Weinstein diagram". We will also refer to the carving out of a Lagrangian disk as "antisurgery" along its boundary Legendrian. We hope this abuse of terminology will not confuse the reader but, in fact, make the proof easier to read.

Proof of Theorem 1.2 Recall that $L \subset X$ is of the form $X = T^*D^n \cup H_i^n$ and $L = D^n \subset T^*D^n$. Then the Weinstein sector $X \setminus L$ is obtained by antisurgery along the Legendrian knot ∂L which corresponds to the unknot $\partial D^n \subset T^*D^n$. From now onward, we will denote the knot ∂L by Λ_+, and depict it in red in all figures. Let $\Lambda = \bigcup_i \Lambda_i$ denote the link consisting of all the attaching spheres of the H_i^n.

We may isotope Λ_+ so that the front projection of Λ is disjoint from the cusps of Λ_+, giving us a surgery diagram, say diagram A, for $\partial(X \setminus L)$, as depicted in the top left of Figure 14. We introduce a canceling pair to diagram A — an $(n-1)$-handle and a critical handle Λ_- such that Λ_- is a parallel pushoff of the bottom arc of Λ_+ — to obtain a new surgery diagram, say diagram B, again for $\partial(X \setminus L)$, as depicted in top right of Figure 14. We observe that diagram B is in turn equivalent to a "diagram C" of the form of the bottom right of Figure 14, where Λ_+ and Λ_- both traverse across the $n-1$ handle exactly once each.

We note that, in diagram C, Λ_+ and Λ_- are parallel except in how they are linked with Λ. So we can apply Lemma 4.3 to obtain a surgery presentation that contains only (-1)-surgery along Legendrian submanifolds, say diagram D. Then this diagram D corresponds to a Weinstein handle diagram where each (-1)-surgery corresponds to a critical handle attachment.

Note that the Weinstein diagram in diagram D has exactly one more index-$(n-1)$ handle than the Weinstein presentation for X. Further, if we denote the attaching spheres of the index-n handles H_i in diagram D by Λ_i', respectively, the construction in Lemma 4.3 implies that the Λ_i' are exactly as described in the theorem statement. □

As a corollary, we explain how to describe the P-loose Legendrian unknot in $\mathbb{R}^{2n-1} \subset S^{2n-1} = \partial B^{2n}$. To do so, we first describe a front diagram of the knot in $S^{n-1} \times \mathbb{R}^n \subset \partial(B^{2n} \cup H^{n-1})$ and then attach a flexible handle to make the ambient space ∂B^{2n}.

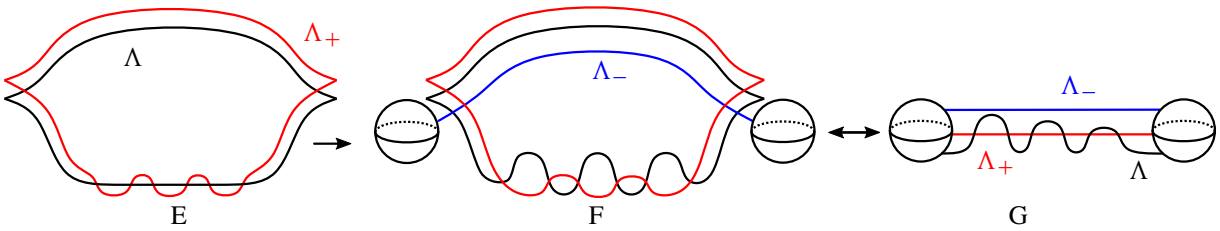


Figure 15: A canceling pair of handles, represented by a pair of spheres and a blue Legendrian, is added just below the cusps of the linked red and black Legendrians. This diagram is equivalent to the final one via two handleslides.

Proof of Corollary 1.4 Recall from Section 2.4 and [20] that the P-loose Legendrian unknot,

$$\Lambda_P \subset S^{2n-1} = \partial B^{2n}$$

is obtained as follows. First, to construct a P-Moore space, consider the CW complex $S^1 \cup_P D^2$, the result of attaching D^2 to S^1 along the degree-p map $\partial D^2 = S^1 \to S^1$. If $n \geq 5$, then this CW complex embeds into S^{n-1}, as observed in [1], and we let U be a neighborhood of this CW complex. Next, we observe that U has a Morse function (with gradient outward-pointing near the boundary of U) that has three critical points, one of index 0 and 1 for the S^1, and one of index 2 for the D^2.

Next, we carve out a disk $D_U \subset T^*D^n$ from T^*D^n to obtain $T^*D^n \setminus D_U$, which as an unstopped Weinstein domain and is equivalent to $B^{2n} \cup H^{n-1}$. Then Λ_P is

$$\partial D \subset \partial(B^{2n} \cup H^{n-1} \cup H_{\text{flex}}),$$

where H_{flex} is a flexible Weinstein handle attached along a Legendrian which is loose in the complement of $\partial D \subset B^{2n} \cup H^{n-1}$. Additionally, H_{flex} is in canceling position with H^{n-1}. We will explicitly construct the Weinstein handle decomposition of this $B^{2n} \cup H^{n-1}$ to obtain an explicit front diagram for Λ_P.

We begin by using the construction from Proposition 2.8. In the boundary $S^{2n-1} = \partial B^{2n}$, we consider a pair of canceling contact surgeries along parallel $(n-1)$-dimensional Legendrian unknots. We label the Legendrian unknot corresponding to the (-1)-surgery Λ and the Legendrian unknot corresponding to the $(+1)$-surgery Λ_+. Here Λ_+ and Λ represent Legendrian boundaries ∂D_U and ∂D, respectively. Note that there exists a P-Moore space $U \subset \Lambda_+$ by our assumption. We perturb Λ_+ past the Λ by "pushing" the subdomain U so that Λ_+ and Λ are no longer parallel. Let us refer to this surgery diagram as diagram E.

To diagram E, we insert a canceling pair of handles consisting of an $(n-1)$-handle H^{n-1} and an n-handle H^n, just below the cusps of Λ_+ and Λ; see Figure 15. Note that as a contact surgery curve on the boundary, the n-handle attachment corresponds to a (-1)-surgery along a Legendrian knot. Let us denote the Legendrian knot corresponding to this (-1)-surgery by Λ_-, and refer to this Weinstein diagram as diagram F. We see from Figure 15 that diagram F is equivalent to diagram G after some handleslides and Legendrian isotopies.

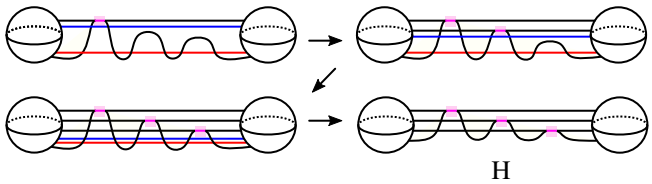


Figure 16: Applying Lemma 4.3 to diagram G involves three instances of a boat move isotopy followed by a handleslide. Then Λ_+ and Λ_- are completely parallel; thus the respective surgeries on these knots cancel.

We are now set up exactly as in the proof of Theorem 1.2, namely diagram G is of the form of diagram C in the proof of Theorem 1.2. So we similarly apply Lemma 4.3 to obtain the required Weinstein diagram, say diagram H. Since the P-Moore space has three critical points, Λ in diagram G also has three critical points, of index 0, 1, and 2 when considering the Morse function f of Lemma 4.3. Thus in applying the lemma, we perform three instances of a boat move and a cusp connect sum to Λ in diagram G to obtain diagram H; see Figure 16.

Finally, to obtain Λ_P in B^{2n}, that is, to make the ambient manifold B^{2n}, we attach a flexible handle, H_{flex}, to cancel out H^{n-1} in diagram H. This amounts to attaching a loose Legendrian Λ_{flex} that winds around the $(n-1)$-handle once, and is in the complement of Λ in diagram H. This gives us diagram I, which is depicted in the top left of Figure 17, with Λ_{flex} denoted in pink.

Next, in diagram I, we slide Λ repeatedly over H_{flex} until Λ no longer passes through the H^{n-1}. We then cancel H^{n-1} with H_{flex}. We are left with a surgery diagram, say diagram J, that consists of a single Legendrian Λ_P in B^{2n}. This process is illustrated in Figure 17. We see that Λ_P consists of four loose Legendrian unknots which are completely parallel, away from a bounded region where they are linked (in a way that depends on p) and are connected via three boat moves and cusp connected sum gluings.

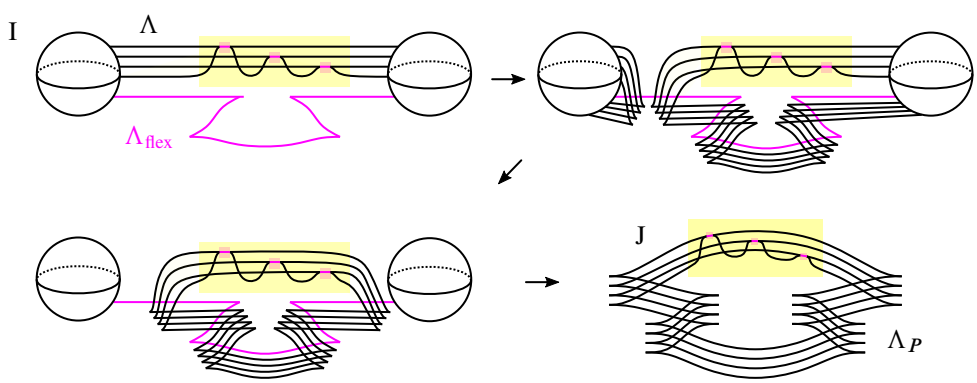


Figure 17: The final step in the construction: In the complement of the black Legendrian Λ which traverses the $(n-1)$-handle some number of times, we attach a flexible handle along the pink loose Legendrian Λ_{flex}. We then repeatedly slide Λ over Λ_{flex} to detach it from the $(n-1)$-handle. Finally, we cancel Λ_{flex} with the $(n-1)$-handle.

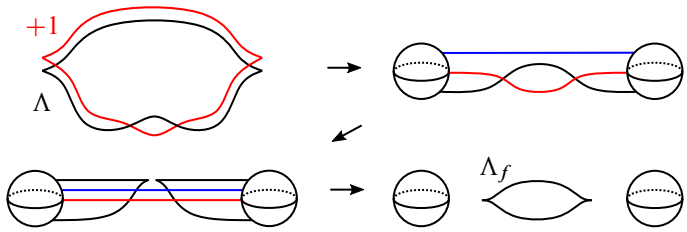


Figure 18: Our construction applied to the case where $U = D^k$. We first add a canceling pair and slide the red and black Legendrians over the blue Legendrian. Then at the maximum on the black Legendrian, we slide over the blue Legendrian. Passing the cusps over the $n-1$ handle and canceling the red with the blue, we obtain a max tb unknotted sphere in the complement of the $n-1$ handle.

As a Weinstein diagram, diagram J depicts the Weinstein sector (B^{2n}, Λ_P), where Λ_P is a P-loose Legendrian unknot. □

4.2 Explicit examples

We now construct several exotic Weinstein manifolds as applications of the construction from Theorem 1.2 and Corollary 1.4. In all these examples, we consider antisurgery on the Lagrangian disk obtained by perturbing the boundary $S^{n-1} = \partial D^n$ of the zero-section $D^n \subset T^*D^n$ in a neighborhood $U \subset S^{n-1}$, as in the construction of P-loose Legendrians (see Section 2.4).

Example 4.7 Suppose $U = D^{n-1} \subset S^{n-1}$ is a disk. Then we may choose the Morse function $g: U \to \mathbb{R}$ to have a single critical point of index 0. So we obtain Λ_U with a single maximum. Applying the construction from Theorem 1.2, we obtain a Legendrian Λ_f by a single handleslide (see Figure 18). The resulting Legendrian, Λ_f, is a standard Legendrian unknot in the complement of the $(n-1)$-handle, H^{n-1}.

Here Λ_f is not loose, and is the Legendrian unknot in the subcritical domain $B^{2n} \cup H^{n-1}$. Indeed, we have constructed the Weinstein diagram of $T^*D^n \cup H^{n-1}$, ie T^*D^n with a subcritical handle attached along a subcritical isotropic sphere in a Darboux chart. Λ_+ is the Legendrian unknot and bounds the Lagrangian unknot; carving out the Lagrangian unknot is equivalent to attaching an index-$(n-1)$ handle.

Example 4.8 Let $U \subset S^{n-1}$ be the disconnected union of a codimension-zero submanifold U' and a disk D^{n-1}. In this case, after cancellation, we expect the remaining Legendrian to be loose. This is because $D^n_{U' \coprod D^{n-1}}$ is Lagrangian isotopic to $D^n_{U'} \natural T^*_0 D^n$, and by [19], $T^*D^n \setminus (D_{U'} \natural T^*_0 D^n)$ is obtained from the subcritical sector $T^*D^n \setminus (D_{U'} \coprod T^*_0 D^n)$ by attaching a flexible handle. We can also see this from our explicit construction as follows. Since U is disconnected, we may isotope Λ_U so that the Reeb chords coming from $g|_{U'}$, the U'-perturbation, are shorter than the Reeb chord coming from the maximum of $g|_{D^{n-1}}$, the D^{n-1}-perturbation (see Figure 19).

Next, we apply the construction of Theorem 1.2 to obtain the Legendrian attaching sphere Λ_f as follows. A combination of boat moves and handleslides will first remove all the critical points coming from U'.

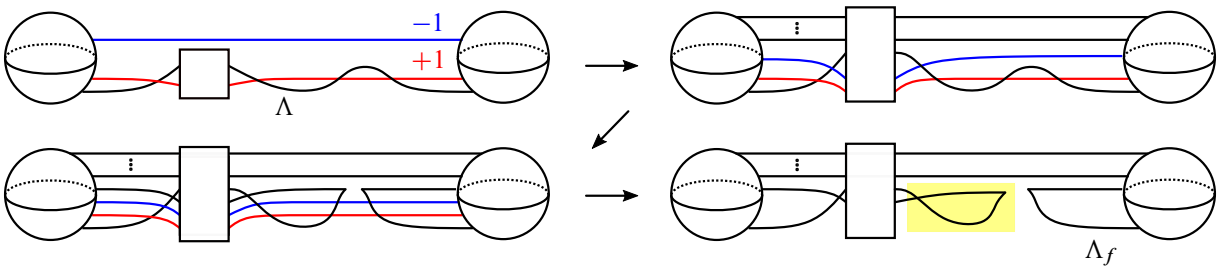


Figure 19: Our construction applied to the case where U is a union of some submanifold (the rectangular region) and a disk (the maximum on the right). We first use boat moves and handleslides to remove all obstructing Reeb chords coming from the boxed region. A single handleslide then puts the red and blue Legendrians in a canceling position. The yellow highlighted region is a loose chart.

Then a single handleslide will remove the critical point coming from D^{n-1} (see Figure 19). We are now in position to cancel Λ_- and Λ_+ as in the proof of Theorem 1.2.

After this cancellation, the resulting Legendrian knot Λ_f has a loose fishtail chart in a transverse slice. By [21], this implies that Λ_f is loose.

Example 4.9 When $0 \in P$, the P-loose Legendrian Λ_P is loose. To see this, first recall that

$$D_U = \Gamma(df) \cap B^{2n} \quad \text{for a function } f : D^n \to \mathbb{R}$$

which is an extension of a Morse function $f : S^{n-1} \to R$ that is negative on U and positive on the closure of the complement. It is enough to prove that Λ_P is loose for $P = \{0\}$; as discussed in [20, Section 2.2.2], $\Lambda_{P \coprod Q}$ is isotopic to $\Lambda_P \# \Lambda_Q$ and the connected sum of any Legendrian with a loose Legendrian (in a separate Darboux chart) is loose.

Let $U = S^k$, that is, U is a P-Moore space for $P = \{0\}$ since its relative cohomology is $\mathbb{Z} \cong \mathbb{Z}/\{0\}$ in positive degree. Consider U to be embedded in S^{n-1} as the intersection $D^{k+1} \cap S^{n-1}$. Then one can take the function $f : D^n \to \mathbb{R}$ to be a perturbation of

$$-(x_1^2 + \cdots + x_{k+1}^2) + x_{k+2}^2 + \cdots + x_n^2.$$

Then $D_U = \Gamma(df)$ is Hamiltonian isotopic to the cotangent fiber, $T_0^* D^n$, in $T^* D^n$. So $T^* D^n \setminus T_0^* D^n$ is just $B^{2n} \cup H^{n-1}$. Further, the stop ∂D_U is a Legendrian that passes through the H^{n-1} exactly one time, ie it is a loose Legendrian.

Next, we will explain how to see this explicitly by following the construction in the proof of Theorem 1.2. We use the notation from the proof of Theorem 1.2 as well. On $U = S^k$, f has two critical points, namely, a maximum and a saddle point. Following the construction, to get an explicit Weinstein diagram we perform a single boat move and two handleslides. After canceling Λ_- and Λ_+, the resulting Legendrian (diagram H) intersects the subcritical $(n-1)$-handle H^{n-1} three times. This is because initially ∂D^n

intersected H^{n-1} once, and each handleslide introduces one new intersection. For $n = 3$, this process is illustrated in Figure 20. In higher dimensions the construction follows analogously.

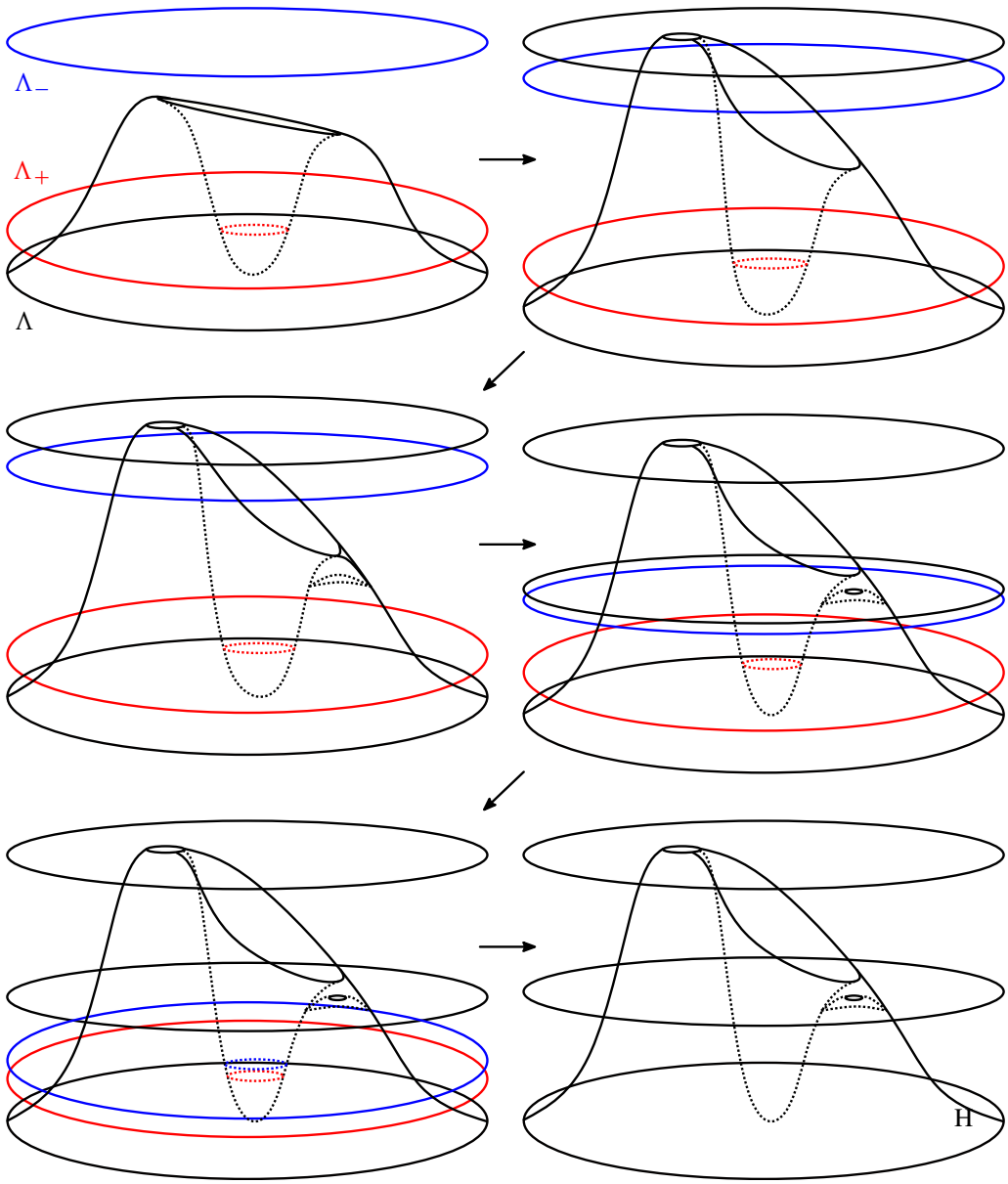


Figure 20: The process to obtain diagram H when $n = 3$. The black Legendrian Λ begins with a maximum and a saddle point between the blue Λ_- and the red Λ_+. Away from the pictured region, Λ_- and Λ_+ are parallel. We perform the following moves: (1) At the maximum, we perform a handleslide. (2) At the saddle point, we perform a boat move. (3) We then isotope Λ_- downwards, and handleslide Λ over Λ_-. (4) We further isotope Λ_- so that it is in canceling position with Λ_+. (5) Finally, we cancel Λ_- and Λ_+, so that only Λ remains.

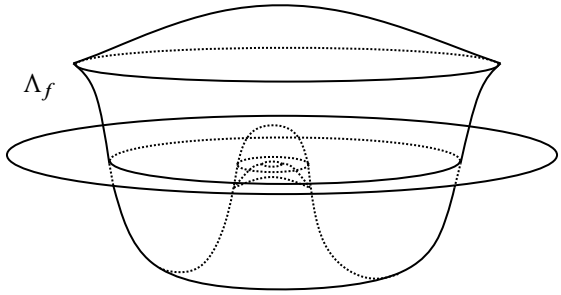


Figure 21: After a Legendrian isotopy which passes the top circle of cusps across the 2-handle, the connected black Legendrian resulting from the recipe in Figure 20 is in canceling position with the subcritical handle whose attaching region is a torus around the middle disk in this dimension.

With diagram H, instead of continuing with the construction in the proof of Theorem 1.2 as is, we first do an additional step. We perform a Legendrian isotopy that passes the circle of cusps over H^{n-1}. The resulting Legendrian, $\Lambda_f = \partial D_U$, passes through H^{n-1} exactly one time.

To conclude, note that Λ_f is now in canceling position with H^{n-1}; see Figure 21. Hence it is loose, and remains loose when the flexible handle H_{flex} is attached alongside it to cancel H^{n-1}. Thus we obtain that Λ_0 is loose.

4.3 Questions

One can construct a loose Legendrian unknot by pushing through any codimension-zero subdomain U (with boundary) past the Legendrian unknot *near a cusp*. As observed in [21], Λ_U is always loose; see Figure 7. If the Euler characteristic of U is 0, then Λ_U is formally Legendrian isotopic to Λ (but not genuinely Legendrian isotopic), and hence called the loose Legendrian unknot. So by the h-principle for loose Legendrians, Λ_U and Λ_V are isotopic if $\chi(U) = \chi(V)$.

Our construction of the P-loose Legendrians also involves pushing through certain codimension-zero subdomains (neighborhoods of P-Moore spaces). However, here the construction is less concrete; one must first push through to create the Lagrangian disk D_U, then carve out D_U, and then attach a flexible handle, ultimately resulting in our recipe above. Hence it is natural to ask whether there is a more direct route towards the construction of these P-loose Legendrians, analogous to the construction of loose Legendrians by Murphy.

Question 4.10 *Can a P-loose Legendrian unknot be constructed more directly by pushing through a P-Moore space past a region of the Legendrian unknot (not near a cusp), after Legendrian isotopy of the unknot?*

For example, this pushing operation, if it exists, must have the property that if U is disconnected then Λ_U must be loose, as discussed in Figure 21.

Another line of inquiry is to see whether our algorithm can provide an alternative proof of the Ganatra–Pardon–Shende [14] localization formula from the point of view of Legendrian invariants. The localization formula [14] computes the wrapped Fukaya of the subdomain $X \setminus D$ as

$$\mathcal{W}(X \setminus D) \cong \mathcal{W}(X)/D,$$

where $\mathcal{W}(X)/D$ is the algebraic localization of $\mathcal{W}(X)$ by the divisor D. They describe a concrete formula computing the morphism chain complexes, $\mathrm{Hom}_{\mathcal{W}(X)/D}(L, K)$, via a dg bar construction that depends on the morphism $\mathrm{Hom}_{\mathcal{W}(X)}(L, K)$ as well as $\mathrm{Hom}_{\mathcal{W}(X)}(L, D)$ and $\mathrm{Hom}_{\mathcal{W}(X)}(D, K)$. For certain Lagrangians, like the cocores of X, these complexes can all be computed using the Legendrian dga's of the attaching spheres of X and ∂D. Hence the Ganatra–Pardon–Shende surgery formula can be used a priori to describe the Legendrian dga's of the attaching spheres of $X \setminus D$. On the other hand, here we give an explicit geometric Weinstein presentation for $X \setminus D$ and an explicit depiction of its Legendrian attaching spheres.

Question 4.11 *Can one compute the Legendrian dga's of the attaching spheres of $X \setminus D$ produced by Theorem 1.2 directly, giving an alternative direct proof of the Ganatra–Pardon–Shende localization formula*?

References

[1] **M Abouzaid**, **P Seidel**, *Altering symplectic manifolds by homologous recombination*, preprint (2010) arXiv 1007.3281

[2] **B Acu**, **O Capovilla-Searle**, **A Gadbled**, **A Marinković**, **E Murphy**, **L Starkston**, **A Wu**, *Weinstein handlebodies for complements of smoothed toric divisors*, 1561, Amer. Math. Soc., Providence, RI (2025) MR

[3] **V I Arnold**, *Singularities of caustics and wave fronts*, Math. Appl. (Soviet Ser.) 62, Kluwer, Dordrecht (1990) MR

[4] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR

[5] **F Bourgeois**, **J M Sabloff**, **L Traynor**, *Lagrangian cobordisms via generating families: construction and geography*, Algebr. Geom. Topol. 15 (2015) 2439–2477 MR

[6] **R Casals**, **E Murphy**, *Legendrian fronts for affine varieties*, Duke Math. J. 168 (2019) 225–323 MR

[7] **K Cieliebak**, **Y Eliashberg**, *From Stein to Weinstein and back: symplectic geometry of affine complex manifolds*, Amer. Math. Soc. Colloq. Publ. 59, Amer. Math. Soc., Providence, RI (2012) MR

[8] **G Dimitroglou Rizell**, *Knotted Legendrian surfaces with few Reeb chords*, Algebr. Geom. Topol. 11 (2011) 2903–2936 MR

[9] **G Dimitroglou Rizell**, **R Golovko**, *On Legendrian products and twist spuns*, Algebr. Geom. Topol. 21 (2021) 665–695 MR

[10] **F Ding**, **H Geiges**, *Handle moves in contact surgery diagrams*, J. Topol. 2 (2009) 105–122 MR

[11] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 177–305 MR

[12] **T Ekholm**, **J Etnyre**, **M Sullivan**, *Non-isotopic Legendrian submanifolds in $\mathbb{R}^{2n+1}$*, J. Differential Geom. 71 (2005) 85–128 MR

[13] **Y Eliashberg**, **S Ganatra**, **O Lazarev**, *Flexible Lagrangians*, Int. Math. Res. Not. 2020 (2020) 2408–2435 MR

[14] **S Ganatra**, **J Pardon**, **V Shende**, *Sectorial descent for wrapped Fukaya categories*, J. Amer. Math. Soc. 37 (2024) 499–635 MR

[15] **R Golovko**, *A note on the infinite number of exact Lagrangian fillings for spherical spuns*, Pacific J. Math. 317 (2022) 143–152 MR

[16] **V Goryunov**, **S Alsaeed**, *Local invariants of framed fronts in 3-manifolds*, Arnold Math. J. 1 (2015) 211–232 MR

[17] **K Honda**, **Y Huang**, *Convex hypersurface theory in contact topology*, preprint (2019) arXiv 1907.06025

[18] **O Lazarev**, *Maximal contact and symplectic structures*, J. Topol. 13 (2020) 1058–1083 MR

[19] **O Lazarev**, *Symplectic flexibility and the Grothendieck group of the Fukaya category*, J. Topol. 15 (2022) 204–237 MR

[20] **O Lazarev**, **Z Sylvan**, *Prime-localized Weinstein subdomains*, Geom. Topol. 27 (2023) 699–737 MR

[21] **E Murphy**, *Loose Legendrian embeddings in high dimensional contact manifolds*, PhD thesis, Stanford University (2012) Available at `https://www.proquest.com/docview/245466004`

[22] **E Murphy**, **K Siegel**, *Subflexible symplectic manifolds*, Geom. Topol. 22 (2018) 2367–2401 MR

[23] **J M Sabloff**, **M G Sullivan**, *Families of Legendrian submanifolds via generating families*, Quantum Topol. 7 (2016) 639–668 MR

[24] **K Sackel**, *Getting a handle on contact manifolds*, PhD thesis, Massachusetts Institute of Technology (2019) Available at `https://dspace.mit.edu/handle/1721.1/122165`

[25] **J Świątkowski**, *On the isotopy of Legendrian knots*, Ann. Global Anal. Geom. 10 (1992) 195–207 MR

[26] **A Weinstein**, *Contact surgery and symplectic handlebodies*, Hokkaido Math. J. 20 (1991) 241–251 MR

ID: *Department of Mathematics, ETH Zürich*
Zürich, Switzerland

OL: *Department of Mathematics, University of Massachusetts Boston*
Boston, MA, United States

CM: *Mathematics Department, Fordham University*
New York, NY, United States

AW: *Department of Mathematics, Bucknell University*
Lewisburg, PA, United States

`ipsita.datta@math.ethz.ch`, `oleg.lazarev@umb.edu`, `cmohanakumar@fordham.edu`, `a.wu@bucknell.edu`

Received: 15 October 2023 Revised: 4 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3737–3754
DOI: 10.2140/agt.2025.25.3737
Published: 1 October 2025

Singular Legendrian unknot links and relative Ginzburg algebras

JOHAN ASPLUND

We associate to a quiver and a subquiver (Q, F) a stopped Weinstein manifold X whose Legendrian attaching link is a singular Legendrian unknot link Λ. We prove that the relative Ginzburg algebra of (Q, F) is quasi-isomorphic to the Chekanov–Eliashberg dg-algebra of Λ. It follows that the Chekanov–Eliashberg dg-algebra of Λ relative to its boundary dg-subalgebra, and the Orlov functor associated to the partially wrapped Fukaya category of X both admit a strong relative smooth Calabi–Yau structure.

53D35, 53D42; 16E45

1. Introduction 3737
2. Relative Ginzburg algebras 3741
3. Stopped Weinstein manifold associated to a quiver and a frozen subquiver 3742
4. Proof of the main theorem 3747
References 3753

1 Introduction

Let Q be a quiver and let $F \subset Q$ be a subquiver called the *frozen subquiver*. Define $Z_0(Q)$ to be the subcritical Weinstein manifold obtained by taking Weinstein connected sum of copies of B^{2n} equipped with the standard Liouville form $\lambda = \frac{1}{2}\sum_{i=1}^{n}(x_i dy_i - y_i dx_i)$ according to the arrows of Q. We define a singular Legendrian $\Lambda(Q, F)$ in the contact boundary of $Z_0(Q)$ by first taking a copy of the $(n-1)$-dimensional Legendrian unknot contained in a Darboux chart in the boundary of each copy of B^{2n}. Next, for each arrow in Q we push part of the Legendrian unknot associated to its tail through the Weinstein 1-handle to the copy of B^{2n} associated with the head of the arrow. If the arrow is nonfrozen we make the two Legendrians link once and if the arrow is frozen we make them intersect transversely once. The result is a singular Legendrian unknot link in the boundary of $Z_0(Q)$; see Figure 1. The pair $(Z_0(Q), \Lambda(Q, F))$ specifies a stopped Weinstein manifold $X(Q, F)$ by attaching top Weinstein handles to nonfrozen components of $\Lambda(Q, F)$ and a stop to each frozen component.

We let $CE^*(\Lambda(Q, F); Z_0(Q))$ denote the Chekanov–Eliashberg dg-algebra of $\Lambda(Q, F)$ over a field, which is defined as the Chekanov–Eliashberg dg-algebra after taking a Weinstein neighborhood of each

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

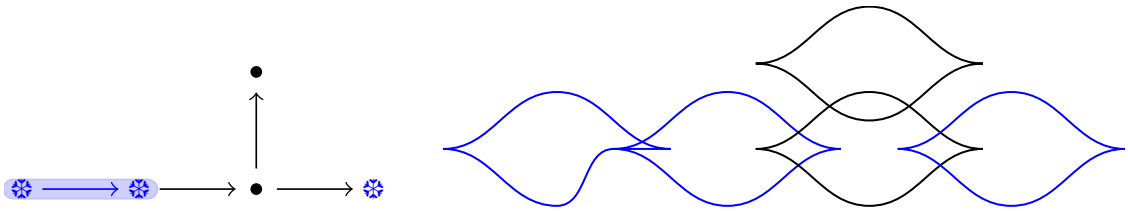


Figure 1: Left: A quiver Q and frozen subquiver F marked with blue snowflakes and blue arrows. Right: Front projection of the corresponding singular Legendrian unknot link $\Lambda(Q,F)$ with frozen components marked in blue.

frozen component of $\Lambda(Q,F)$; see [5, Section 3.1]. There is a canonical geometrically defined dg-subalgebra $\mathcal{B} \hookrightarrow CE^*(\Lambda(Q,F);Z_0(Q))$ that in particular is empty if $F=\varnothing$. Let $\mathcal{G}_n^*(Q,F)$ denote the n-dimensional relative Ginzburg algebra of (Q,F) (see Definition 2.2). There is a map of dg-algebras $G\colon \mathcal{G}_{n-1}^*(F) \to \mathcal{G}_n^*(Q,F)$ (see Lemma 2.3), where $\mathcal{G}_{n-1}^*(F) := \mathcal{G}_{n-1}^*(F,\varnothing)$.

Theorem 1.1 *Let $n \geq 4$ and consider a quiver Q together with a subquiver $F \subset Q$. There are quasi-isomorphisms of dg-algebras $CE^*(\Lambda(Q,F);Z_0(Q)) \cong \mathcal{G}_n^*(Q,F)$ and $\mathcal{B} \cong \mathcal{G}_{n-1}^*(F)$, such that the following diagram commutes*:

$$\begin{array}{ccc} \mathcal{B} & \xrightarrow{\text{incl.}} & CE^*(\Lambda(Q,F);Z_0(Q)) \\ \downarrow\cong & & \downarrow\cong \\ \mathcal{G}_{n-1}^*(F) & \xrightarrow{G} & \mathcal{G}_n^*(Q,F) \end{array}$$

Remark 1.2 (1) Theorem 1.1 also holds for $n=3$ if every vertex of F has valency less than three.

(2) If $F_1=\varnothing$, then $\mathcal{B} \cong C_{-*}(\Omega\Lambda(F))$ where $\Lambda(F) \subset \Lambda(Q,F)$ is the Legendrian sublink consisting of the frozen components only, and the dg-algebra $CE^*(\Lambda(Q,F);Z_0(Q))$ is the Chekanov–Eliashberg dg-algebra of $\Lambda(Q,F)$ with loop space coefficients in $C_{-*}(\Omega\Lambda(F))$ as defined by Ekholm and Lekili [13, Section 3.2].

(3) Theorem 1.1 was proven in the case of $F=\varnothing$ in [3, Corollary 4.16].

Brav and Dyckerhoff [9] introduced the notion of a (strong) relative Calabi–Yau structure. In their terminology a *left* Calabi–Yau structure is a *smooth* Calabi–Yau structure in the terminology of Kontsevich and Soibelman [22]. A relative smooth Calabi–Yau structure is called *weak* if the quasi-isomorphisms in [9, (1.13)] does not necessarily come from a class in relative negative cyclic homology, and *strong* if it does.

Corollary 1.3 *The canonical inclusion $\mathcal{B} \hookrightarrow CE^*(\Lambda(Q,F);Z_0(Q))$ admits a strong relative smooth n-Calabi–Yau structure.*

Proof The map $G\colon \mathcal{G}_{n-1}^*(F) \to \mathcal{G}_n^*(Q,F)$ (see Lemma 2.3) admits a strong relative smooth n-Calabi–Yau structure by Lemma 2.4. Apply Theorem 1.1. □

Let $V(Q,F)$ denote a Weinstein neighborhood of the singular Legendrian $\Lambda(Q,F)\subset\partial Z_0(Q)$. It follows from [5, Theorem 1.1] that there is a derived equivalence

$$\operatorname{Perf} CE^*(\Lambda(Q,F);Z_0(Q))\cong\mathcal{W}(X(Q,F);V(Q,F)),$$

where $\mathcal{W}(X(Q,F))$ denotes the partially wrapped Fukaya category of $X(Q,F)$ stopped at the Weinstein hypersurface $V(Q,F)$ and where Perf denotes the category of perfect modules; see Ganatra, Pardon and Shende [17] and Sylvan [30]. The Orlov functor $\iota_{V(Q,F)}\colon\mathcal{W}(V(Q,F))\hookrightarrow\mathcal{W}(X(Q,F);V(Q,F))$ (see Sylvan [31, Section 2.4] for the definition) fits in a commutative diagram

$$\begin{array}{ccc}\operatorname{Perf}\mathcal{B} & \hookrightarrow & \operatorname{Perf} CE^*(\Lambda(Q,F);Z_0(Q))\\ \downarrow\cong & & \downarrow\cong\\ \mathcal{W}(V(Q,F)) & \xhookrightarrow{\iota_{V(Q,F)}} & \mathcal{W}(X(Q,F);V(Q,F))\end{array}\tag{1-1}$$

leading to the following.

Corollary 1.4 *The Orlov functor $\iota_{V(Q,F)}$ admits a strong relative smooth n-Calabi–Yau structure.*

1.1 Related results

As far as we know, the novelty of Corollary 1.3 is the existence of a *strong* relative smooth Calabi–Yau structure. Quite a bit is known in the nonrelative case or in the weak case, as we list below.

- In the case $F=\varnothing$ it abstractly follows from the Bourgeois–Ekholm–Eliashberg surgery isomorphism [7; 13; 11] together with the construction of a strong smooth n-Calabi–Yau structure on the wrapped Fukaya category due to Ganatra [15, Theorem 1.3; 16, Theorem 1.16].
  Another proof of this fact goes via [3, Corollary 4.16] and the fact that the (nonrelative) Ginzburg algebra is a strong smooth n-Calabi–Yau algebra [20, Theorem 6.3].
- Assume $Q=\bullet$, $F=\varnothing$ and let Λ be a horizontally displaceable Legendrian sphere (not necessarily an unknot) in the contactization of a Liouville domain. Then Legout [24] proved that the Chekanov–Eliashberg dg-algebra of Λ admits a *weak* n-Calabi–Yau structure via an explicit geometric construction.
  If $\Lambda\subset\mathbb{R}^3$ is not necessarily an unknot, Chen proves in forthcoming work that the augmentation category $\mathcal{A}\mathrm{ug}_-(\Lambda;\mathbb{Z}_2)$ defined by Bourgeois and Chantraine [6, Section 3.2] carries a strong proper 2-Calabi–Yau structure. In the terminology of Brav and Dyckerhoff [9] this is a *right* Calabi–Yau structure.
- Assume $Q=F=\bullet$, and let $\Lambda\subset\mathbb{R}^3$ not necessarily an unknot. In forthcoming work by Ma and Sabloff it is proven that the A_∞-functor $\mathcal{A}\mathrm{ug}_+(\Lambda;\mathbb{Z}_2)\to\mathcal{C}(\Lambda;\mathbb{Z}_2)$ that is defined on morphisms as precomposition with the canonical inclusion $C_{-*}(\Omega\Lambda)\hookrightarrow CE^*(\Lambda;C_{-*}(\Omega\Lambda))$ carries a weak relative proper 2-Calabi–Yau structure, where $\mathcal{A}\mathrm{ug}_+$ denotes the augmentation category as defined by Ng, Rutherford, Shende, Sivek and Zaslow [27, Section 4].

- For general subcritical Weinstein manifolds and for general Legendrian submanifolds, Corollary 1.3 is proven in the weak case in forthcoming work by Dimitroglou Rizell and Legout via an explicit geometric construction

Corollary 1.4 is not new, but our proof is. It follows in the general case of stopped Weinstein manifolds admitting a stable polarization (which is always true for $X(Q,F)$) by combining work of Shende and Takeda [29, Theorem 1.14] and Ganatra, Pardon and Shende [18, Theorem 1.4].

1.2 Outline of the proof

The proof of Theorem 1.1 is a generalization of the computation in [3, Section 4.3], where the Chekanov–Eliashberg dg-algebra of the Legendrian attaching link of a plumbing of copies of T^*S^n for $n \geq 3$ was computed, which corresponds to the case $F = \varnothing$. A rough sketch of the proof of Theorem 1.1 is as follows.

(1) Find a 1-dimensional *simplicial decomposition* of the subcritical Weinstein manifold $Z_0(Q)$ in the sense of [3, Section 2.2.4].

(2) Find a singular Legendrian $\boldsymbol{\Lambda}(Q,F)$ relative to the simplicial decomposition of $Z_0(Q)$ whose completion is the singular Legendrian $\Lambda(Q,F)$ (see [3, Definitions 4.1 and 4.2]).

(3) Compute the Chekanov–Eliashberg dg-algebra of each part of $\boldsymbol{\Lambda}(Q,F)$. Each such piece is either a Legendrian unknot with boundary, a Legendrian Hopf link with boundary or two Legendrian unknots with a transverse intersection, with boundary.

(4) Using the local computation and the gluing formula for Chekanov–Eliashberg dg-algebras of singular Legendrians [4, Theorem 2.43] we recover $CE^*(\Lambda(Q,F); Z_0(Q))$ up to quasi-isomorphism.

(5) Write down an explicit chain homotopy equivalence between the resulting Chekanov–Eliashberg dg-algebra and the relative Ginzburg algebra of (Q,F).

1.3 Future directions

Even though the relative Ginzburg algebra is known to not be a strong smooth Calabi–Yau algebra, it is expected to admit a pre-Calabi–Yau structure in the sense of Kontsevich, Takeda and Vlassopoulos [23]. This is consistent with the folklore conjecture that the partially wrapped Fukaya category should admit a pre-Calabi–Yau structure, which is motivated by work of Seidel [28, Section 3.3]. In view of the surgery formula for partially wrapped Fukaya categories [5, Theorem 1.1] we thus have the following.

Conjecture 1.5 *The Chekanov–Eliashberg dg-algebra of a singular Legendrian submanifold in the boundary of a stopped Weinstein manifold admits a geometrically defined pre-Calabi–Yau structure.*

Recently, Ng [26] defined an L_∞-structure on a commutative version of the Chekanov–Eliashberg dg-algebra of a Legendrian in $\mathbb{R}^3$ with loop space coefficients, which we expect to be a shadow of the conjectured pre-Calabi–Yau structure.

Throughout this paper, every quiver comes equipped with the trivial potential. An interesting problem that we hope to return to in the future is realizing the Ginzburg algebra of any (Q, F) equipped with a nontrivial potential as a Chekanov–Eliashberg dg-algebra. The only example of such a relation known to date is due to Li [25, Theorem 1.2].

Acknowledgments

The author is grateful to Georgios Dimitroglou Rizell and Noémie Legout for helpful comments on a draft of this paper and for explaining their work in progress, Merlin Christ for his guidance in the literature on relative Ginzburg algebras, Lenhard Ng for sharing a draft of recent work [26], Joshua Sabloff for sharing his forthcoming results with Jiajie Ma, and Zhenyi Chen for sharing and explaining his forthcoming results. Finally the author thanks the referee for valuable suggestions, improving the exposition of the paper. The author was supported by the Knut and Alice Wallenberg Foundation.

2 Relative Ginzburg algebras

Let $\boldsymbol{k}$ be a field. Let Q be a quiver (with trivial potential), and let $F \subset Q$ be a subquiver. Let $\overline{Q}_n(F)$ denote the graded quiver with vertex set Q_0 and arrow set Q_1 consisting of

- an arrow $g\colon v \to w$ in degree 0 for each $g\colon v \to w$ in Q_1;
- an arrow $g^*\colon w \to v$ in degree $2-n$ for each $g\colon v \to w$ in $Q_1 \smallsetminus F_1$;
- an arrow $h_v\colon v \to v$ in degree $1-n$ for each $v \in Q_0 \smallsetminus F_0$.

We will use the notation $\overline{Q}_n := \overline{Q}_n(\varnothing)$ and

$$[x, y]_v^A := \sum_{A \ni e\colon v \to \bullet} xy - \sum_{A \ni e\colon \bullet \to v} yx,$$

where $A \subset Q_1$ is a subset.

Definition 2.1 (n-dimensional Ginzburg algebra) The *n-dimensional Ginzburg algebra* $\mathcal{G}_n^*(Q)$ of the quiver Q is the path algebra of $\overline{Q}_n$ equipped with the differential d given on arrows by

$$dg = dg^* = 0, \quad dh_v = [g, g^*]_v^{Q_1}, \quad v \in Q_0,$$

and extended to the whole path algebra by linearity and the Leibniz rule.

Definition 2.2 (n-dimensional relative Ginzburg algebra) The *n-dimensional relative Ginzburg algebra* $\mathcal{G}_n^*(Q, F)$ of the tuple (Q, F) is the path algebra of $\overline{Q}_n(F)$ equipped with the differential d given on arrows by

$$dg = dg^* = 0, \quad dh_v = [g, g^*]_v^{Q_1 \smallsetminus F_1}, \quad v \in Q_0 \smallsetminus F_0,$$

and extended to the whole path algebra by linearity and the Leibniz rule.

Lemma 2.3 *There is a map of dg-algebras $G\colon \mathcal{G}^*_{n-1}(F) \to \mathcal{G}^*_n(Q,F)$ that is defined on generators as*

$$g \mapsto g, \quad g^* \mapsto 0, \quad h_v \mapsto [g, g^*]^{Q_1 \smallsetminus F_1}_v$$

and extended to the whole of $\mathcal{G}^*_{n-1}(F)$ by linearity and multiplicativity.

Proof This is immediate from the definitions. □

The following result seems well known to experts and implicitly alluded to in [32, Section 7.2] and [21, Section 8] but we did not find an explicit reference in the literature.

Lemma 2.4 *The dg-algebra map $G\colon \mathcal{G}^*_{n-1}(F) \to \mathcal{G}^*_n(Q,F)$ defined in Lemma 2.3 admits a strong relative smooth n-Calabi–Yau structure.*

Proof It is well known that $\mathcal{G}^*_{n-1}(F) \cong \Pi_n(\boldsymbol{k}F)$ by [20, Theorem 6.3], where $\Pi_n(\boldsymbol{k}F)$ denotes the n-Calabi–Yau completion of the path algebra of F. Yeung defined the relative n-Calabi–Yau completion $\Pi_n(\boldsymbol{k}Q, \boldsymbol{k}F)$ in [33, Section 7], and Wu defined a reduced version $\Pi^{\mathrm{red}}_n(\boldsymbol{k}Q, \boldsymbol{k}F)$ in [32, Section 3.6], which is quasi-isomorphic to $\Pi_n(\boldsymbol{k}Q, \boldsymbol{k}F)$ by [32, Proposition 3.18]. It was proven by Yeung [33, Theorem 7.1], Bozec, Calaque and Scherotzke [8, Corollary 5.24] and Wu [32, Proposition 3.18] that the inclusion $\Pi_n(\boldsymbol{k}F) \hookrightarrow \Pi_n(\boldsymbol{k}Q, \boldsymbol{k}F)$ admits a strong relative smooth n-Calabi–Yau structure.

Thus it suffices to show that $\Pi^{\mathrm{red}}_n(\boldsymbol{k}Q, \boldsymbol{k}F) \cong \mathcal{G}^*_n(Q,F)$, which is well known to experts. In the same way that the n-Calabi–Yau completion in [20] is defined as the tensor algebra of a shift of the inverse dualizing bimodule, the relative n-Calabi–Yau completion is defined as the tensor algebra of a shift of the *relative* inverse dualizing bimodule. An explicit description of the reduced version is described in [32, page 37], and the underlying module is generated by the dual arrows g^* in degree $2-n$ for every $g \in Q_1 \smallsetminus F_1$ and loops h_v in degree $1-n$ for every $v \in Q_0 \smallsetminus F_0$. A similar description for $n=2$ is found in [8, Section 5.3.2] but holds for general n in a similar manner. □

Remark 2.5 The dg-algebra $\Pi_n(\boldsymbol{k}Q, \boldsymbol{k}F)$ was also defined in [21] under the name *relative derived preprojective algebra*.

3 Stopped Weinstein manifold associated to a quiver and a frozen subquiver

3.1 Weinstein manifolds and stops

Let us briefly recall some basic geometric definitions.

Definition 3.1 (Weinstein hypersurface) A *Weinstein hypersurface* is a Weinstein embedding

$$(V^{2n-2}, \lambda|_V) \hookrightarrow (X \smallsetminus \operatorname{Core} X, \lambda)$$

such that the induced map $V \to \partial X$ is an embedding. We will denote Weinstein embeddings by $V \hookrightarrow \partial X$.

Definition 3.2 (Weinstein pair) A *Weinstein pair* is a tuple (X^{2n}, V^{2n-2}) consisting of a Weinstein manifold X and a Weinstein hypersurface $V \hookrightarrow \partial X$.

Definition 3.3 (stop associated to a Weinstein hypersurface) Let (X, V) be a Weinstein pair. The *stop associated to the Weinstein hypersurface* $V \hookrightarrow \partial X$ is the Weinstein cobordism σ_V obtained by gluing the Weinstein cobordism $(V \times D_\varepsilon T^*\Delta^1, \lambda_V + 2x\,dy + y\,dx)$ to $\mathbb{R} \times (V \times \mathbb{R})$ along

$$V \times (-\varepsilon, \varepsilon) \hookrightarrow \partial(\mathbb{R} \times (V \times \mathbb{R})).$$

Definition 3.4 (Weinstein manifold stopped at a Weinstein hypersurface) Let (X, V) be a Weinstein pair. The *Weinstein manifold X stopped at V* is the Weinstein cobordism obtained by gluing σ_V to X along $V \times (-\varepsilon, \varepsilon)$.

Remark 3.5 What we call a "stop" deviates slightly from the original definition in [30]. Our usage of the word stop originates from its usage in [2; 5; 13].

Definition 3.6 (Weinstein manifold stopped at a singular Legendrian) Let X be a Weinstein manifold and $\Lambda \subset \partial X$ a singular Legendrian. We define the *Weinstein manifold X stopped at Λ* to be the Weinstein manifold X stopped at a Weinstein neighborhood $T^*\Lambda$ of Λ.

Definition 3.7 (Legendrian relative to Weinstein pair) Let (X, V) be a Weinstein pair. A Legendrian submanifold relative to V is an embedded Legendrian submanifold-with-boundary $\Lambda \subset \partial X$ such that $\partial\Lambda \subset \partial V$ is Legendrian.

3.2 A simplicial decomposition

Let Q be a quiver and let $F \subset Q$ be a subquiver called the *frozen subquiver*. Recall the definition of the pair $(Z_0(Q), \Lambda(Q, F))$ in Section 1. We pick a Weinstein neighborhood of each frozen component of $\Lambda(Q, F)$ together with a handle decomposition that consists of one Weinstein 0-handle per frozen vertex of the component and one Weinstein 1-handle per frozen edge of the component. Then let $P_0(F)$ denote the union of the subcritical parts of each of these Weinstein neighborhoods. The purpose of this section is to describe a simplicial decomposition of $(Z_0(Q), P_0(F))$ (see [3, Definition 2.15]) and a Legendrian submanifold $\boldsymbol{\Lambda}(Q, F)$ relative to this simplicial decomposition (see [4, Definition 2.38]). The construction closely follows [3, Section 4.3].

Let C be the graph that is obtained by taking the underlying graph of Q, and adding one extra vertex in the middle of each edge (thus splitting the edge into two). For convenience, we consider a partition $C_0 = V \cup F_0 \cup E \cup F_1$ where V corresponds to elements of $Q_0 \setminus F_0$ (nonfrozen vertices) and E corresponds to elements in $Q_1 \setminus F_1$. The set $E \cup F_1$ consists of the newly added vertices (edge vertices); see Figure 2.

A simplicial decomposition of $(Z_0(Q), P_0(F))$ is a tuple $((C, C'), (\boldsymbol{V}, \boldsymbol{V}'))$ with C as above and $C' = F_0 \cup F_1$ (where $F_0 \cup F_1$ is viewed as a subset of the vertex set C_0). We call $e \in C_1$ a *frozen edge* if e is adjacent to any frozen edge vertex. All other edges are called nonfrozen.

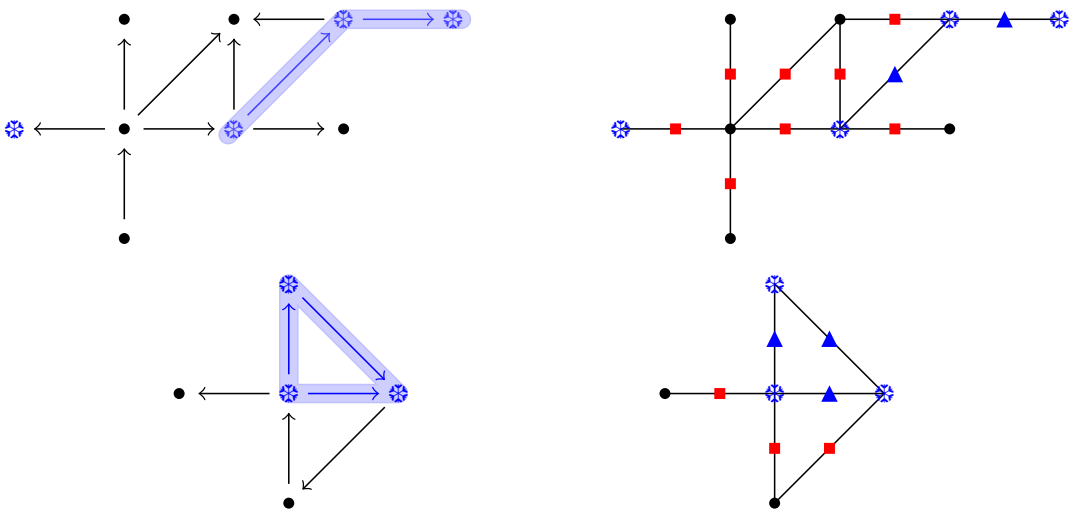


Figure 2: Left: Plumbing quivers Q with a set F_0 of frozen vertices and a set F_1 of frozen edges. Right: The corresponding graph C with vertex set $V \cup F_0 \cup E \cup F_1$. The vertices represented by black dots belong to V, the vertices represented by blue snowflakes belong to F_0, the vertices represented by red squares belong to E and the vertices represented by blue triangles belong to F_1.

Both $\boldsymbol{V}$ and $\boldsymbol{V}'$ are sets consisting of Weinstein manifolds and Weinstein hypersurfaces of varying dimensions. For simplicity we denote by B^{2n}_{std} the Weinstein manifold consisting of B^{2n} equipped with its standard Liouville one-form $\lambda = \frac{1}{2}\sum_{i=1}^{n}(x_i dy_i - y_i dx_i)$. The set $\boldsymbol{V}$ consists of

(1) one copy of $V_e^{2n-2} := B^{2n-2}_{\text{std}}$ for each edge $e \in C_1$;

(2) one copy of $V_v^{2n} := B^{2n}_{\text{std}}$ for each vertex $v \in C_0$;

(3) one Weinstein hypersurface $\iota_v \colon \bigsqcup_{C_1 \ni e \supset v} V_e^{2n-1} \hookrightarrow \partial V_v^{2n}$ for each vertex $v \in C_0$.

The set $\boldsymbol{V}'$ consists of

(1) one copy of $(V')_v^{2n-2} := B^{2n-2}_{\text{std}}$ for each frozen vertex and each frozen edge vertex $v \in F_0 \cup F_1 \subset C_0$;

(2) one copy of $(V')_e^{2n-4} := B^{2n-4}_{\text{std}}$ for each frozen edge $e \in F_1$;

(3) one Weinstein hypersurface $\bigsqcup_{C_1 \ni e \supset v}(V')_e^{2n-4} \hookrightarrow \partial(V')_v^{2n-2}$ for each frozen vertex $v \in F_0$ and frozen edge vertex $v \in F_1$.

We have that $(C, \boldsymbol{V})$ is a simplicial decomposition of $Z_0(Q)$, and $(C', \boldsymbol{V}')$ is a simplicial decomposition of P_0. To complete the construction, we have a hypersurface inclusion $(C', \boldsymbol{V}') \hookrightarrow (C, \boldsymbol{V})$ (see [4, Definition 2.12]) which consists of

(1) the inclusion $C' = F \hookrightarrow C$ of the (underlying graph of the) frozen subquiver into C;

(2) one Weinstein hypersurface $(V')_e^{2n-4} \hookrightarrow V_e^{2n-2}$ for each edge $e \in C_1$;

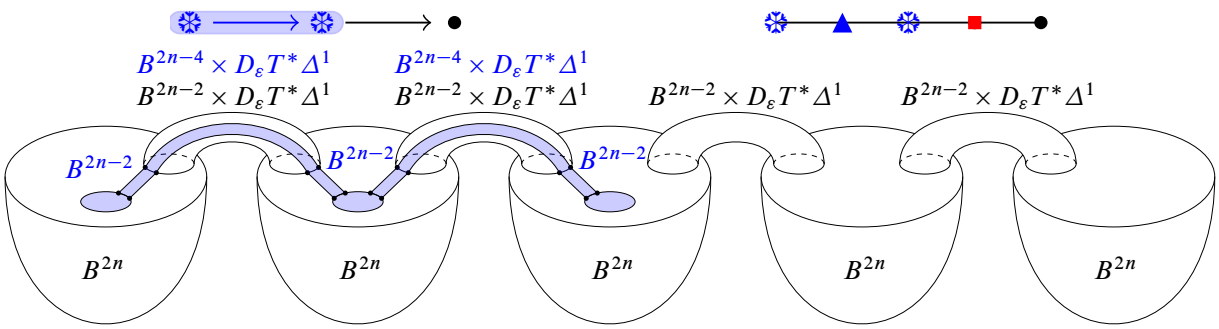


Figure 3: Top left: A quiver with a frozen subquiver (Q, F). Top right: The simplicial complex C associated to (Q, F). Bottom: The simplicial decomposition of the pair $(Z_0(Q), P_0(F))$.

(3) one Weinstein hypersurface

$$(V')_v^{2n-2} \#_{(V')_e^{2n-4}} V_e^{2n-2} \hookrightarrow \partial V_v^{2n}$$

for each pair of a frozen (edge) vertex $v \in F_0 \cup F_1$ and an edge $e \in C_1$ adjacent to v.

An example of the simplicial decomposition of $(Z_0(Q), P_0(F))$ is depicted in Figure 3. We now define a Legendrian submanifold relative to $((C', C), (V, V'))$ denoted by $\Lambda(Q, F)$ (see [4, Definition 2.38]) as follows:

(1) For each nonfrozen edge vertex $v \in E$ let Λ_v be the $(n-1)$-dimensional Hopf link $\Lambda_1 \cup \Lambda_2$ with boundary such that $\partial\Lambda_i \subset \partial V_{e_i}^{2n-2}$ is the $(n-2)$-dimensional Legendrian unknot, where $e_1, e_2 \in C_1$ are the two edges adjacent to v; see Figure 4.

(2) For each frozen edge vertex $v \in F_1$ let Λ_v be two $(n-1)$-dimensional Legendrian disks $\Lambda_1 \cup \Lambda_2$ with boundary in $\partial(V_{e_1}^{2n-2} \#_{(V')_{e_1}^{2n-4}} (V')_v^{2n-2} \#_{(V')_{e_2}^{2n-4}} V_{e_2}^{2n-2})$ such that

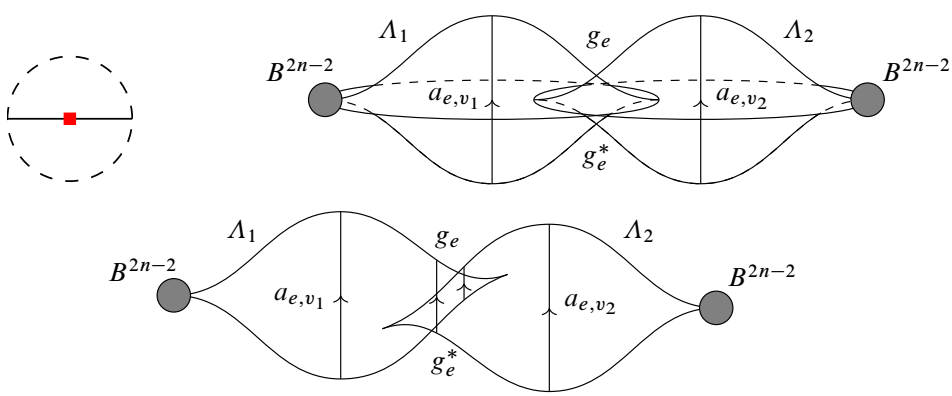


Figure 4: Top left: Local depiction of a vertex $v \in E$. Top right: Two copies of B^{2n-2} with Λ_v which becomes the $(n-1)$-dimensional Hopf link with boundary in $\partial(B^{2n-2} \sqcup B^{2n-2})$ after a small perturbation in a Darboux chart. Bottom: A slice of Λ_v in generic position after a small perturbation, where the Reeb chords g_e and g_e^* are visible.

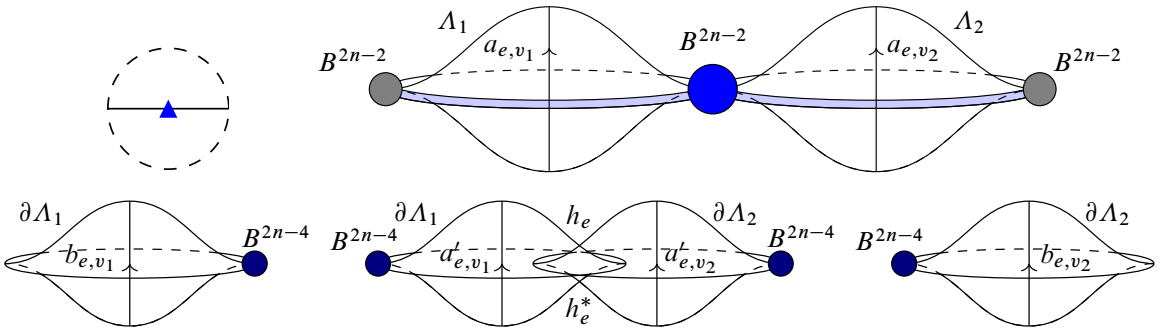


Figure 5: Top left: Local depiction of a vertex $v \in F_1$. Top right: Three copies of B^{2n-2} with Λ_v being two $(n-1)$-dimensional Legendrian disks with boundary in $\partial(B^{2n-2} \#_{B^{2n-4}} B^{2n-2} \#_{B^{2n-4}} B^{2n-2})$ in a Darboux chart. Bottom left: The boundary of Λ_1 in the leftmost copy of B^{2n-2}. Bottom middle: The boundary of $\Lambda_1 \cup \Lambda_2$ in the middle blue copy of B^{2n-2}. Bottom right: The boundary of Λ_2 in the rightmost copy of B^{2n-2}.

(a) $\partial\Lambda_i \cap \partial V_{e_i}^{2n-2} \subset \partial V_{e_i}^{2n-2}$ is the $(n-2)$-dimensional Legendrian unknot with boundary an $(n-3)$-dimensional Legendrian unknot in $\partial(V')_{e_i}^{2n-2}$, where $e_1, e_2 \in C_1$ are the two edges adjacent to v;

(b) $(\partial\Lambda_1 \cup \partial\Lambda_2) \cap \partial(V')_v^{2n-2} \subset \partial(V')_v^{2n-2}$ is the $(n-2)$-dimensional Hopf link with two boundary components, one in each $\partial(V')_{e_i}^{2n-4}$ looking like the $(n-3)$-dimensional unknot, where $e_1, e_2 \in C_1$ are the two edges adjacent to v.

See Figure 5.

(3) For each $v \in V$, let Λ_v be the $(n-1)$-dimensional Legendrian unknot with k open disks removed, where k is the valency of the vertex v such that each of the k components of $\partial\Lambda_v$ is the $(n-2)$-dimensional Legendrian unknot in $\partial V_{e_i}^{2n-2}$ where $e_1, \dots, e_k \in C_1$ are the k edges that are adjacent to v; see Figure 6.

(4) For each $v \in F$, let Λ_v be the $(n-1)$-dimensional Legendrian unknot with one boundary component being an $(n-2)$-dimensional Legendrian unknot in ∂B^{2n-2} for each nonfrozen edge adjacent to v.

The new copy of B^{2n-2} coming from the fact that $v \in F$ (see the definition of V') is connected via a Weinstein 1-handle to every other B^{2n-2} corresponding to frozen edges of C that are adjacent to v. The boundary Λ_v in this new B^{2n-2} is an $(n-2)$-dimensional unknot with some number of disks removed (corresponding to the number of frozen edges adjacent to v); see Figure 7.

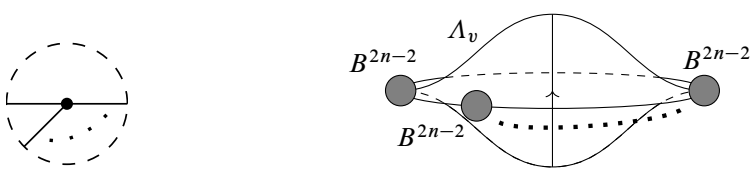


Figure 6: Left: Local depiction of a vertex $v \in V$ with valency k. Right: k copies of B^{2n-2} with Λ_v being the $(n-1)$-dimensional Legendrian unknot with boundary in $\partial\big(\bigsqcup_{i=1}^{k} B^{2n-2}\big)$ in a Darboux chart.

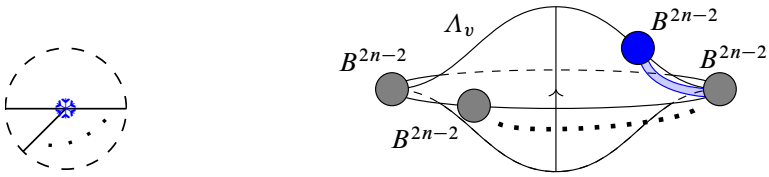


Figure 7: Left: Local depiction of a frozen vertex $v \in F$ with valency k. Right: $k+1$ copies of B^{2n-2} with Λ_v being the $(n-1)$-dimensional Legendrian unknot with boundary in $\partial\big(\bigsqcup_{\text{adjacent nonfrozen edges}} B^{2n-2} \sqcup (B^{2n-2} \#_{B^{2n-4}} B^{2n-2} \#_{B^{2n-4}} \cdots \#_{B^{2n-4}} B^{2n-2})\big)$ in a Darboux chart.

The subcritical part $Z_0(Q)$ is now constructed from (C, V) via gluing as described in [3, Section 1.3], and similarly $(Z_0(Q), P_0(F))$ is constructed from $((C, C'), (V, V'))$ via gluing as described in [4, Section 2.1].

4 Proof of the main theorem

In Section 3.2 we defined a simplicial decomposition of the subcritical Weinstein pair $(Z_0(Q), P_0(F))$, and a Legendrian $\boldsymbol{\Lambda}(Q, F)$ relative to the simplicial decomposition. In this section we compute the Chekanov–Eliashberg dg-algebra of $\boldsymbol{\Lambda}(Q, F)$, which by [4, Definition 2.41] is the Chekanov–Eliashberg dg-algebra of the singular Legendrian $\Lambda(Q, F) \subset \partial Z_0(Q)$.

4.1 Computation of the Chekanov–Eliashberg dg-algebra

We denote by $CE^*(\Lambda; Z)$ the Chekanov–Eliashberg dg-algebra of Λ computed in the boundary of the Weinstein manifold Z.

Lemma 4.1 *Let X_0 be a subcritical stopped Weinstein manifold and $\Lambda \subset \partial X_0$ a Legendrian submanifold such that Λ is contained in an arbitrarily small Darboux chart $D \subset \partial X_0$. Then there is a quasi-isomorphism $CE^*(\Lambda; \partial X_0) \cong CE^*(\Lambda; D)$.*

Proof This follows from a well-known energy argument and follows the same idea as in the proof of [3, Lemma 2.37] which we give an outline of here. Without loss of generality we may assume that the Darboux chart D is disjoint from every subcritical Weinstein handle and stop for dimension reasons.

Let $A, \varepsilon \in \mathbb{R}_{>0}$. Denote by X_0^ε the subcritical stopped Weinstein manifold X_0 where all its subcritical handles and stops have attaching regions of size at most ε (see [3, Definition 2.11]). We call any Reeb chord of Λ that leaves D external, and all other Reeb chords internal. For any external Reeb chord c of Λ of length $\ell(c) < A$ we can find ε small enough that c is disjoint from every subcritical Weinstein handle and stop of X_0^ε (cf [3, Lemma 2.24]). Finally we have that J-holomorphic disks in $\mathbb{R} \times \partial X_0^\varepsilon$ asymptotic to an internal Reeb chord at infinity do not escape $\mathbb{R} \times D$ which follows from a monotonicity argument from which we conclude the result; see [14, Lemmas 5.9 and 5.10]. □

The main tool we use to compute the Chekanov–Eliashberg dg-algebra of $\boldsymbol{\Lambda}(Q,F)$ is the following gluing formula.

Lemma 4.2 *There is a quasi-isomorphism of dg-algebras*

$$CE^*(\boldsymbol{\Lambda}(Q,F);(Z_0(Q),P_0(F)))\cong \operatorname*{colim}_{x\in C_0\cup C_1}\mathcal{A}_x =: \mathcal{A},$$

where $\mathcal{A}_x$ is a dg-algebra generated by the Reeb chords of Λ_x that are completely contained in a Darboux chart, where $\Lambda_x := \partial\Lambda_v$ for $x\in C_1$ and $v\in C_0$ is any vertex adjacent to x (see the definition of $\boldsymbol{\Lambda}(Q,F)$ in Section 3.2). The differential on $\mathcal{A}_x$ is the one induced from $CE^*(\boldsymbol{\Lambda}(Q,F);(Z_0(Q),P_0(F)))$.

Proof This follows from [4, Theorem 2.43], which applies in our situation because $P_0(F)$ is subcritical (cf [3, Remark 4.7]), together with Lemma 4.1. As remarked in the proof of [4, Theorem 2.43], the proof is almost identical to that of [3, Theorem 2.33]. □

Remark 4.3 The colimit in Lemma 4.2 should be interpreted as follows: for each $x\in C_0\cup C_1$ we have a dg-algebra $\mathcal{A}_x$ and for each vertex inclusion $C_0\ni v\subset e\in C_1$ there is an inclusion of dg-algebras $\mathcal{A}_e\subset\mathcal{A}_v$, and the colimit is taken of this diagram of dg-algebras.

All dg-algebras are by definition semifree and hence the colimit is again a semifree dg-algebra that is generated by the union of the generators. See [3, Remark 1.3 and Section 2.5] for further details regarding coefficient rings and in which category the colimit is taken.

Remark 4.4 If Q is a tree, we can by a Legendrian isotopy fit the entire $\Lambda(Q,F)$ inside a single arbitrarily small Darboux ball in $\partial Z_0(Q)\cong S^{2n-1}$. In this case the Chekanov–Eliashberg dg-algebra is computable directly without employing the gluing formula in Lemma 4.2.

We now compute each $\mathcal{A}_x$ appearing in the colimit in Lemma 4.2. The computation for $v\in V\cup E$ is identical to the computation in [3, Section 4.3.2].

For $v\in V$ The dg-subalgebra $\mathcal{A}_v$ for $v\in V$ is generated by a_v in degree $1-n$ and $b_{e,v}$ in degree $2-n$ for each $e\in C_1$ adjacent to v. The differential on the generators is given by

$$\partial a_v=\sum_{e\supset v} b_{e,v},\qquad \partial b_{e,v}=0,$$

where the sum is taken over all edges adjacent to v. The Morse flow trees counted in ∂a_v are depicted in Figure 9(A).

For $v\in F_0$ The dg-subalgebra $\mathcal{A}_v$ for $v\in F_0$ is generated by a_v in degree $1-n$, b_v in degree $2-n$, $b_{e,v}$ in degree $2-n$ for each $e\in C_1$, and $d_{e,v}$ in degree $3-n$ for each frozen $e\in C_1$ that is adjacent to v. The differential on the generators is given by

$$\partial a_v=-b_v+\sum_{e\supset v} b_{e,v},\qquad \partial b_v=\sum_{\substack{e\supset v\\ \text{frozen}}} d_{e,v},\qquad \partial b_{e,v}=\begin{cases} d_{e,v} & \text{if } e \text{ is frozen,}\\ 0 & \text{if } e \text{ is nonfrozen,}\end{cases}\qquad \partial d_{e,v}=0,$$

where the sum is taken over all edges adjacent to v. The appearance of the extra term b_v and the generators $d_{e,v}$ in the differential comes from the extra unknot boundary component of Λ_v that lives in the boundary of $B^{2n-2}\#_{B^{2n-4}}\cdots\#_{B^{2n-4}}B^{2n-2}$ (see Figure 7). The Morse flow trees counted in ∂a_v are depicted in Figure 9(B) and (C). The Morse flow trees counted in ∂b_v and $\partial b_{e,v}$ are similar to those in Figure 9(A) except that they appear in the boundary of the corresponding gray or blue copy of B^{2n-2}.

For $e \in E$ The dg-subalgebra $\mathcal{A}_e$ for $e \in E$ is generated by a_{e,v_1} and a_{e,v_2} both in degree $1-n$, b_{e,v_1} and b_{e,v_2} both in degree $2-n$, g_e in degree 0 and g_e^* in degree $2-n$. By construction $e \in E$ corresponds to an arrow e in Q, and we let $v_1, v_2 \in C_0$ denote the two vertices adjacent to e. The differential on the generators is given by

$$\partial a_{e,v_1} = \begin{cases} b_{e,v_1} - g_e g_e^* & \text{if } e\colon v_1 \to v_2, \\ b_{e,v_1} + g_e^* g_e & \text{if } e\colon v_2 \to v_1, \end{cases} \qquad \partial a_{e,v_2} = \begin{cases} b_{e,v_2} + g_e^* g_e & \text{if } e\colon v_1 \to v_2, \\ b_{e,v_2} - g_e g_e^* & \text{if } e\colon v_2 \to v_1, \end{cases}$$
$$\partial b_{e,v_1} = \partial b_{e,v_2} = \partial g_e = \partial g_e^* = 0.$$

The Morse flow trees counted by $\partial a_{e,v_1}$ and $\partial a_{e,v_1}$ are depicted in Figure 9(D)–(G).

For $e \in F_1$ The dg-subalgebra $\mathcal{A}_e$ for $e \in F_1$ is generated by the set

$$\{a_{e,v_1}, a_{e,v_2}, c, b_{e,v_1}, b_{e,v_2}, d_{e,v_1}, d_{e,v_2}, a'_{e,v_1}, a'_{e,v_2}, h_e^*, h_e\}$$

in degrees

$$|a_{e,v_i}| = 1-n, \quad |c| = |b_{e,v_i}| = |a'_{e,v_i}| = 2-n, \quad |h_e^*| = |d_{e,v_i}| = 3-n, \quad |h_e| = 0,$$

where by construction $e \in F_1$ corresponds to a frozen arrow e in Q and we let $v_1, v_2 \in C_0$ denote the two vertices adjacent to e. The differential on the generators is given by

$$\partial a_{e,v_1} = \begin{cases} b_{e,v_1} - h_e c - a'_{e,v_1} & \text{if } e\colon v_1 \to v_2, \\ b_{e,v_1} + c h_e - a'_{e,v_1} & \text{if } e\colon v_2 \to v_1, \end{cases} \qquad \partial a_{e,v_2} = \begin{cases} b_{e,v_2} + c h_e - a'_{e,v_2} & \text{if } e\colon v_1 \to v_2, \\ b_{e,v_2} - h_e c - a'_{e,v_2} & \text{if } e\colon v_2 \to v_1, \end{cases}$$
$$\partial a'_{e,v_1} = \begin{cases} d_{e,v_1} - h_e h_e^* & \text{if } e\colon v_1 \to v_2, \\ d_{e,v_1} + h_e^* h_e & \text{if } e\colon v_2 \to v_1, \end{cases} \qquad \partial a'_{e,v_2} = \begin{cases} d_{e,v_2} + h_e^* h_e & \text{if } e\colon v_1 \to v_2, \\ d_{e,v_2} - h_e h_e^* & \text{if } e\colon v_2 \to v_1, \end{cases}$$
$$\partial c = h_e^*, \qquad \partial b_{e,v_1} = \partial b_{e,v_2} = \partial h_e = \partial h_e^* = 0.$$

This dg-algebra is computed as follows. The more nondegenerate front projection of $\Lambda_1 \cup \Lambda_2$ is shown in Figure 8 and it is the high dimensional version of the singular Legendrian A_2-bouquet exhibited in [1, Section 8.4.2]. The Reeb chord c goes from Λ_2 to Λ_1 if e is oriented as $v_1 \to v_2$ in Q, and comes from the fact that the boundary of $\Lambda_1 \cup \Lambda_2$ in the blue copy of B^{2n-2} is an $(n-2)$-dimensional standard Hopf link (see Figure 5). The generators d_{e,v_i} are those of the corresponding $(n-3)$-dimensional unknot in the boundary of the B^{2n-4}. The generators a'_{e,v_1}, a'_{e,v_2}, h_e^* and h_e are the generators of the standard $(n-2)$-dimensional standard Hopf link in the boundary of the blue copy of B^{2n-2}.

The Morse flow trees counted in $\partial a_{e,v_1}$ and $\partial a_{e,v_2}$ are depicted in Figure 9(H)–(M), and those counted in ∂c are depicted in Figure 9(N). The Morse flow trees counted in $\partial a'_{e,v_1}$ and $\partial a'_{e,v_2}$ are similar to those depicted in Figure 9(D)–(G) except that they appear in the boundary of the blue copy of B^{2n-2}.

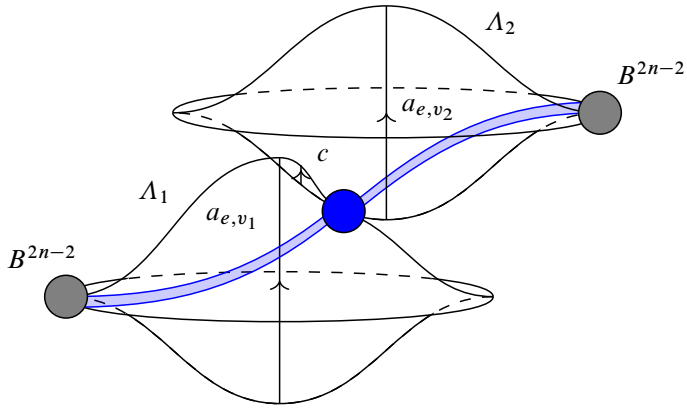


Figure 8: The more accurate front projection of the Legendrian $\Lambda_1 \cup \Lambda_2$ with boundary used to compute $\mathcal{A}_e$ for $e \in F_1$.

Remark 4.5 We make a few comments regarding gradings and signs in the above computation of $\mathcal{A}_x$.

(1) We use the cohomological grading convention in $\mathcal{A}_x$. The degree of a Reeb chord is given by the negative of its Conley–Zehnder index plus one, and the differential increases degree by one. This grading is the negative of the usual homological grading used in eg [10; 12; 14]. The main reason for this choice is to ensure that the statement of Corollary 1.4 makes sense, since the so-called surgery formula [5, Theorem 1.1; 7, Theorem 5.8; 13, Theorem 83] holds with this choice of grading.

(2) The Maslov class of every Legendrian involved vanishes, and to define the degree of the generators corresponding to Reeb chords between two different components we have made an implicit choice of basepoint on each component. We have further chosen the same Maslov potential on every component.

(3) The signs appearing in $\mathcal{A}_x$ a priori depend on a number of choices, all of which we have left implicit. We however note that there is a unique spin structure, decreasing the number of choices by one.

Let us compare the signs appearing in $\partial a_{e,v_1}$ and $\partial a_{e,v_2}$ for a nonfrozen edge $e\colon v_1 \to v_2$. Having fixed choices so that $\partial a_{e,v_2} = b_{e,v_1} + g_e g_e^*$, the sign in front of $g_e^* g_e$ (which is denoted by ν_{end} in [19, Theorem 1.1]) only depends on the order of the generators and differs from the sign in front of $g_e g_e^*$ in $\partial a_{e,v_2}$. The same reasoning applies to $\partial a'_{e,v_1}$ and $\partial a'_{e,v_2}$ for $e \in F_1$. This is also the reason for the sign difference between the terms $h_e c$ and $c h_e$ present in $\partial a_{e,v_1}$ and $\partial a_{e,v_2}$. The sign on front of a'_{e,v_i} in $\partial a_{e,v_i}$ is determined by demanding $\partial^2 = 0$, using the assumption that we have equipped the moduli spaces of holomorphic curves with coherent orientations. This is also how the signs in ∂a_v for $v \in F_0$ and ∂c are determined.

Since we have made no explicit choices and only computed relative sign differences, the actual Chekanov–Eliashberg dg-algebra may differ from $\mathcal{A}_x$ by a composition of automorphisms $z \mapsto -z$ for some generator $z \in \mathcal{A}_x$. The isomorphism class is however unchanged, and the relative signs suffices for our purposes.

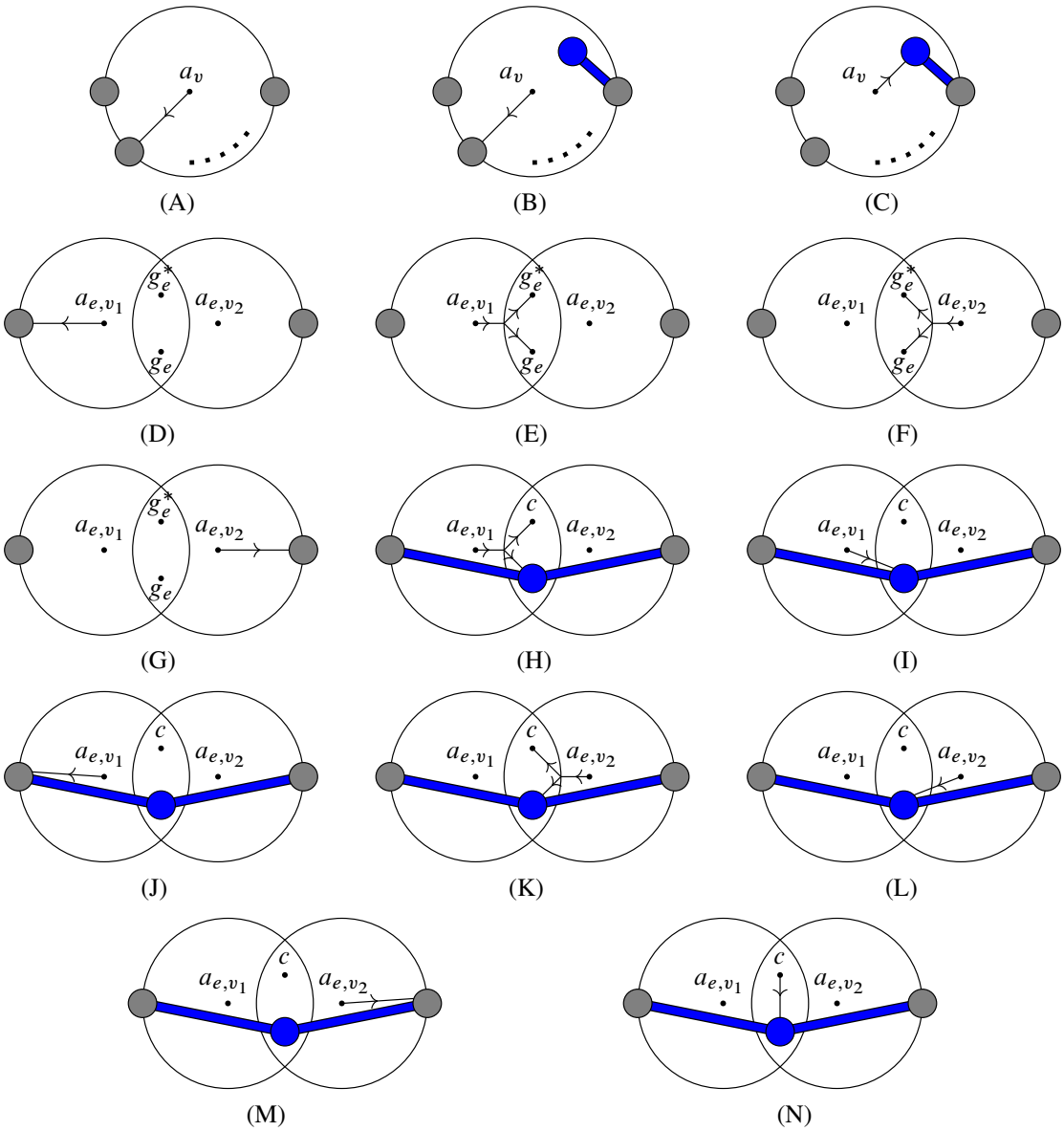


Figure 9: Morse flow trees describing the differential of $\mathcal{A}_x$.

Proof of Theorem 1.1 By Lemma 4.2 it suffices to prove that $\mathcal{A}$ is quasi-isomorphic to the relative Ginzburg algebra $\mathcal{G}_n^*(Q,F)$ (see Definition 2.2). Consider the map $\tau\colon \mathcal{A} \to \mathcal{G}_n^*(Q,F)$ defined on generators as

$$\tau(b_{e,v}) = \begin{cases} g_e g_e^* & \text{if } e\colon v \to \bullet, \\ -g_e^* g_e & \text{if } e\colon \bullet \to v, \end{cases} \quad \text{for } e \text{ nonfrozen}, \qquad \tau(b_v) = [g_e, g_e^*]_v^{Q_1 \smallsetminus F_1} \quad \text{for } v \text{ frozen}$$

$$\tau(a_v) = a_v \quad \text{for } v \text{ nonfrozen}, \qquad \tau(g_e) = g_e, \qquad \tau(g_e^*) = g_e^*, \qquad \tau(h_e) = h_e,$$

$$\tau(x) = 0 \quad \text{for all other generators},$$

and extended to the whole of $\mathcal{A}$ by linearity and multiplicativity. Furthermore define $\varphi\colon \mathcal{G}_n^*(Q,F)\to\mathcal{A}$ on generators by

$$\varphi(a_v)=a_v-\sum_{e\supset v}a_{e,v}\quad \text{for } v \text{ nonfrozen},\qquad \varphi(x)=x\text{for all other generators},$$

and extended to the whole of $\mathcal{G}_n^*(Q,F)$ by linearity and multiplicativity. Finally define $K\colon\mathcal{A}\to\mathcal{A}$ on generators by

$$K(b_{e,v})=a_{e,v},\qquad K(d_{e,v})=\begin{cases}h_e c+a'_{e,v} & \text{if } e\colon v\to\bullet,\\ -ch_e+a'_{e,v} & \text{if } e\colon \bullet\to v,\end{cases}\qquad K(b_v)=-a_v+\sum_{e\supset v}a_{e,v}$$
$$K(h_e^*)=c,\qquad K(x)=0\quad \text{for all other generators},$$

and extended to the whole of $\mathcal{A}$ by linearity and as an $(\mathrm{id},\varphi\tau)$-derivation. By an easy but tedious check we have that τ and φ are chain maps and K is a chain homotopy $\varphi\tau\simeq\mathrm{id}$. Applying the gluing formula in Lemma 4.2 finishes the first part of the proof.

By [5, Remark 2.8] there is a canonical dg-subalgebra $\mathcal{B}\subset CE^*(\Lambda(Q,F);Z_0(Q))$ that is quasi-isomorphic to the dg-subalgebra generated by the boundary $\partial\overline{\Lambda}(Q,F)\subset\partial P_0(F)$. By the computation of the dg-algebra $\mathcal{A}$ we have that $\mathcal{B}$ is generated by $\{b_v,d_{e,v},a'_{e,v},h_e,h_e^*\}$. By construction, the boundary $\partial\overline{\Lambda}(Q,F)\subset\partial P_0(F)$ is a smooth Legendrian unknot link such that when attaching a top Weinstein handle to each component of $\partial\overline{\Lambda}(Q,F)$ and calling the result $P(F)$, we have that $P(F)$ is a disjoint collection of plumbings of copies of T^*S^{n-1} according to the quiver F. Repeating the above argument with a trivial frozen subquiver with $Q=F$ (or appealing to [3, Corollary 4.16]) yields the result. To spell it out explicitly, we define maps $\sigma\colon\mathcal{B}\to\mathcal{G}_{n-1}^*(F)$, $\varepsilon\colon\mathcal{G}_{n-1}^*(F)\to\mathcal{B}$ and $J\colon\mathcal{B}\to\mathcal{B}$ on generators as

$$\sigma(d_{e,v})=\begin{cases}h_e h_e^* & \text{if } e\colon v\to\bullet,\\ -h_e^*h_e & \text{if } e\colon\bullet\to v,\end{cases}\qquad \sigma(a'_{e,v})=0,\qquad \sigma(x)=x\quad\text{for all other generators},$$
$$\varepsilon(b_v)=b_v-\sum_{e\supset v}a'_{e,v},\qquad \varepsilon(x)=0\quad\text{for all other generators},$$
$$J(d_{e,v})=a'_{e,v},\qquad J(x)=0\quad\text{for all other generators}.$$

We extend σ and ε to all of $\mathcal{B}$ and $\mathcal{G}_{n-1}^*(F)$ by linearity and multiplicativity, and extend J to all of $\mathcal{B}$ by linearity and as an $(\mathrm{id},\varepsilon\sigma)$-derivation. As above, one can check that σ and ε are chain maps and J specifies a chain homotopy equivalence $\varepsilon\sigma\simeq\mathrm{id}$.

Finally a straightforward check shows that the diagram

$$\begin{array}{ccc}\mathcal{B} & \xhookrightarrow{\text{incl.}} & CE^*(\Lambda(Q,F);Z_0(Q))\\ \downarrow{\scriptstyle\sigma} & & \downarrow{\scriptstyle\tau}\\ \mathcal{G}_{n-1}^*(F) & \xrightarrow{G} & \mathcal{G}_n^*(Q,F)\end{array}$$

commutes, where $G\colon\mathcal{G}_{n-1}^*(F)\to\mathcal{G}_n^*(Q,F)$ is the map defined in Lemma 2.3. □

References

[1] **B H An**, **Y Bae**, *A Chekanov–Eliashberg algebra for Legendrian graphs*, J. Topol. 13 (2020) 777–869 MR Zbl

[2] **J Asplund**, *Fiber Floer cohomology and conormal stops*, J. Symplectic Geom. 19 (2021) 777–864 MR Zbl

[3] **J Asplund**, *Simplicial descent for Chekanov–Eliashberg dg-algebras*, J. Topol. 16 (2023) 489–541 MR Zbl

[4] **J Asplund**, *Tangle contact homology*, Internat. J. Math. 35 (2024) art. id. 2450044 MR Zbl

[5] **J Asplund**, **T Ekholm**, *Chekanov–Eliashberg dg-algebras for singular Legendrians*, J. Symplectic Geom. 20 (2022) 509–559 MR Zbl

[6] **F Bourgeois**, **B Chantraine**, *Bilinearized Legendrian contact homology and the augmentation category*, J. Symplectic Geom. 12 (2014) 553–583 MR Zbl

[7] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR Zbl

[8] **T Bozec**, **D Calaque**, **S Scherotzke**, *Relative critical loci and quiver moduli*, Ann. Sci. École Norm. Sup. 57 (2024) 553–614 MR Zbl

[9] **C Brav**, **T Dyckerhoff**, *Relative Calabi–Yau structures*, Compos. Math. 155 (2019) 372–412 MR Zbl

[10] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR Zbl

[11] **T Ekholm**, *Holomorphic curves for Legendrian surgery*, preprint (2019) arXiv 1906.07228

[12] **T Ekholm**, **J Etnyre**, **M Sullivan**, *The contact homology of Legendrian submanifolds in* $\mathbb{R}^{2n+1}$, J. Differential Geom. 71 (2005) 177–305 MR Zbl

[13] **T Ekholm**, **Y Lekili**, *Duality between Lagrangian and Legendrian invariants*, Geom. Topol. 27 (2023) 2049–2179 MR Zbl

[14] **T Ekholm**, **L Ng**, *Legendrian contact homology in the boundary of a subcritical Weinstein 4-manifold*, J. Differential Geom. 101 (2015) 67–157 MR Zbl

[15] **S Ganatra**, *Symplectic cohomology and duality for the wrapped Fukaya category*, PhD thesis, Massachusetts Institute of Technology (2012)

[16] **S Ganatra**, *Cyclic homology,* S^1*–equivariant Floer cohomology and Calabi–Yau structures*, Geom. Topol. 27 (2023) 3461–3584 MR Zbl

[17] **S Ganatra**, **J Pardon**, **V Shende**, *Covariantly functorial wrapped Floer theory on Liouville sectors*, Publ. Math. Inst. Hautes Études Sci. 131 (2020) 73–200 MR Zbl

[18] **S Ganatra**, **J Pardon**, **V Shende**, *Microlocal Morse theory of wrapped Fukaya categories*, Ann. of Math. 199 (2024) 943–1042 MR Zbl

[19] **C Karlsson**, *To compute orientations of Morse flow trees in Legendrian contact homology*, preprint (2017) arXiv 1704.05156

[20] **B Keller**, *Deformed Calabi–Yau completions*, J. Reine Angew. Math. 654 (2011) 125–180 MR Zbl

[21] **B Keller**, **Y Wang**, *An introduction to relative Calabi–Yau structures*, from "Representations of algebras and related structures", Eur. Math. Soc., Berlin (2023) 279–304 MR Zbl

[22] **M Kontsevich**, **Y Soibelman**, *Notes on* A_∞*-algebras,* A_∞*-categories and non-commutative geometry*, from "Homological mirror symmetry", Lecture Notes in Phys. 757, Springer (2009) 153–219 MR Zbl

[23] **M Kontsevich**, **A Takeda**, **Y Vlassopoulos**, *Pre-Calabi–Yau algebras and topological quantum field theories*, preprint (2021) arXiv 2112.14667

[24] **N Legout**, *Calabi–Yau structure on the Chekanov–Eliashberg algebra of a Legendrian sphere* (2023) arXiv 2304.03014 To appear in Algebr. Geom. Topol.

[25] **Y Li**, *Koszul duality via suspending Lefschetz fibrations*, J. Topol. 12 (2019) 1174–1245 MR Zbl

[26] **L Ng**, *An L-infinity structure for Legendrian contact homology*, preprint (2023) arXiv 2311.14614

[27] **L Ng**, **D Rutherford**, **V Shende**, **S Sivek**, **E Zaslow**, *Augmentations are sheaves*, Geom. Topol. 24 (2020) 2149–2286 MR Zbl

[28] **P Seidel**, *Fukaya A_∞-structures associated to Lefschetz fibrations, I*, J. Symplectic Geom. 10 (2012) 325–388 MR Zbl

[29] **V Shende**, **A Takeda**, *Calabi–Yau structures on topological Fukaya categories*, preprint (2016) arXiv 1605.02721

[30] **Z Sylvan**, *On partially wrapped Fukaya categories*, J. Topol. 12 (2019) 372–441 MR Zbl

[31] **Z Sylvan**, *Orlov and Viterbo functors in partially wrapped Fukaya categories*, preprint (2019) arXiv 1908.02317

[32] **Y Wu**, *Relative cluster categories and Higgs categories*, Adv. Math. 424 (2023) art. id. 109040 MR Zbl

[33] **W-K Yeung**, *Relative Calabi–Yau completions*, preprint (2016) arXiv 1612.06352

Department of Mathematics, Stony Brook University
Stony Brook, NY, United States

johan.asplund@stonybrook.edu

Received: 6 November 2023 Revised: 15 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3755–3774
DOI: 10.2140/agt.2025.25.3755
Published: 1 October 2025

An RBG construction of integral surgery homeomorphisms

QIANHE QIN

We generalize the RBG construction of Manolescu and Piccirillo to produce pairs of knots with the same n-surgery, and investigate the possibility of constructing exotic definite four-manifolds using n-surgery homeomorphisms.

57K10, 57K18, 57K40

1 Introduction

Manolescu and Piccirillo [12] introduced RBG links, a kind of 3-component framed links in S^3 that produce knot pairs with the same 0-surgery. (Similar constructions appeared earlier in Akbulut's work [2; 3].) RBG links are relevant for an approach to constructing exotic definite 4-manifolds. The strategy is to find a knot pair (K, K'), such that $S^3_0(K) \cong S^3_0(K')$, K is H-slice in some 4-manifold W (ie bounds a null-homologous disk in $W \setminus \operatorname{int} B^4$) and K' is not H-slice in W. Then, one can construct a new 4-manifold W' (an exotic copy of W) by carving out a neighborhood of the slice disk bounded by K, and gluing back the trace of 0-surgery on K' using some 0-surgery homeomorphism.

In [12], they focused on a class of RBG links called special RBG links, and experimented on a 6-parameter family of RBG links. They used Rasmussen's s-invariant to obstruct K' from being H-slice in W, and collected several knots K where the usual invariants obstructing H-sliceness vanish. Later on, however, Nakamura showed that these knots K are not slice. He developed a method in [13] to stably relate the traces of K and K', and obstruct K from being H-slice using $s(K') \neq 0$.

In this paper, for $n \in \mathbb{Z}$, we generalize RBG links to $|n|$-RBG links, which can be used to produce knot pairs (K, J) such that $S^3_l(K) \cong S^3_m(J)$ with $l, m \in \{n, -n\}$.

Definition 1.1 An *$|n|$-RBG link* $L = \{(R, r), (B, b), (G, g)\}$ is a 3-component framed link in S^3, with framings $r \in \mathbb{Q}$ and $b, g \in \mathbb{Z}$, together with homeomorphisms

$$\psi_B \colon S^3_{r,g}(R, G) \to S^3 \quad \text{and} \quad \psi_G \colon S^3_{r,b}(R, B) \to S^3,$$

such that $H_1(S^3_{r,b,g}(R, B, G); \mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$.

Remark 1.2 A certain type of $|n|$-RBG links was defined in the context of Legendrian knots by Casals, Etnyre and Kegel [6, Definition 3.19].

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

Theorem 1.3 *Any $|n|$-RBG link $L=\{(R,r),(B,b),(G,g)\}$ has an associated knot pair (K_B, K_G) and a homeomorphism $\phi_L\colon S^3_{f_b}(K_B)\to S^3_{f_g}(K_G)$ with $f_b, f_g\in\{n,-n\}$. Conversely, given a homeomorphism $\phi\colon S^3_l(K)\to S^3_m(J)$ with $l,m\in\{n,-n\}$, there exists an $|n|$-RBG link L_ϕ such that the associated knot pair is (K,J) and $\phi_{L_\phi}=\phi$ up to isotopy.*

As in [12], one can attempt to use n-surgery homeomorphisms to construct exotic 4-manifolds. A knot K is said to be *n-slice* in W, if $K\subset\partial(W\setminus \operatorname{int} B^4)$ bounds a properly embedded disk D with self-intersection number $-n$. If there is another knot K' with $S^3_n(K)\cong S^3_n(K')$, by removing a tubular neighborhood of D and gluing back the trace of n-surgery on K', we obtain a new 4-manifold W' such that K' is n-slice in W'.

For $W=\#^l\,\overline{\mathbb{CP}}^2$, we have that W' is homeomorphic to W. If K' is not n-slice in W, then W' is not diffeomorphic to W. Moreover, there is an adjunction inequality for the Rasmussen's s-invariant for knots which are n-slice in W; this was conjectured by Manolescu, Marengon, Sarkar and Willis [11] and was proved by Ren [18]. Thus, one can try to use the s-invariant and a pair of knots with the same n-surgery to construct an exotic $\#^l\,\overline{\mathbb{CP}}^2$.

1.1 *n*-special RBG links

We define n-special RBG links, for which the associated knot pairs are easier to find diagrammatically and the associated knots have the same n-surgery.

Definition 1.4 A link $L=\{(R,r),(B,b),(G,g)\}$ with linking matrix M_L, is called an *n-special RBG link*, if $b=g=0$, $n=-\det(M_L)$, and there exist link isotopies

$$R\cup B\cong R\cup\mu_R\cong R\cup G,$$

where μ_R is the meridian of R.

As in [12], one can experiment on parametrized families of n-special RBG links to look for knot pairs that share the same n-surgery, such that one of the knots is n-slice in some $\#^l\,\overline{\mathbb{CP}}^2$ and the other is not n-slice in any $\#^m\,\overline{\mathbb{CP}}^2$.

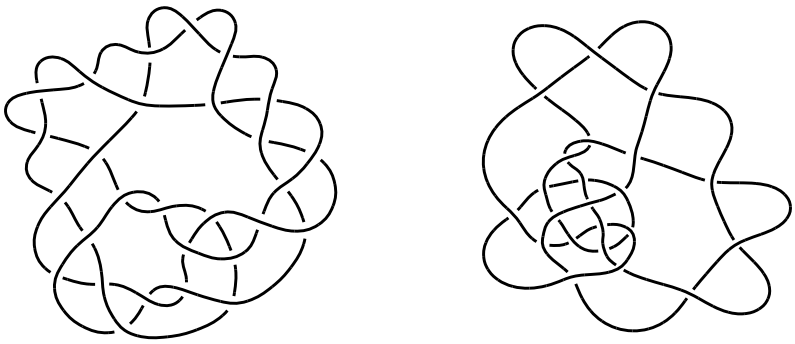
Nakamura [13] showed that for special RBG links such that $R=U$ and K_B is H-slice in W, Rasmussen's s-invariant cannot be used to obstruct K_G from being H-slice. We generalize Nakamura's theorem to n-special RBG links.

Theorem 1.5 *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link with n nonnegative.*

(a) *If R is r-slice in some $\#^m\,\mathbb{CP}^2$ and K_B is n-slice in $\#^l\,\overline{\mathbb{CP}}^2$, then $s(K_G)\le n-\sqrt{n}$.*

(b) *If R is $(r-1)$-slice in some $\#^m\,\overline{\mathbb{CP}}^2$, then $s(K_G)\le n+1-\sqrt{n+1}$.*

If $n>0$, the above theorem leaves open the possibility of using the s-invariant to detect exotic pairs of definite 4-manifolds from n-special RBG links where R is the unknot. For example, we have:



$K(-2,1,3)$ $K(-2,1,2)$

Figure 1

Theorem 1.6 *If the knot $K(-2,1,3)$ from the left-hand side of Figure 1 is 3-slice in some $\#^m\,\overline{\mathbb{CP}}^2$, then there exists an exotic $\#^m\,\overline{\mathbb{CP}}^2$.*

Theorem 1.5 also gives a new way of obstructing knots from being n-slice, by finding another knot with the same n-surgery and checking its s-invariant. (This generalizes the $n=0$ method, which was used by Piccirillo [17] in her proof that the Conway knot is not slice, and then extended by Nakamura [13].) For example, we prove the following theorem.

Theorem 1.7 *The knots in Figure 1 are not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$.*

1.2 n-peculiar RBG links

We will also define a different class of $|n|$-RBG links called *n-peculiar RBG links* (see Section 5), for which the red components R are rationally framed and the associated knot pairs can be obtained diagrammatically.

Definition 1.8 A link $\{(R,r),(B,b),(G,g)\}$ is called an n-peculiar RBG link, if there exists $t\in\mathbb{Z}$ such that

- $R=U$ and B,G are meridians of R,
- $b=g=1/r+1/t$,
- $n=(g+b-2l)-t(l-b)^2$,

where $l=\mathrm{lk}(B,G)$ under an orientation of L such that $\mathrm{lk}(B,R)=\mathrm{lk}(G,R)=1$.

This gives a new construction of RBG links when $n=0$, for which Nakamura's obstruction in [13] does not immediately apply; so, in principle, they can potentially produce exotic 4-spheres.

Organization of the paper In Section 2, we generalize the RBG construction of zero surgeries to integral surgeries. We discuss the construction of a potential exotic pair by cutting and pasting of n-traces. In Section 3, we introduce n-special RBG links, for which the associated knot pairs can be obtained diagrammatically. In Section 4, we generalize Nakamura's sliceness obstruction to an n-sliceness obstruction using n-special RBG links, and we give examples where the n-sliceness of knots in $\#^m \overline{\mathbb{CP}}^2$ is obstructed by this new method. In Section 5, we discuss n-peculiar RBG links.

Acknowledgements The author is grateful to her advisor, Ciprian Manolescu, for his suggestion to work on this project and his continued guidance and support. We would also like to thank Anthony Conway, Nathan Dunfield, Kyle Hayden, Maggie Miller, Lisa Piccirillo, Qiuyu Ren, Ali Naseri Sadr and Charles Stine for helpful suggestions and conversations.

2 $\langle n \rangle$-surgery homeomorphisms

Let n be an integer and $|n|$ be its absolute value. We use $\langle n \rangle$ to denote the set $\{n, -n\}$.

Definition 1.1 An *$|n|$-RBG link* $L = \{(R, r), (B, b), (G, g)\}$ is a 3-component framed link in S^3, with framings $r \in \mathbb{Q}$ and $b, g \in \mathbb{Z}$, together with homeomorphisms

$$\psi_B \colon S^3_{r,g}(R, G) \to S^3 \quad \text{and} \quad \psi_G \colon S^3_{r,b}(R, B) \to S^3,$$

such that $H_1(S^3_{r,b,g}(R, B, G); \mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$.

Remark 2.1 In [12], RBG links are defined to be rationally framed for 0-surgeries. In the case of $\pm n$-surgeries with $n \neq 0$, we restrict the framings b, g of B, G to be integers, so that we can pin down the surgery coefficient to $\pm n$ from the homological condition: $H_1(S^3_{r,b,g}(R, B, G); \mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$.

Definition 2.2 Given a pair of framed knots $\{(K, f_K), (J, f_J)\}$, a homeomorphism $\phi \colon S^3_{f_K}(K) \to S^3_{f_J}(J)$ is called an *$\langle n \rangle$-surgery homeomorphism*, if $f_K, f_J \in \{n, -n\}$. If $f_K = f_J = n$, then we call ϕ an *n-surgery homeomorphism*.

We generalize Theorem 1.2 of [12] to $|n|$-RBG links as in the following theorem, which is a rephrasing of Theorem 1.3 from the introduction.

Theorem 2.3 *Any $|n|$-RBG link $L = \{(R, r), (B, b), (G, g)\}$ has an associated knot pair (K_B, K_G) and an $\langle n \rangle$-surgery homeomorphism $\phi_L \colon S^3_{f_b}(K_B) \to S^3_{f_g}(K_G)$. Conversely, given an $\langle n \rangle$-surgery homeomorphism $\phi \colon S^3_l(K) \to S^3_m(J)$, there exists an $|n|$-RBG link L_ϕ such that the associated knot pair is (K, J) and $\phi_{L_\phi} = \phi$ up to isotopy.*

Proof Given an $|n|$-RBG link $L = \{(R, r), (B, b), (G, g)\}$, we associate to it two framed knots $(K_B, f_b), (K_G, f_g)$ and an $\langle n \rangle$-surgery homeomorphism ϕ_L as follows: First, let (K_B, f_b) be $\psi_B(B, b)$,

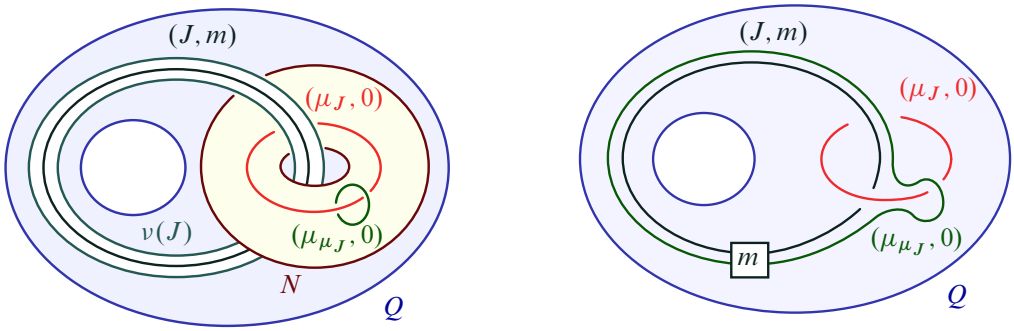


Figure 2

and let (K_G, f_g) be $\psi_G(G, g)$. Since a homeomorphism maps an integer framing to an integer framing, f_b, f_g are integers. Then, extend ψ_B (resp. ψ_G) to $\widetilde{\psi}_B\colon S^3_{r,b,g}(R, B, G) \to S^3_{f_b}(K_B)$ (resp. $\widetilde{\psi}_G\colon S^3_{r,b,g}(R, B, G) \to S^3_{f_g}(K_G)$) by gluing back tubular neighborhoods of B and K_B (resp. G and K_G) according to the framings. Since $H_1(S^3_{r,b,g}(R, B, G); \mathbb{Z}) = \mathbb{Z}/n\mathbb{Z}$, we have $f_b, f_g \in \{n, -n\}$. Finally, define ϕ_L as

$$\phi_L := \widetilde{\psi}_G \circ \widetilde{\psi}_B^{-1}\colon S^3_{f_b}(K_B) \to S^3_{f_g}(K_G).$$

Conversely, given an $\langle n\rangle$-surgery homeomorphism $\phi\colon S^3_l(K) \to S^3_m(J)$, define an $|n|$-RBG link L_ϕ as follows.

Fix setwise representatives of $S^3_l(K), S^3_m(J)$ by specifying the knots $K, J \subset S^3$ and the surgery tubular neighborhoods $\nu(K), \nu(J)$. Pick a meridian $\mu_J \subset S^3 \setminus \nu(J)$. Up to isotopy, we can assume that ϕ^{-1} maps μ_J into $S^3 \setminus \nu(K)$. Choose a tubular neighborhood Q of J which contains μ_J, and pick a tubular neighborhood N of μ_J such that $N \subset Q \setminus \nu(J)$ and $\phi^{-1}(N) \subset S^3 \setminus \nu(K)$. Pick a meridian μ_{μ_J} of μ_J in N. See Figure 2, left. Let L_ϕ be $\{\phi^{-1}(\mu_J, 0), (K, l), (\phi^{-1}(\mu_{\mu_J}), 0)\}$.

Let $\widetilde{N}$ be the manifold obtained by surgery on N along $\{(\mu_J, 0), (\mu_{\mu_J}, 0)\}$, and let $\widetilde{Q}$ be the manifold obtained by surgery on Q along $\{(\mu_J, 0), (J, m), (\mu_{\mu_J}, 0)\}$. Extend $\phi\colon S^3_l(K) \to S^3_m(J)$ to $\widetilde{\phi}\colon S^3_{l,r,0}(K, R, G) \to S^3_{m,0,0}(J, \mu_J, \mu_{\mu_J})$. Let $\widetilde{\psi}_B$ be the slam-dunk map (as in Figure 5.30 in [8]) in $\widetilde{\phi}^{-1}(\widetilde{N})$ and the identity map on $S^3_l(K) \setminus \phi^{-1}(N)$.

Let η be the composition map $\phi \circ \widetilde{\psi}_B \circ (\widetilde{\phi})^{-1}$, which is identity outside of $\widetilde{N}$. Slide μ_{μ_J} over J and cancel the pair (J, μ_J) in $\widetilde{Q}$, as in Figure 2, right. Together, they induce a homeomorphism ψ which is identity on $S^3 \setminus Q$. Let $\widetilde{\psi}_G$ be $\psi \circ \widetilde{\phi}$. Thus, we have the following commutative diagram:

$$\begin{array}{ccccc}
 & & S^3_{r,l,0}(R, B, G) & \xrightarrow{\widetilde{\psi}_B} & S^3_l(K) \\
 & \swarrow^{\widetilde{\psi}_G} & \downarrow{\scriptstyle \widetilde{\phi}} & & \downarrow{\scriptstyle \phi} \\
S^3_m(J) & \xleftarrow[\psi]{} & S^3_{0,m,0}(\mu_J, J, \mu_{\mu_J}) & \xrightarrow{\eta} & S^3_m(J)
\end{array}$$

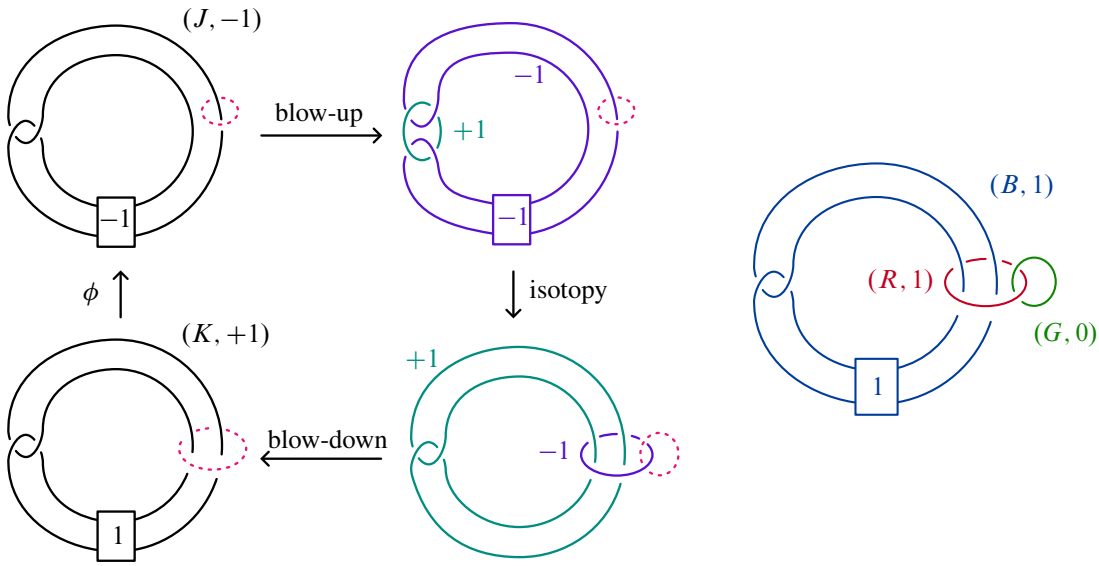


Figure 3

Since $\psi \circ \eta^{-1}$ is identity outside of Q and $\mathrm{MCG}(S^1 \times D^2, S^1 \times S^1)$ is trivial, $\psi \circ \eta^{-1}$ is isotopic to the identity. Therefore, $\widetilde{\psi}_G \circ \widetilde{\psi}_B^{-1}$ is isotopic to ϕ. Undo the surgery on B (resp. G), we obtain ψ_B (resp. ψ_G) from $\widetilde{\psi}_B$ (resp. $\widetilde{\psi}_G$). □

Example 2.4 Consider a homeomorphism ϕ between the 1-surgery on the figure-eight knot K and the (-1)-surgery on the right-handed trefoil J in Figure 3, left, which is an analogue of Figure 23 in [4].

We can construct the corresponding $|1|$-RBG link (Figure 3, right) by chasing the image of $(\mu_J, 0)$ under the map ϕ^{-1}.

For the rest of the paper, we will mostly be concerned with n-surgery homeomorphisms (rather than those that change an n-surgery to a $(-n)$-surgery).

2.1 n-slice knots from n-surgery homeomorphisms

Let X be a smooth, closed, oriented 4-manifold, and let $X^\circ = X \setminus \mathrm{int}(B^4)$. Let K be a knot in $\partial X^\circ \cong S^3$. Suppose K bounds a properly embedded disk D in X°. There exists a tubular neighborhood $\nu(D) \cong D^2 \times D^2$, where D is identified with $D^2 \times \{0\}$. Pick a point $p \in \partial D^2$ and denote $S^1 \times \{p\}$ by K_D. Following Section 2.2 of [13], we make the following definition.

Definition 2.5 The knot $K \subset \partial X^\circ$ is *n-slice*, if $\mathrm{lk}(K, K_D) = n$ in ∂X°.

Denote $X^\circ \setminus \nu(D)$ by $E(D)$ and denote the trace of the n-surgery along K by $X_n(K)$. By the trace embedding lemma [9, Lemma 3.3], if K is n-slice in X, then $-X_n(K)$ is smoothly embedded in X. In particular, we have that $[D] \cdot [D] = -n$.

Now, given an n-surgery homeomorphism $\phi\colon S^3_n(K)\to S^3_n(J)$, we define

$$X_{(D,\phi)} = -X_n(J)\cup_\phi E(D).$$

For $n=0$, if the disk D is not null-homologous, then it is possible that $X_{(D,\phi)}$ is not homeomorphic to X. Note that if X is definite, such as $\#^m\,\mathbb{CP}^2$, every 0-slice disk is null-homologous.

Example 2.6 [10, Example 5.3] Let X be $S^2\times S^2$. Since $S^2\times D^2\cong X_0(U)$, $S^2\times D^2$ is the exterior of some disk in X°. Let $\phi\colon S^3_0(U)\to S^3_0(U)$ be a homeomorphism that maps the 0-framed meridian of U to a 1-framed meridian of U. Then $X_{(D,\phi)}\cong \mathbb{CP}^2\#\overline{\mathbb{CP}}^2$.

However, for $X=\#^m\,\overline{\mathbb{CP}}^2$, we have that $X_{(D,\phi)}$ is homeomorphic to X.

Proposition 2.7 *Let $\phi\colon S^3_n(K)\to S^3_n(J)$ be an n-surgery homeomorphism and D be an n-slice disk bounded by K. If X is simply connected and negative definite with $n\neq 0$, then $X_{(D,\phi)}$ is homeomorphic to X.*

Proof Denote the intersection pairing of a 4-manifold M on $H_2(M;\mathbb{Q})$ by Q_M. Since $n\neq 0$, we have $H_2(S^3_n(K);\mathbb{Q})=H_1(S^3_n(K);\mathbb{Q})=0$. Thus, by the Mayer–Vietoris sequence for X,

$$H_2(X;\mathbb{Q})\cong H_2(E(D);\mathbb{Q})\oplus H_2(-X_n(K);\mathbb{Q}).$$

Similarly, for $X_{(D,\phi)}$, we have $H_2(X_{(D,\phi)};\mathbb{Q})\cong H_2(E(D);\mathbb{Q})\oplus H_2(-X_n(J);\mathbb{Q})$. Thus, Q_X is isomorphic to $Q_{E(D)}\oplus Q_{-X_n(K)}$ over $\mathbb{Q}$, and $Q_{X_{(D,\phi)}}$ is isomorphic to $Q_{E(D)}\oplus Q_{-X_n(J)}$ over $\mathbb{Q}$. Since $Q_{X_n(K)}\cong Q_{X_n(J)}\cong(n)$ and X is negative definite, we have that $X_{(D,\phi)}$ is also negative definite. By Donaldson's theorem, the intersection forms of X and $X_{(D,\phi)}$ are diagonalizable over $\mathbb{Z}$. As in the proof of Lemma 3.3 in [12] for the case $n=0$, $X_{(D,\phi)}$ is simply connected. Hence, by Freedman's theorem $X_{(D,\phi)}$ is homeomorphic to X. □

Proposition 2.8 *The knot J is n-slice in $X_{(D,\phi)}$.*

Proof The knot trace is $X_{-n}(-J)=B^4\cup_{(-J,-n)}\{\text{2-handle}\}$. Remove the B^4 from $X_{(D,\phi)}$, and the core of the 2-handle gives an n-slice disk of J. □

2.2 Extendability over n-traces

For $(X, X_{(D,\phi)})$ to be a potential exotic pair, we need that the n-surgery homeomorphism

$$\phi\colon S^3_n(K)\to S^3_n(J)$$

does not extend smoothly to an n-trace diffeomorphism $\Phi\colon X_n(K)\to X_n(J)$. In some cases, one can see that ϕ actually extends smoothly over n-traces.

Example 2.9 Given an $|n|$-RBG link L such that $(R,r)=(U,0)$, B and G are meridians of R and ψ_B (resp. ψ_G) is induced by sliding B (resp. G) over R and a slam-dunk. Replacing (R,r) by a dotted circle and doing the same diagram calculus, we obtain a diffeomorphism from $X_n(K_B)$ to $X_n(K_G)$, extending ϕ_L. Note that L is an n-special RBG link with $(R,r)=(U,0)$ (see Definition 1.4).

Generalizing Definition 3.12 in [12], we say that an n-surgery homeomorphism $\phi\colon S^3_n(K)\to S^3_n(J)$ has *property U*, if there exists a choice of surgery diagrams of $S_n(K)$ and $S_n(J)$, such that ϕ sends a 0-framed meridian of K to a 0-framed curve γ which appears unknotted in the diagram of $S_n(J)$.

Theorem 2.10 *If ϕ has property U, then there exists a diffeomorphism $\Phi\colon X_n(K)\to X_n(J)$ with $\Phi|_\partial=\phi$.*

Proof This is a generalization of Theorem 3.13 in [12]. □

For example, all links in Example 2.9 have property U.

If the mapping class group of the n-surgeries is trivial, then if there exists some $\phi\colon S^3_n(K)\to S^3_n(J)$ which does not extend over the trace, then the traces $X_n(K)$ and $X_n(J)$ are not diffeomorphic. In general, it is hard to obstruct extensibility smoothly, but we have obstructions for extending homeomorphically over the traces.

Proposition 2.11 *Let $\phi\colon S^3_n(K)\to S^3_n(J)$ be an n-surgery homeomorphism, with $n\neq 0$, which induces $f_*\colon H_1(S^3_n(K))\to H_1(S^3_n(J))$. Then, f extends to a trace homeomorphism $\Phi\colon X_n(K)\to X_n(J)$ if and only if $f_*([\mu_K])=\pm[\mu_J]$.*

Proof If $f_*([\mu_K])=\pm[\mu_J]$, then we can lift f_* to an isometry $\Lambda\colon H_2(X_n(K))\to H_2(X_n(J))$ such that the following diagram commutes:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & H_2(X_n(K)) & \longrightarrow & H_2(X_n(K),S^3_n(K)) & \longrightarrow & H_1(S^3_n(K)) & \longrightarrow & 0\\
 & & \downarrow{\scriptstyle\Lambda} & & {\scriptstyle\Lambda^*}\uparrow & & \downarrow{\scriptstyle f_*} & & \\
0 & \longrightarrow & H_2(X_n(J)) & \longrightarrow & H_2(X_n(K),S^3_n(K)) & \longrightarrow & H_1(S^3_n(K)) & \longrightarrow & 0
\end{array}$$

Since $n\neq 0$, the geometric obstruction $\theta(f,\Lambda)$ vanishes for any morphism (f,Λ). Moreover, the Kirby–Siebenmann invariants $\Delta(X_n(K))\equiv\Delta(X_n(J))\equiv 0 \pmod 2$. Then the result follows from [5, Corollary 0.8(i)]. □

Remark 2.12 There exist homeomorphisms of n-surgeries that do not map meridian to meridian. For instance, consider an n-special RBG link with $l=2$, $r=3$ (see Section 3). One can chase the meridian within the link diagram and construct a homeomorphism f which maps meridian $[\mu_{K_G}]$ to $3[\mu_{K_B}]$. By Proposition 2.11, we have that f is not extensible over the n-trace.

Proposition 2.13 *Let n be an integer such that $\{l \mid l^2 = 1\} = \{\pm 1\} \subset \mathbb{Z}/n\mathbb{Z}$. Every n-surgery homeomorphism extends over traces.*

Proof Consider the linking form Q of M. Since n-traces have intersection form (n), we have $Q([\mu_K],[\mu_K]) = Q([\mu_J],[\mu_J]) = 1/n \pmod{\mathbb{Z}}$. Let $f_*([\mu_K]) = l[\mu_J]$. Since Q is invariant under f_*, we have that $l^2 = 1 \pmod{n}$. The result follows from Proposition 2.11. □

For example, when $n = 1, 2, 4, p^k$ or $2p^k$ with p an odd prime, every n-surgery homeomorphism extends over the traces homeomorphically. For other n, it is possible that $\phi\colon S^3_n(K) \to S^3_n(J)$ does not map meridian to meridian homologically; in such a case, if we also have that the mapping class group of $S^3_n(K)$ is trivial, then we conclude that their n-traces $X_n(K)$ and $X_n(J)$ are not homeomorphic.

3 *n*-special RBG links

Definition 1.4 A link $L = R \cup B \cup G$, with framings $r, b, g \in \mathbb{Z}$, respectively, and a linking matrix M_L, is called an *n-special RBG link*, if

- $b = g = 0$,
- there exist link isotopies $R \cup B \cong R \cup \mu_R \cong R \cup G$,
- $n = -\det(M_L)$.

We obtain a knot diagram of (K_G, f_g) as follows. Choose a diagram of an n-special RBG link L. Isotope L such that $B = \mu_R$. Slide G over R such that G does not intersect some meridian disk Δ_B of R bounded by B. Cancel the pair (B, R). Denote the image of (G, g) by (G', g'). Similarly, we can obtain (K_B, f_b).

Remark 3.1 Since there is only one orientation preserving homeomorphism of S^3 up to isotopy, in the standard (empty) diagram of S^3, the framed knot (K_G, f_g) is isotopic to (G', g') and (K_B, f_b) is isotopic to (B', b').

Notice that isotopies and slides do not change the determinant of the linking matrix, and a slam-dunk changes the sign of the linking matrix. Therefore, the homeomorphism ϕ_L induced by L is an n-surgery homeomorphism.

Let ψ_B (ψ_G) be the homeomorphism induced by a sequence of Kirby moves. Note that an n-special RBG link, together with ψ_B and ψ_G, is an $|n|$-RBG link.

Orient the link L such that $\mathrm{lk}(R, G) = \mathrm{lk}(R, B) = 1$ and denote $\mathrm{lk}(B, G)$ by l. Then

$$n = -\det M_L = l(rl - 2).$$

We can get arbitrary n by setting $l = \pm 1$ and changing r.

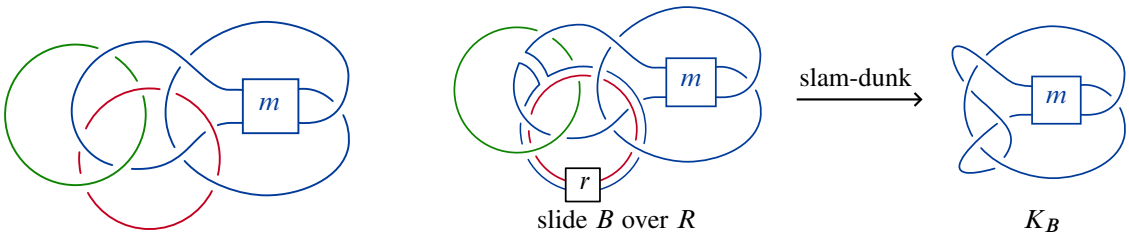


Figure 4: Left: a one parameter family of n-special RBG links. Right: obtaining the diagram of K_B with one slide.

Example 3.2 Consider a family of special RBG links with $l = -1$ in Figure 4, left.

The knot diagram for K_B can be computed by sliding B over R and canceling the pair (G, R) (see Figure 4, right). Similarly, to get K_G, we isotope the link diagram such that B becomes a circle, and let Δ_B be the inner domain bounded by B. Slide G over R along Δ_B, and cancel the pair (B, R) (see Figure 5, left).

If $m = 1$, K_B is the figure-eight knot, and K_G is also the figure-eight knot for any r. For $r = -1$, $m = 0$, we have $S^3_1(K_B) \cong S^3_1(K_G)$. Using SnapPy [7], we identify K_B as 6_2, which is the mirror of 6_2 in the Rolfsen knot table, and K_G as $K13n3596$. This gives an example of small knots which have the same 1-surgery.

Remark 3.3 By Theorem 3.7 in [1], for a knot K with an annulus presentation, one can construct another knot K' via the $(*n)$ operation, such that $S^3_n(K) \cong S^3_n(K')$.

In Example 3.2 with $m = 0$, the knot K_B has an annulus presentation as shown in Figure 5, right. The knot K' obtained by applying $(*(-1))$ operation on the mirror of K_B, is the mirror of the knot K_G in Example 3.2.

There are, however, some cases where an n-special RBG link produces identical knots.

Proposition 3.4 *Let $L = \{(R, r), (B, 0), (G, 0)\}$ be an n-special RBG link. If $(R, r) = (U, 0)$ and R bounds a disk Δ_R such that $|\Delta_R \cap B| = |\Delta_R \cap G| = 1$, then $K_B = K_G$.*



Figure 5: Left: the diagram of K_G. Right: an annulus presentation of K_B in Figure 4, right, with $m = 0$.

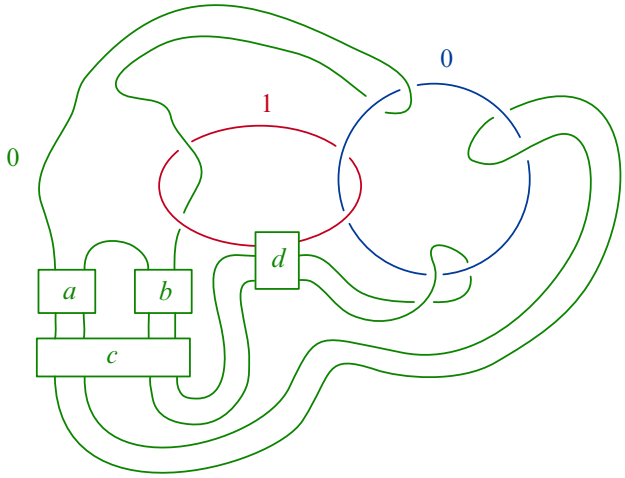


Figure 6

Proof Slide B over G such that the resulting blue component B' does not intersect with Δ_R. Since R has framing 0, we can cancel the pair (R, G). Since these induce a homeomorphism from $S^3_{0,0}(R, G)$ to S^3, the knot K_B is isotopic to B'. Similarly, the knot K_G is obtained by sliding G over B using the same band as above and therefore has the same diagram as K_B. □

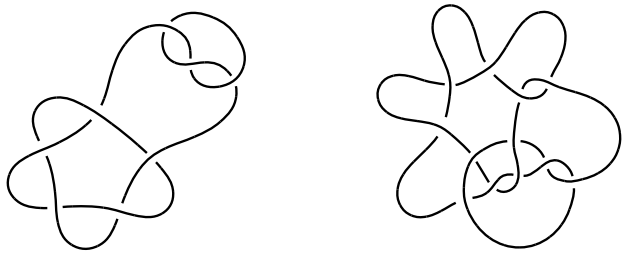
Proposition 3.5 *Let L be an n-special RBG link. If B bounds a properly embedded disk Δ_B such that $|\Delta_B \cap R| = 1$, $|\Delta_B \cap G| < 2$, and if G bounds a properly embedded disk Δ_G such that $|\Delta_G \cap R| = 1$, $|\Delta_G \cap B| \leq 2$, then $K_B = K_G$. (All intersections are required to be transverse.)*

Proof This is a generalization of Proposition 4.11 in [12], which was for $n = 0$. The proof in [12] is independent of the framings of the RBG link. □

Remark 3.6 From Example 3.2, we see that there exists a special RBG link with disks Δ_G, Δ_B, such that $|\Delta_G \cap R| = 1$, $|\Delta_B \cap R| = 1$, and $|\Delta_G \cap B| = 1$, $|\Delta_B \cap G| = 3$, but the associated knots K_B, K_G are not isotopic.

Example 3.7 Consider a family of special RBG links with four twisting boxes as in Figure 6. Since the linking number l between B and G is -1, if $r = 1$, then $n = 3$. Therefore, we obtain a family of 3-special RBG links parametrized by the numbers of twists (a, b, c, d). For each choice of (a, b, c, d), we denote the green knot associated to the link by $K_G(a, b, c, d)$, and the corresponding blue knot by $K_B(a, b, c, d)$.

For example, let $(a, b, c, d) = (-2, 1, -1, -1)$. The knot $K_G(-2, 2, -1, -1)$ is the nonhyperbolic knot $T(-2, 3) \# T(2, 5)$, and the knot $K_B(-2, 2, -1, -1)$ is recognized as $K12n121$ by SnapPy [7] (see Figure 7). Since $K_G(-2, 2, -1, -1)$ and $K_B(-2, 2, -1, -1)$ are generated by a 3-special RBG link, they have the same 3-surgery. Thus, 3 is a noncharacterizing slope for $T(-2, 3) \# T(2, 5)$, recovering an example in [20].



$K_G(-2,2,-1,-1)$ $K_B(-2,2,-1,-1)$

Figure 7

This family also produces knot pairs with small crossing numbers (summing up to 20). When $(a,b,c,d) = (1,2,-1,-1)$, we have that $S^3_3(6_3) \cong S^3_3(K14n15962)$. For $(a,b,c,d) = (0,1,-1,-1)$, we obtain that $S^3_3(6_2) \cong S^3_3(K14n10164)$, which can also be obtained from Example 3.2 with $r = 1$, $m = 0$. Finally, when $(a,b,c,d) = (-1,2,-1,-1)$, the link generates the knot pair $(K_G, K_B) = (10_{125}, 10_{132})$, whose 3-surgeries are isometric hyperbolic 3-manifolds. All of the knot pairs in this example have the property that $s(K_B) = s(K_G)$.

4 Potential exotica

In this section, we follow the recipe given in Section 2.1 and look for exotica using n-special links and the Rasmussen's s-invariant. More specifically, we can use n-special RBG links to find knot pairs (K,J) that share the same n-surgery. If K is n-slice in $\#^m\,\overline{\mathbb{CP}}^2$ and J is not n-slice in $\#^m\,\overline{\mathbb{CP}}^2$ for some m, then we can build an exotic negative-definite 4-manifold (see Proposition 2.7). We first discuss the use of the s-invariant in obstructing a knot from being n-slice and then present an example which generates potential exotica.

4.1 n-special RBG link and the s-invariant

We generalize the results of Nakamura [13] to n-special RBG links.

Proposition 4.1 (analogue of Lemma 3.1 in [13]) *Let $L = \{(R,r),(B,0),(G,0)\}$ be an n-special RBG link. If R is r-slice in a smooth oriented closed 4-manifold W, then*

$$X_n(K_B)\#W \cong X_n(K_G)\#W.$$

Proof Let Z be the 4-manifold obtained by attaching two 2-handles to $W^\circ \setminus \nu(D)$ along $\{(B,0),(G,0)\}$, where D is an r-slice disk of R. Since $(B,0)$ is isotopic to $(\mu_R,0)$, the 2-handle attached along $(B,0)$ fills $\nu(D)$. Slide G over R such that G does not intersect some meridian disk Δ_B bounded by B. This changes $(G,0)$ to (K_G,n) and induces $Z \cong W^\circ \cup_{(K_G,n)} 2h \cong X_n(K_G)\#W$. Similarly, $Z \cong X_n(K_B)\#W$, and therefore $X_n(K_B)\#W \cong X_n(K_G)\#W$. □

Corollary 4.2 (analogue of Corollary 3.2 in [13]) *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link, such that R is r-slice in W. If K_B is n-slice in X, then K_G is n-slice in $X\#-W$.*

Proof If K_B is n-slice in X, then $-X_n(K_B)$ smoothly embeds in X, which implies that $-X_n(K_B)\#-W$ smoothly embeds in $X\#-W$. By Proposition 4.1, $-X_n(K_G)\#-W$ also smoothly embeds in $X\#-W$. Therefore, K_G is n-slice in $X\#-W$. □

Proposition 4.3 (analogue of Lemma 3.11 in [13]) *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link. If R is $(r-1)$-slice in some closed 4-manifold W, then K_B, K_G are $(n+1)$-slice in $W\#\overline{\mathbb{CP}}^2$.*

Proof We use $X_{f_K,f_J}(K,J)$ to denote the 4-manifold obtained by attaching two 2-handles along the framed link $\{(K,f_K),(J,f_J)\}$ to a 4-ball.

Consider the slides of B over R which are applied to L in obtaining the knot K_B. These induce a diffeomorphism between $X_{r,n+1}(R,K_B)$ and $X_{r,1}(R,B)$. Next, Slide R over B to separate B and R, and obtain a diffeomorphism $X_{r,1}(R,B)\cong X_{r-1}(R)\#\mathbb{CP}^2$. Since R is $(r-1)$-slice in W, the trace $-X_{r-1}(R)$ smoothly embeds in W, which implies that $-X_{r-1}(R)\#\overline{\mathbb{CP}}^2$ smoothly embeds in $W\#\overline{\mathbb{CP}}^2$. Since $-X_{n+1}(K_B)\subset -X_{r,n+1}(R,K_B)$, the trace $-X_{n+1}(K_B)$ smoothly embeds in $W\#\overline{\mathbb{CP}}^2$, ie K_B is $(n+1)$-slice in $W\#\overline{\mathbb{CP}}^2$. □

From now on, we assume that n is a nonnegative integer.

Theorem 4.4 *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link.*

(a) *If R is r-slice in some $\#^m\mathbb{CP}^2$ and K_B is n-slice in some $\#^l\overline{\mathbb{CP}}^2$, then K_G is n-slice in $\#^{l+m}\overline{\mathbb{CP}}^2$.*

(b) *If R is $(r-1)$-slice in some $\#^m\overline{\mathbb{CP}}^2$, then K_B, K_G are $(n+1)$-slice in $\#^{m+1}\overline{\mathbb{CP}}^2$.*

Proof (a) If R is r-slice in some $\#^m\mathbb{CP}^2$, and K_B is n-slice in some $\#^l\overline{\mathbb{CP}}^2$, then by Corollary 4.2, K_G is n-slice in $\#^{l+m}\overline{\mathbb{CP}}^2$.

(b) If R is $(r-1)$-slice in some $\#^m\overline{\mathbb{CP}}^2$, then by Proposition 4.3, K_B, K_G are $(n+1)$-slice in $\#^{m+1}\overline{\mathbb{CP}}^2$. □

For obstructing n-sliceness in $\#^l\overline{\mathbb{CP}}^2$, we will use the following adjunction inequality for the s-invariant in [18].

Theorem 4.5 [18, Corollary 1.4] *Let $W=\#^l\overline{\mathbb{CP}}^2$. Let $K\subset\partial W^\circ=S^3$ be a knot, and $\Sigma\subset W^\circ$ a properly, smoothly embedded oriented surface with no closed components, such that $\partial\Sigma=K$. Then*

$$s(K)\le 2g(\Sigma)-|[\Sigma]|-[\Sigma]\cdot[\Sigma].$$

Proposition 4.6 *If K is n-slice in some $\#^l\,\overline{\mathbb{CP}}^2$, then $s(K)\le n-\sqrt{n}$. If J is $(-n)$-slice in some $\#^l\,\mathbb{CP}^2$, then $s(J)\ge -n+\sqrt{n}$.*

Proof Let $\{e_1,\dots,e_l\}$ be an orthonormal basis of $H_2(\#^l\,\overline{\mathbb{CP}}^2)$, and $[D]=a_1e_1+a_2e_2+\cdots+a_le_l$ in $H_2(\#^l\,\overline{\mathbb{CP}}^2)$. Since K is n-slice, $-[D]\cdot[D]=a_1^2+\cdots+a_l^2=n$. Since

$$|[D]|^2=(|a_1|+\cdots+|a_l|)^2\ge a_1^2+\cdots+a_l^2=n,$$

we have $s(K)\le n-\sqrt{n}$ by Theorem 4.5. For the second part, consider the mirror of J, which is n-slice in $\#^l\,\overline{\mathbb{CP}}^2$, and obtain that $s(J)\ge -n+\sqrt{n}$. □

Note that a similar inequality holds for the τ-invariant (see [13, Corollary 2.13]).

Theorem 1.5 *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link.*

(a) *If R is r-slice in some $\#^m\,\mathbb{CP}^2$ and K_B is n-slice in some $\#^l\,\overline{\mathbb{CP}}^2$, then $s(K_G)\le n-\sqrt{n}$.*

(b) *If R is $(r-1)$-slice in some $\#^m\,\overline{\mathbb{CP}}^2$, then $s(K_G)\le n+1-\sqrt{n+1}$.*

Proof If R is r-slice in some $\#^m\,\mathbb{CP}^2$ and K_B is n-slice in some $\#^l\,\overline{\mathbb{CP}}^2$, then by Corollary 4.2, K_G is n-slice in $\#^{l+m}\,\overline{\mathbb{CP}}^2$, then by Proposition 4.6, $s(K_G)\le n-\sqrt{n}$.

If R is $(r-1)$-slice in some $\#^m\,\overline{\mathbb{CP}}^2$, then by Proposition 4.3, K_G is $(n+1)$-slice in $\#^{m+1}\,\overline{\mathbb{CP}}^2$. Then, by Proposition 4.6, $s(K_G)\le n+1-\sqrt{n+1}$. □

Given an n-special RBG link L, one can modify the link to L' by decreasing the framing r of R. Notice that the diagrams of K_G can be obtained by adding full twists on parallel strands of the diagram of K'_G of the modified link. We will use the following lemma, which generalize Proposition 7.6(8) in [11], to get more precise bounds of $s(K_G)$.

Lemma 4.7 *Let K be a knot. Let K' be obtained by adding one positive twist to parallel strands of K. If the algebraic intersection between the parallel stands and the twisting disk Δ_T is l, then we have $s(K')\le s(K)-|l|+l^2$.*

Proof Let T be the boundary of Δ_T. We build $W=\overline{\mathbb{CP}}^2\setminus(\operatorname{int}B^4\cup\operatorname{int}B^4)$ by attaching a (-1)-framed 2-handle along T to $S^3\times\{1\}\subset S^3\times I$. Consider the annulus $A=K\times I\subset S^3\times I$. Choose a point $p\in K$ which is not on the twisting disk Δ_T. Pick a tubular neighborhood N of $\{p\}\times I$ in W, such that $N\cap(\Delta_T\times\{1\})=\varnothing$. Let (D,X) be $(A\setminus N, W\setminus N)\cong(D^2,\overline{\mathbb{CP}}^2\setminus\operatorname{int}B^4)$. The boundary of D is $-K\#K'$. Since K intersect Δ algebraically l-times, D represents the class $l[\overline{\mathbb{CP}}^1]$ in $H^2(\overline{\mathbb{CP}}^2)$. Applying Theorem 4.5 to (D,X), we have $s(-K\#K')\le g(D)-|[D]|-[D]\cdot[D]=-|l|+l^2$. Thus, $s(K')\le s(K)-|l|+l^2$. □

Proposition 4.8 *Let $L=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link, such that $R=U$, $r\geq 0$. Suppose there exists a disk Δ_G, such that $\partial\Delta_G=G$, and B, R intersect Δ_G geometrically once. Suppose also that there exists a disk Δ_R, such that $\partial\Delta_R=R$, and B, G intersect Δ_R geometrically once. If K_B is n-slice in some $\#^l\,\overline{\mathbb{CP}}^2$, then $s(K_G)\leq n-\sqrt{n}$.*

Proof Since there exists a disk Δ_G, such that $\partial\Delta_G=G$, and B, R intersect Δ_G geometrically once, the knot K_B can be obtained by sliding B over R once, and therefore the diagram of K_B does not depend on r. Let $L^0=\{(U,0),(B,0),(G,0)\}$. Proposition 3.4 implies that $K_B^0=K_G^0$.

Observe that K_G can be obtained from K_G^0 by adding an r-positive twist box on parallel strands. Since $|\Delta_G\cap B|=1$ implies $|\operatorname{lk}(B,G)|=1$, the parallel strands link the twist box algebraically once. By Lemma 4.7, $s(K_G)\leq s(K_G^0)\leq n-\sqrt{n}$. □

Proposition 4.9 *Let $L[r]=\{(R,r),(B,0),(G,0)\}$ be an n-special RBG link such that $R=U$, $r>1$. Let l be the linking number between B and G. If $l^2\leq n$, then $s(K_G)\leq n-\sqrt{n}$.*

Proof Consider the $(n-l^2)$-special RBG link $L[r-1]=\{(R,r-1),(B,0),(G,0)\}$. Since $r>1$, U is $(r-2)$-slice in some $\#^m\,\overline{\mathbb{CP}}^2$. By Theorem 1.5, $s(K_G[r-1])\leq n-l^2+1-\sqrt{n-l^2+1}$. The knot K_G of $L[r]$ can be obtained by adding one positive full twists along parallel strands from $K_G[r-1]$. Since $\operatorname{lk}(B,G)=l$, the algebraic intersection number between the parallel strands and the twist box is l. By Lemma 4.7, we have that $s(K_G)\leq s(K_G[r-1])-|l|+l^2$. Thus, $s(K_G)\leq n-|l|+1-\sqrt{n-l^2+1}$. If $l^2\leq n$, then $n-|l|+1-\sqrt{n-l^2+1}\leq n-\sqrt{n}$. The result follows. □

4.2 Experiments

Consider an n-special RBG link with $R=U$. If $r\leq 0$, then R is r-slice in $\#^{|r|}\,\mathbb{CP}^2$. By Theorem 1.5, the n-sliceness of K_B implies that $s(K_G)\leq n-\sqrt{n}$. Therefore, we cannot use Proposition 4.6 to obstruct the n-sliceness of K_G. If $r>0$, then R is $(r-1)$-slice in $\#^{r-1}\,\overline{\mathbb{CP}}^2$. By Theorem 1.5, we only have $s(K_G)\leq n+1-\sqrt{n+1}$. Now, pick an integer n such that there exists an even integer $2q$ satisfying $n-\sqrt{n}<2q\leq n+1-\sqrt{n+1}$ (eg $n=3,6,8,11,13,15$ etc). If one can find an n-special RBG link with $s(K_G)=2q$, then since $s(K_G)>n-\sqrt{n}$, we have that K_G is not n-slice by Proposition 4.6. Hence, Theorem 1.5 leaves open the possibility to use the s-invariant to obstruct K_G from being n-slice.

For example, we can consider the case where $n=3$. If there is a 3-special RBG link with $R=U$ and $r>0$, such that K_G is 3-slice in some $\#^m\,\overline{\mathbb{CP}}^2$ and $s(K_B)=2$, then by Proposition 4.6, K_B is not 3-slice in any $\#^m\,\overline{\mathbb{CP}}^2$. Thus, such a link would produce an exotic $\#^m\,\overline{\mathbb{CP}}^2$. Note that $3=n=l(rl-2)$, so $(r,l)\in\{(5,1),(1,-1),(1,3)\}$. If $l=1$, then $r=5>1$, and by Proposition 4.9, we have that $s(K_G)\leq n-\sqrt{n}$, which cannot obstruct the 3-sliceness of K_G. Thus, all potentially useful 3-special RBG links are those with $r=1$ and $l\in\{-1,3\}$.

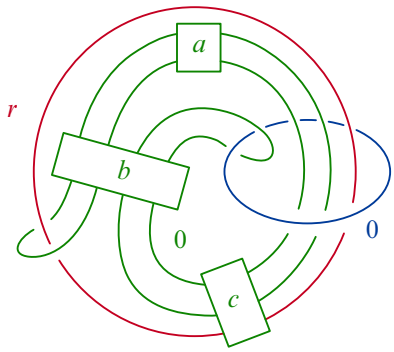


Figure 8: The diagram of $(r+2)$-special RBG links $L(a,b,c;r)$.

Here is a potential situation that could lead to an exotic $\#^m\,\overline{\mathbb{C}\mathbb{P}}^2$ using the s-invariant. Consider the family $L(a,b,c;r)$ of special RBG links in Figure 8. Since $l=-1$, the link $L(a,b,c;r)$ is an $(r+2)$-special RBG link. Figure 9 gives the diagrams of $K_B(a,b,c;r)$ and $K_G(a,b,c;r)$. Note that two diagrams in Figure 9 are isotopic for all choices of a,b and c. Thus, we denote $K_B(a,b,c;r)$ by $K(a,b,c)$, and $K_G(a,b,c;r)$ by $K(a,b,r-c)$.

Example 4.10 The link $L(-2,1,-2;1)$ (see Figure 8) is a 3-special RBG link. Using KnotJob [19], one can compute that $s(K_G)=0$ and $s(K_B)=2$. The knot $K_B(-2,1,-2;1)=K(-2,1,-2)$ is recognized by SnapPy [7] as $K9_533$, and $K_G(-2,1,-2;1)=K(-2,1,3)$ has a diagram with 28 crossings as in Figure 1, left, by KnotJob [19].

From Example 4.10, we obtain the following theorem.

Theorem 1.6 *If the knot $K(-2,1,3)$ from the left-hand side of Figure 1 is 3-slice in some $\#^m\,\overline{\mathbb{C}\mathbb{P}}^2$, then there exists an exotic $\#^m\,\overline{\mathbb{C}\mathbb{P}}^2$.*

Therefore, we are interested in whether $K(-2,1,3)$ is 3-slice in some $\#^m\,\overline{\mathbb{C}\mathbb{P}}^2$. Notice that if we change the colored crossings in Figure 9 by adding 2-handles, we get ribbon knots for any choice of $(a,b,c;r)$.

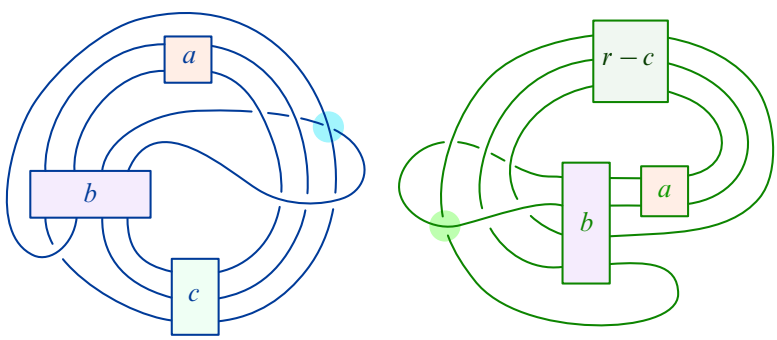


Figure 9: The diagrams of $K_B(a,b,c;r)$ and $K_G(a,b,c;r)$.

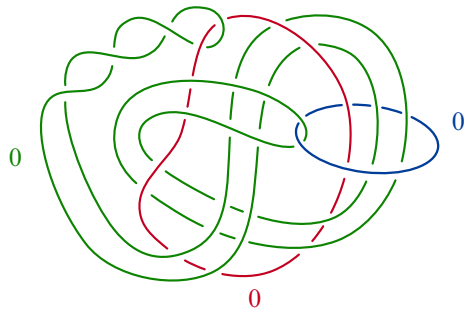


Figure 10: A diagram of the 2-special RBG link $L(-2,1,2;0)$ from the family in Example 4.10.

Therefore, both K_B and K_G are H-slice in $\mathbb{CP}^2$ and 4-slice in $\overline{\mathbb{CP}}^2$. Using SnapPy [7], we compute that $\tau(K(-2,1,3))=0$, which does not obstruct 3-sliceness. However, by considering $L(-2,1,-3;0)$, we can obstruct 2-sliceness.

Proposition 4.11 *The knot $K(-2,1,3)$ in Figure 1, left, is not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$.*

Proof The knot $K(-2,1,3)$ has the same diagram as $K_G(-2,1,-3;0)$. Consider the 2-special RBG link in the 4-parameter family in Figure 8 with $(a,b,c;r)=(-2,1,-3;0)$. Using KnotJob, we compute that $s(K_B(-2,1,-3;0))=2$. However, by Theorem 1.5(a), if $K_G(-2,1,-3;0)$ is 2-slice, then $s(K_B(-2,1,-3;0))\le 2-\sqrt{2}<2$. Thus, $K(-2,1,3)=K_G(-2,1,-3;0)$ is not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$. □

Although we cannot determine whether $K(-2,1,3)$ is 3-slice, Proposition 4.11 gives an example where we can use an n-special RBG link to obstruct the n-sliceness of a knot. Another example of this type is the 2-special RBG link in Figure 10.

Example 4.12 The 2-special RBG link $L(-2,1,2;0)$ in Figure 10 gives another example where $s(K_B)\neq s(K_G)$. Using SnapPy [7], we recognize K_G as $K9_533$. Figure 1, right, gives a diagram of K_B. Using KnotJob [19], we compute that $s(K_B)=0$, $s(K_G)=2$. Also, using SnapPy [7], we have that $\tau(K_B)=\tau(K_G)=0$.

Using the 2-RBG link in Figure 10, we obtain:

Proposition 4.13 *The knot $K(-2,1,2)$ in Figure 1, right, is not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$.*

Proof Since $K(-2,1,2)$ is the blue knot $K_B(-2,1,2;0)$ associated with the 2-special RBG link $L(-2,1,2;0)$ in Example 4.12 (see Figure 10), we apply Theorem 1.5(a) with $n=2$. Since $s(K_G)=2$ is larger than $n-\sqrt{n}=2-\sqrt{2}$, the associated knot K_B is not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$. Thus, $K(-2,1,2)$ is not 2-slice in any $\#^m\,\overline{\mathbb{CP}}^2$. □

So far, we have proved Theorem 1.7, which is a combination of Propositions 4.11 and 4.13 using Theorem 1.5. We do not know other proofs of Theorem 1.7. In general, there are two other approaches to obstruct knots from being slice n-slice in any $\#^m \overline{\mathbb{CP}}^2$. The first one is to use the τ invariant from [16], where Ozsváth and Szabó proved that the τ invariant of a knot K satisfies the adjunction inequality in a negative definite 4-manifold W with $b_1(W) = 0$; namely

$$2\tau(K) \leq 2g(\Sigma) - |[\Sigma]| - [\Sigma]\cdot[\Sigma],$$

where Σ is a properly, smoothly embedded surface in W without closed components, such that $\partial\Sigma = K$. Similarly to the proof of Proposition 4.6, we have that if K is n-slice in some $\#^m \overline{\mathbb{CP}}^2$, then $2\tau(K) \leq n - \sqrt{n}$. However, we have $s = \tau = 0$ for the knots in Figure 1, so this method does not apply to Theorem 1.7.

The second approach is to use the d-invariant from [15]. Theorem 9.6 in [15] gives a lower bound on the d-invariants of a rational homology sphere that bounds a definite 4-manifold. Moreover, by Proposition 1.6 in [14], we can compute the d-invariant of $S^3_2(K)$ by looking at the full knot Floer complex of K. A calculation of $\widehat{\mathrm{HFK}}$ with SnapPy [7] does not quite determine the full knot complexes of the knots in Figure 1, but suggests that they could be CFK-equivalent to the unknot. Therefore, this method does not apply either.

5 n-peculiar RBG links

In this section, we consider a different construction of $|n|$-RBG links (that are usually not n-special). Let $n \in \mathbb{Z}$ and $r \in \mathbb{Q}$. Consider a two component link $\{(K, n), (\mu_K, r)\}$, where μ_K is a meridian of K. By a Rolfsen twist, there exists a homeomorphism from $S^3_{n,r}(K, \mu_K)$ to $S^3_{n-1/r}(K)$, which restricts to the identity outside a tubular neighborhood of K. If $K = U$ and $n - 1/r = 1/t$ for some $t \in \mathbb{Z}$, then $\{(K, n), (\mu_K, r)\}$ is a surgery diagram for S^3.

Definition 1.8 A link $\{(R, r), (B, b), (G, g)\}$ is called an n-peculiar RBG link, if there exists $t \in \mathbb{Z}$ such that

- $R = U$ and B, G are meridians of R,
- $b = g = 1/r + 1/t$,
- $n = (g + b - 2l) - t(l - b)^2$,

where $l = \mathrm{lk}(B, G)$ under an orientation of L such that $\mathrm{lk}(B, R) = \mathrm{lk}(G, R) = 1$.

We can read out the diagram of K_G from its RBG link diagram in three steps. Let Δ_R be a meridian disk of B such that $\partial\Delta_R = R$. First, isotope G away from Δ_R by sliding G over B. Then, slam dunk R into B, and obtain a link diagram where R is deleted and B has framing $1/t$. Finally, blow down the blue component B by Rolfsen twists. We can get a diagram for K_B in a similar way.

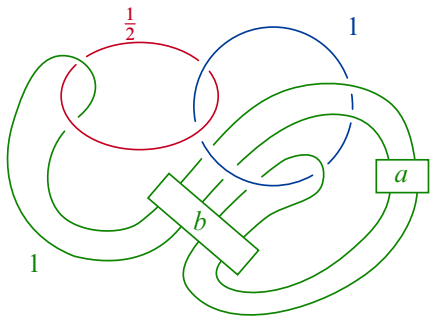


Figure 11: A family of 0-peculiar links.

Lemma 5.1 *An n-peculiar RBG link induces an n-surgery homeomorphism $\phi\colon S^3_n(K_B) \to S^3_n(K_G)$.*

Proof We keep track of the linking matrix M_L of the link L under those three diagram changes. We start with

$$M_L = \begin{bmatrix} r & 1 & 1 \\ 1 & b & l \\ 1 & l & g \end{bmatrix}.$$

Sliding G over B so that $|\Delta_R \cap G| = 0$, the linking matrix becomes

$$\begin{bmatrix} r & 1 & 0 \\ 1 & b & l-b \\ 0 & l-b & g+b-2l \end{bmatrix}.$$

After the slam-dunk of R into B, the linking matrix is

$$\begin{bmatrix} b-1/r & l-b \\ l-b & g+b-2l \end{bmatrix}.$$

Now, blow down the blue component by a Rolfsen twist along B. The framing of K_G is

$$f_g = (g+b-2l) - t(l-b)^2 = n. \qquad \square$$

Example 5.2 Consider the family of peculiar RBG links in Figure 11, parametrized by two twisting boxes. Since $l = 1$ and $b = g = 1$, we have $n = 0$. When (a, b) is $(2, -1)$ or $(3, -2)$, the peculiar link $L(a, b)$ generates a knot pair such that $s(K_G) = -2$ and $s(K_B) = 0$. (When $(a, b) = (2, -1)$, the knot K_B is 11_{270}.) However, since the signature of K_B is 2 in each case, the knots K_B are not H-slice in any $\#^m \mathbb{CP}^2$, so the knot pairs do not produce any exotic $\#^m \mathbb{CP}^2$.

References

[1] **T Abe**, **I D Jong**, **J Luecke**, **J Osoinach**, *Infinitely many knots admitting the same integer surgery and a four-dimensional extension*, Int. Math. Res. Not. 2015 (2015) 11667–11693 MR Zbl

[2] **S Akbulut**, *On 2-dimensional homology classes of 4-manifolds*, Math. Proc. Cambridge Philos. Soc. 82 (1977) 99–106 MR Zbl

[3] **S Akbulut**, *Knots and exotic smooth structures on 4-manifolds*, J. Knot Theory Ramifications 2 (1993) 1–10 MR Zbl

[4] **S Akbulut**, **R Kirby**, *Mazur manifolds*, Michigan Math. J. 26 (1979) 259–284 MR Zbl

[5] **S Boyer**, *Simply-connected 4-manifolds with a given boundary*, Trans. Amer. Math. Soc. 298 (1986) 331–357 MR Zbl

[6] **R Casals**, **J Etnyre**, **M Kegel**, *Stein traces and characterizing slopes*, Math. Ann. 389 (2024) 1053–1098 MR Zbl

[7] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy, a computer program for studying the geometry and topology of 3-manifolds* Available at `http://snappy.computop.org`

[8] **R E Gompf**, **A I Stipsicz**, *4-manifolds and Kirby calculus*, Grad. Stud. Math. 20, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[9] **K Hayden**, **L Piccirillo**, *The trace embedding lemma and spinelessness* (2019) arXiv 1912.13021 To appear in J. Differential Geom.

[10] **O A Ivanov**, **N Y Netsvetaev**, *On the intersection form of the result of gluing manifolds with degenerate intersection forms*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 231 (1995) 169–179 MR Zbl In Russian; translated in J. Math. Sci. 91 (1998) 3440–3447

[11] **C Manolescu**, **M Marengon**, **S Sarkar**, **M Willis**, *A generalization of Rasmussen's invariant, with applications to surfaces in some four-manifolds*, Duke Math. J. 172 (2023) 231–311 MR Zbl

[12] **C Manolescu**, **L Piccirillo**, *From zero surgeries to candidates for exotic definite 4-manifolds*, J. Lond. Math. Soc. 108 (2023) 2001–2036 MR Zbl

[13] **K Nakamura**, *Trace embeddings from zero surgery homeomorphisms*, J. Topol. 16 (2023) 1641–1664 Zbl

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

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

[16] **P Ozsváth**, **Z Szabó**, *Knot Floer homology and the four-ball genus*, Geom. Topol. 7 (2003) 615–639 MR Zbl

[17] **L Piccirillo**, *The Conway knot is not slice*, Ann. of Math. 191 (2020) 581–591 MR Zbl

[18] **Q Ren**, *Lee filtration structure of torus links*, Geom. Topol. 28 (2024) 3935–3960 MR Zbl

[19] **D Schütz**, *KnotJob*, Java software (2015) Available at `https://www.maths.dur.ac.uk/users/dirk.schuetz/knotjob.html`

[20] **K Varvarezos**, *Certain connect sums of torus knots with infinitely many non-characterizing slopes*, preprint (2023) arXiv 2302.05068

Department of Mathematics, Stanford University
Stanford, CA, United States

`qqhe@stanford.edu`

Received: 8 November 2023 Revised: 4 May 2024

msp

Algebraic & Geometric Topology **25**:6 (2025) 3775–3787
DOI: 10.2140/agt.2025.25.3775
Published: 6 October 2025

Powell's conjecture on the Goeritz group of S^3 is stably true

MARTIN SCHARLEMANN

In 1980, J Powell proposed that, for every genus g, five specific elements suffice to generate the Goeritz group $\mathcal{G}_g$ of genus g Heegaard splittings of S^3. Powell's conjecture remains undecided for $g \geq 4$. Let $\mathcal{P}_g \subset \mathcal{G}_g$ denote the subgroup generated by Powell's elements. Here we show that, for each genus g, the natural function $\mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ is trivial.

57K35, 57M50, 57M60

1 Introduction

Following early work of Goeritz [3], the *genus g Goeritz group* $\mathcal{G}_g$ of the 3-sphere can be described as the isotopy classes of orientation-preserving homeomorphisms of the 3-sphere that leave the standard genus g Heegaard surface T_g invariant, with its orientation preserved. Goeritz identified a finite set of generators for $\mathcal{G}_2$. In 1980, J Powell [6] extended Goeritz's set of generators to a set of five elements that he believed would generate the Goeritz group for any fixed higher-genus splitting, but his proof contained a serious gap. In [2], the Powell conjecture (as it is now called) was confirmed for $\mathcal{G}_3$, and, in [7], it is pointed out that one of Powell's proposed generators is redundant, so in fact only four of Powell's proposed generators need to be considered.

Powell's view of the Goeritz group, which we will adopt, is framed somewhat differently. Following Johnson and McCullough [4] (who extend the notion to arbitrary compact orientable manifolds), consider the space of left cosets $\mathrm{Diff}(S^3)/\mathrm{Diff}(S^3, T_g)$, where $\mathrm{Diff}(S^3, T_g)$ consists of those orientation-preserving diffeomorphisms of S^3 that carry T_g to itself. The fundamental group $\widetilde{\mathcal{G}}_g = \pi_1(\mathrm{Diff}(S^3)/\mathrm{Diff}(S^3, T_g))$ of this space projects to the genus g Goeritz group $\mathcal{G}_g$ (with kernel $\pi_1(\mathrm{SO}(3)) = \mathbb{Z}_2$ [6, page 197]) as follows: a nontrivial element is represented by an isotopy of T_g in S^3 that begins with the identity and ends with a diffeomorphism of S^3 that takes T_g to itself; this diffeomorphism of the pair (S^3, T_g) represents an element of the Goeritz group as defined earlier. The advantage of this point of view is that an element of $\widetilde{\mathcal{G}}_g$ can be viewed quite vividly: it is an excursion of T_g in S^3 that begins and ends with the standard picture of $T_g \subset S^3$. Such excursions are how Powell defines his proposed generators [6, Figure 4].

© 2025 MSP (Mathematical Sciences Publishers). Distributed under the Creative Commons Attribution License 4.0 (CC BY). Open Access made possible by subscribing institutions via Subscribe to Open.

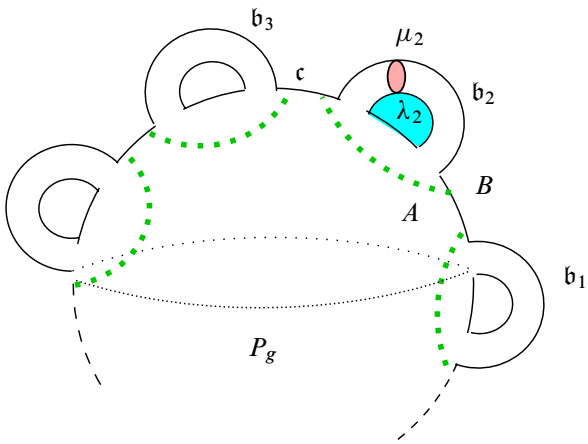


Figure 1: Powell's picture of T_g.

2 Powell's generators reexpressed

Powell's description of $T_g \subset S^3$ for $g \geq 2$ begins with a round 2-sphere in S^3, to which is then connect-summed a standard unknotted torus at each of g points in a circumference $\mathfrak{c}$ of the sphere. These summands $\mathfrak{b}_i$ for $1 \leq i \leq g$ will be called the standard genus one bubbles; their complement is a g-punctured sphere P_g. See Figure 1. For the Heegaard splitting $S^3 = A \cup_{T_g} B$, let $a_i \subset \mathfrak{b}_i \cap T_g$ be a circle (unique up to isotopy in the punctured torus $\mathfrak{b}_i \cap T_g$) that bounds a disk μ_i in the solid torus $A \cap \mathfrak{b}_i$. Similarly, let $b_i \subset \mathfrak{b}_i \cap T_g$ be a circle (unique up to isotopy rel a_i in the punctured torus $\mathfrak{b}_i \cap T_g$) that bounds a disk λ_i in the solid torus $B \cap \mathfrak{b}_i$, chosen so that $|a_i \cap b_i| = 1$. We will call a_i and b_i, respectively, the meridian and longitude of the summand $\mathfrak{b}_i$. Similarly, μ_i and λ_i will be called the meridian and longitudinal disks of $\mathfrak{b}_i$.

In terms of this picture, here are Powell's four proposed generators (the fifth being redundant):

- D_ω is the homeomorphism $(\mathfrak{b}_1, \mathfrak{b}_1 \cap T_g) \to (\mathfrak{b}_1, \mathfrak{b}_1 \cap T_g)$ shown in Figure 2, left. It preserves both μ_1 and λ_1 but reverses the orientation of each. We will call this the *standard flip* on $\mathfrak{b}_1$.
- D_η is the homeomorphism $(S^3, T_g) \to (S^3, T_g)$ shown in Figure 2, center: The punctured sphere P_g is rotated by $2\pi/g$ along the circumference $\mathfrak{c}$ and each standard bubble $\mathfrak{b}_i$ is moved to $\mathfrak{b}_{i+1}$, $i \in \mathbb{Z}/g$, sending each μ_i to μ_{i+1} and λ_i to λ_{i+1}. Fix an orientation for the disks μ_1 and λ_1, and observe that $(D_\eta)^{i-1}$ can be used to fix an orientation on each pair μ_i, λ_i, which is then preserved by D_η.
- Let $v \subset P_g$ be the subarc of $\mathfrak{c}$ connecting $\mathfrak{b}_1$ and $\mathfrak{b}_2$. Let B be the reducing ball for the splitting T_g obtained by attaching a 1-handle regular neighborhood of v to the bubbles $\mathfrak{b}_1$ and $\mathfrak{b}_2$. $D_{\eta_{12}}$ is the homeomorphism $(B, B \cap T_g) \to (B, B \cap T_g)$ shown in Figure 2, right. It exchanges the meridian disks μ_1 and μ_2 and the longitudinal disks λ_1 and λ_2, preserving the orientation of each. We will call this the *standard exchange* of $\mathfrak{b}_1$ and $\mathfrak{b}_2$.

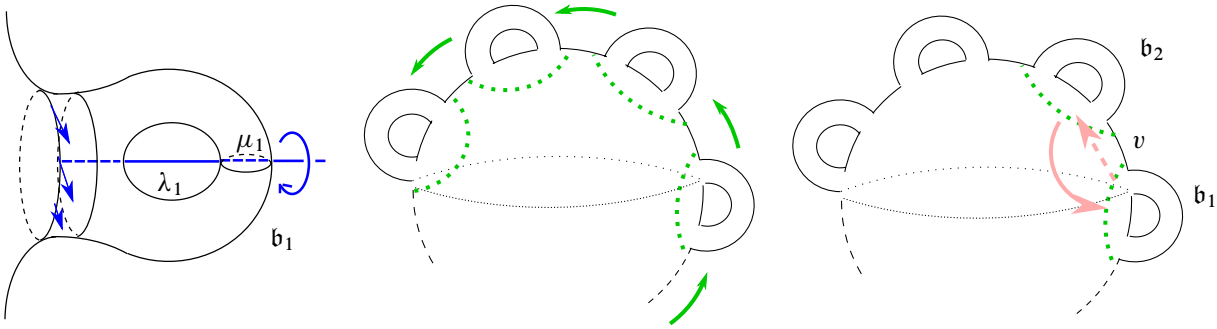


Figure 2: Powell's D_ω, the standard flip on $\mathfrak{b}_1$ (left), D_η (center), and $D_{\eta_{12}}$, the standard exchange of $\mathfrak{b}_1$ and $\mathfrak{b}_2$ (right).

- Let $v \subset T_g$ be an arc connecting $\partial\mu_1$ to $\partial\ell_2$, chosen so that inside the bubbles the interior of v is disjoint from the circles a_1, a_2, b_1, b_2, and outside the bubbles v is a subarc of $\mathfrak{c}$. The complex $\mu_1 \cup \ell_2 \cup v$ is called the *standard eyeglass* in T_g; see Figure 3. Powell's generator D_θ, illustrated in several panels of [6, Figure 4], makes use of this structure; it can be diagrammed more simply as in Figure 3 and is called the *standard eyeglass twist*. This move will be generalized later; see Figure 6.

As noted earlier, in the exact sequence

$$1 \to \pi_1(\mathrm{SO}(3)) \to \widetilde{\mathcal{G}}_g = \pi_1(\mathrm{Diff}(S^3)/\mathrm{Diff}(S^3, T_g) \to \mathcal{G}_g \to 1,$$

Powell's proposed generators actually lie in $\widetilde{\mathcal{G}}_g$, so it makes sense to note the easy:

Proposition 2.1 *Powell's proposed generators generate $\mathcal{G}_g$ if and only if they generate $\widetilde{\mathcal{G}}_g$.*

Proof One direction is obvious; the other follows immediately from the observation that the 2π-rotation $(D_\eta)^g$ is a Powell move that also represents the nontrivial element of $\pi_1(\mathrm{SO}(3))$. □

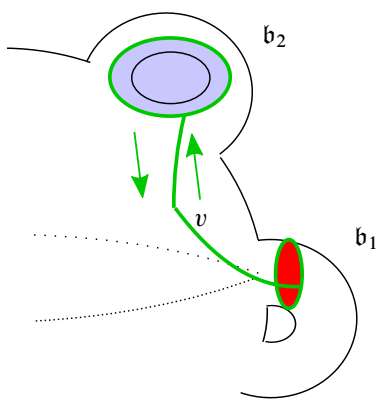


Figure 3: Powell's D_θ, the standard eyeglass twist.

Definition 2.2 The Powell group $\mathcal{P}_g$ is the subgroup of the Goeritz group $\mathcal{G}_g$ that is generated by Powell's four proposed generators D_θ, D_ω, $D_{\eta_{12}}$, D_η. Any element of $\mathcal{P}_g$ (that is, any composition of these generators) will be called a *Powell move*.

Here is another easy line of argument: Both of Powell's moves D_η and $D_{\eta_{12}}$ can be thought of as induced by a homeomorphism on the g-punctured sphere P_g, and such a homeomorphism also defines an element of the g-stranded braid group $\mathcal{B}_g$ of the sphere. Furthermore, $(D_\eta)^{i-1}$ will conjugate the standard exchange $D_{\eta_{12}}$ to a similar exchange (which we will denote by ϕ_i and also call standard) between $\mathfrak{b}_i$ and $\mathfrak{b}_{i+1}$, so each standard exchange ϕ_i is a Powell move. Moreover, viewed as an element of the braid group $\mathcal{B}_g$, each exchange ϕ_i for $1 \le i \le g-1$ just described is exactly one of the $g-1$ canonical half-twist generators σ_i of $\mathcal{B}_g$ (see [1, Chapter 9]). Similarly, $(D_\eta)^{i-1}$ will conjugate D_ω to a flip in the bubble $\mathfrak{b}_i$, which we will call the standard flip in $\mathfrak{b}_i$ and denote by $\omega_i \in \mathcal{P}_g$.

Proposition 2.3 *The Powell group $\mathcal{P}_g$ is also generated by these $g+1$ elements of $\mathcal{P}_g$:*

$$D_\theta, \quad D_\omega, \quad \{\phi_i \mid 1 \le i \le g-1\}.$$

Proof In the braid group $\mathcal{B}_g$, the composition $\sigma_1\sigma_2\cdots\sigma_{g-1}$ visibly has the effect of moving g points evenly distributed in the circumference $\mathfrak{c} \subset S^2$ by a rotation of $2\pi/g$. (See also [1, Chapter 9].) It follows that, for some even k (easily but unimportantly found to be $k=2$; see below), $D_\eta = (D_\omega)^k\phi_1\phi_2\cdots\phi_{g-1}$ in $\mathcal{P}_g$. Thus D_η lies in the subgroup of $\mathcal{P}_g$ generated by D_θ, D_ω and $\{\phi_i \mid 1 \le i \le g-1\}$, as does $D_{\eta_{12}} = \phi_1$. Thus each of the generators D_θ, D_ω, $D_{\eta_{12}}$, D_η lies in the subgroup of $\mathcal{P}_g$ generated by D_θ, D_ω and $\{\phi_i \mid 1 \le i \le g-1\}$, as required. □

Figure 4 demonstrates the equality $D_\eta = (D_\omega)^2\phi_1\phi_2\cdots\phi_{g-1}$ (composed right to left) in the case $g=4$. The red arrows in each $\mathfrak{b}_i$ are meant to show orientation within the bubble: the arrows indicate the normal direction to a_i. In the left panel, $\phi_1\phi_2\phi_3$ rotates the red arrow in $\mathfrak{b}_4$ through an angle $-\frac{3\pi}{2}$ to the red arrow in $\mathfrak{b}_1$; in the right panel, D_η rotates it $\frac{\pi}{2}$.

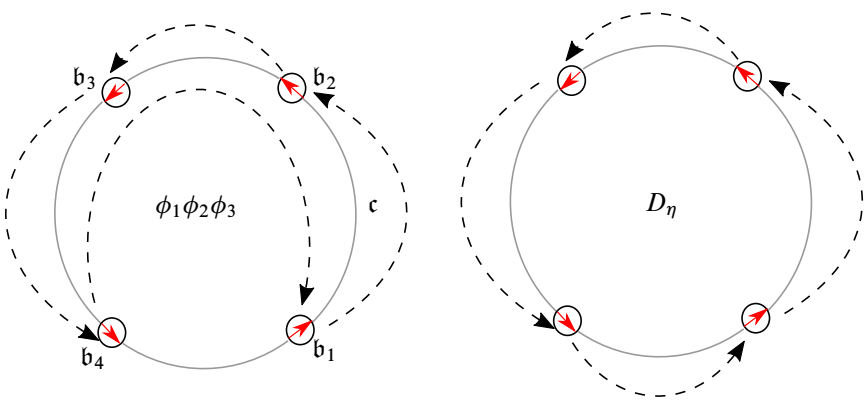


Figure 4: $D_\eta = (D_\omega)^2\phi_1\phi_2\phi_3$.

3 What the Powell group can do

We have seen above that the Powell group $\mathcal{P}_g$ contains any automorphism of T_g obtained by operating on the planar surface P_g by an element of the braid group $\mathcal{B}_g$. In [2, Sections 1–4], broader types of automorphisms are also shown to lie in $\mathcal{P}_g$. For example, [2, Lemma 1.4] says this:

Lemma 3.1 *Any braid move of a collection of standard bubbles over their complementary surface is a Powell move.* □

Figure 5 shows an example in the case of two standard bubbles.

In the case of a single standard bubble $\mathfrak{b}_i$, a braid move is just an isotopy of $\mathfrak{b}_i$ through some path in $T_g - \mathfrak{b}_i$ that returns $\mathfrak{b}_i$ to itself. This will be called a *standard bubble move* of $\mathfrak{b}_i$ and is a Powell move. More generally, if $\mathfrak{b}$ is an arbitrary bubble, such an isotopy of $\mathfrak{b}_i$ through some path in $T_g - \mathfrak{b}_i$ is called a (*generic*) *bubble move*. A generic bubble move is not obviously Powell, but see Corollary 3.5 below.

Here is one consequence: Let $q\colon (S^3, T_{g+1}) \to (S^3, T_g)$ be the quotient map that shrinks the standard bubble $\mathfrak{b}_{g+1} \subset S^3$ to a point $\star \in T_g$. Suppose $\tau\colon (S^3, T_g) \to (S^3, T_g)$ represents an element in $\mathcal{G}_g$ and let α be an embedded path in T_g from $\star$ to $\tau^{-1}(\star)$. Let $\alpha^{\star}_{\tau^{-1}(\star)}\colon (S^3, T_g) \to (S^3, T_g)$ be a homeomorphism whose support lies in a regular neighborhood of α and pushes $\star$ to $\tau^{-1}(\star)$. (See [5, Homogeneity lemma].) Then $\tau\alpha^{\star}_{\tau^{-1}(\star)}(\star) = \star$, so τ induces $\tau_\alpha\colon (S^3, T_{g+1}) \to (S^3, T_{g+1})$ by replacing a neighborhood of $\star$ with the standard bubble $\mathfrak{b}_{g+1}$. (The construction can be extended to the case in which the path α is not necessarily embedded, but is in general position, by breaking α up into a finite sequence of embedded paths.)

Proposition 3.2 *For any genus $g \geq 2$, the construction $\tau \to \tau_\alpha$ described above determines a function*

$$\iota^+\colon \mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$$

with the property that $\iota^+(\mathcal{P}_g)$ is trivial.

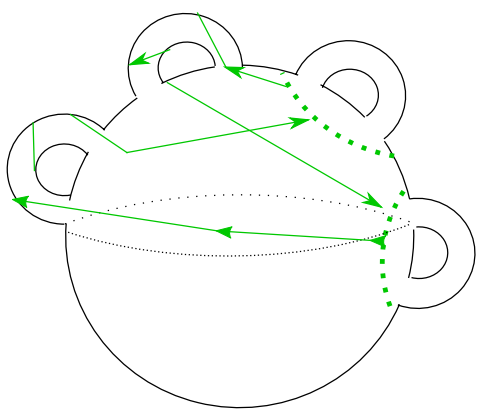


Figure 5

Proof The construction above assigns the homeomorphism $\tau_\alpha : (S^3, T_{g+1}) \to (S^3, T_{g+1})$ to τ; we are interested in the coset that τ_α represents in $\mathcal{G}_{g+1}/\mathcal{P}_{g+1}$. The first observation is that the coset does not depend on the path α. Indeed, suppose β is a different path in T_g from $\star$ to $\tau^{-1}(\star)$. Then τ_α and τ_β differ by a standard bubble move on $\mathfrak{b}_{g+1}$ along the path $\overline{\alpha}\beta$, and by Lemma 3.1 such a bubble move is in $\mathcal{P}_{g+1}$. Thus τ_α and τ_β represent the same coset in $\mathcal{G}_{g+1}/\mathcal{P}_{g+1}$.

Claim *This assigned coset in $\mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ is independent of the choice of τ as a representative of a given element of $\mathcal{G}_g$.*

Proof Suppose that $\tau_1 : (S^3, T_g) \to (S^3, T_g)$ represents the same element in $\mathcal{G}_g$ as $\tau = \tau_0$. This means that there is an isotopy $\tau_t, 0 \le 1$ from τ_0 to τ_1. Such an isotopy defines a path β in T_g from $\tau_0^{-1}(\star)$ to $\tau_1^{-1}(\star)$ given by $\beta(t) = \tau_t^{-1}(\star)$. Then the concatenation $\alpha_1 = \alpha * \beta$ is a path from $\star$ to $\tau_1^{-1}(\star)$ and so determines a coset in $\mathcal{G}_{g+1}/\mathcal{P}_{g+1}$, as described just before the proposition. That is, the coset is represented by $(\tau_1)_{\alpha_1}$, the homeomorphism obtained from $\tau_1(\alpha_1)^*_{\tau_1^{-1}(\star)}$ by replacing a neighborhood of $\star$ with the standard bubble $\mathfrak{b}_{g+1}$.

There is now a straightforward way to construct an isotopy from $(\tau_0)_{\alpha_0}$ to $(\tau_1)_{\alpha_1}$: For each $0 \le t \le 1$, let $\beta_t : I \to T_g$ be the path given by $\beta_t(s) = \beta(st)$. Then β_t is a path from $\beta(0) = \tau_0^{-1}(\star)$ to $\beta(t) = \tau_t^{-1}(\star)$, so, for each $0 \le t \le 1$, $\alpha_t = \alpha * \beta_t$ is a path from $\star$ to $\tau_t^{-1}(\star)$. Using this path, construct a homeomorphism $(\tau_t)_{\alpha_t} : (S^3, T_{g+1}) \to (S^3, T_{g+1})$ in the same way we constructed $(\tau_0)_{\alpha_0}$ and $(\tau_1)_{\alpha_1}$. Then, letting t now run from 0 to 1, $(\tau_t)_{\alpha_t} : (S^3, T_{g+1}) \to (S^3, T_{g+1})$ defines an isotopy from $(\tau_0)_{\alpha_0}$ to $(\tau_1)_{\alpha_1}$, as required. Thus $(\tau_0)_{\alpha_0}$ and $(\tau_1)_{\alpha_1}$ represent the same element of $\mathcal{G}_{g+1}$. This proves the claim. ◁

Since the assigned coset in $\mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ is independent of the choice of α and the choice of representative of a class in $\mathcal{G}_g$, the construction defines a function $\iota^+ : \mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ without ambiguity.

To show that $\iota^+(\mathcal{P}_g)$ lies in $\mathcal{P}_{g+1}$, consider the generators of $\mathcal{P}_g$ described in Proposition 2.3. Each involves at most two standard bubbles, so the same will be true once ι^+ carries them to $\mathcal{G}_{g+1}$. Following Lemma 3.1, arbitrary exchanges of standard bubbles are Powell moves in $\mathcal{G}_{g+1}$, so, up to Powell moves, we can view the generators of $\iota^+(\mathcal{P}_g)$ as each involving only the standard bubbles $\mathfrak{b}_1$ and $\mathfrak{b}_2$, so each is a standard generator of $\mathcal{P}_{g+1}$. □

Another useful lemma requires the following generalization of the standard eyeglass move D_θ, originally [2, Definition 2.1]:

Definition 3.3 For $S^3 = A \cup_T B$ a Heegaard splitting of S^3, an *eyeglass* $\eta \subset S^3$ is the union of two properly embedded disjoint disks $\ell_a \subset A$ and $\ell_b \subset B$ (the *lenses* of η), with an embedded arc $v \subset T - \partial(\ell_a \cup \ell_b)$ (the *bridge* of η) connecting their boundaries. The 1-complex $\eta \cap T$ (the bridge, together with the boundary of the lenses) is called the *frame* of η.

The eyeglass η defines a natural automorphism $(S^3, T) \to (S^3, T)$, with support in a ball neighborhood of η, as illustrated in Figure 6. This automorphism is called an *eyeglass twist* based on η.

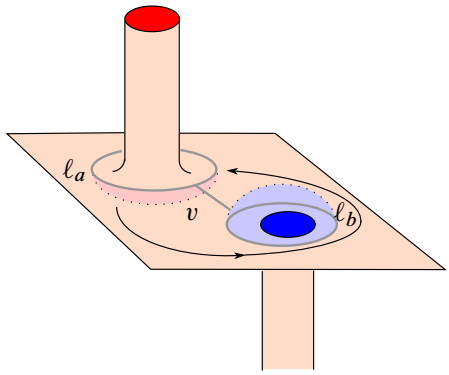


Figure 6: An eyeglass twist.

Since the support of an eyeglass twist lies entirely in a ball, it is isotopic in S^3 to the identity. Thus any eyeglass twist of the standard splitting $A \cup_{T_g} B$ lies in $\mathcal{G}_g$.

Remark Figure 6 shows the lens disk ℓ_a twisting once around the lens disk ℓ_b, but this is an artifact of the figure; it could equally well have been drawn symmetrically, with ℓ_b twisting once around the lens disk ℓ_a.

Suppose $c \subset T_g$ separates T_g into two components: T_A containing the punctured tori in $g_1 < g$ standard bubbles, and T_B containing the punctured tori in the other $g_2 = g - g_1$ standard bubbles.

Lemma 3.4 [2, Lemma 3.4] *Suppose η is an eyeglass in T_g whose lenses consist of a disk $a \subset A$ with $\partial a \subset T_A$ and a disk $b \subset B$ with $\partial b \subset T_B$. Suppose further that the bridge v intersects c exactly once. Then an eyeglass twist along η is a Powell move.*

Corollary 3.5 *Suppose $\mathfrak{b}$ is a generic bubble disjoint from the standard meridian μ_i and standard longitude λ_i of a standard bubble $\mathfrak{b}_i$, and γ is an embedded arc in $T_g - \mathfrak{b}$ from $\partial\mathfrak{b}$ to a_i (resp. b_i) whose interior is disjoint from both a_i and b_i. Then a bubble move of $\mathfrak{b}$ around the path $\overline{\gamma} a_i \gamma$ (resp. $\overline{\gamma} b_i \gamma$) is a Powell move.*

Proof Let $c \subset T_g$ be the boundary of a regular neighborhood of $a_i \cup b_i$, separating the standard bubble $\mathfrak{b}_i$ from the rest of T. Then the arc γ intersects c in one point and the bubble move around $\overline{\gamma} a_i \gamma$, say, can be described as an eyeglass move along an eyeglass with one lens $\mu_i \subset A$, the other lens the disk $\partial\mathfrak{b} \cap B$, and bridge the arc γ. The result then follows from Lemma 3.4. □

4 Stabilization and topological conjugacy

Definition 4.1 Orientation-preserving homeomorphisms $h_1, h_2 : (S^3, T_g) \to (S^3, T_g)$ are *topologically conjugate* if there is an orientation-preserving homeomorphism $g : (S^3, T_g) \to (S^3, T_g) \in \mathcal{G}_g$ such that $g^{-1} h_1 g = h_2$. A homeomorphism topologically conjugate to a standard flip is called a *generic flip*; a homeomorphism topologically conjugate to a standard exchange is called a *generic exchange*.

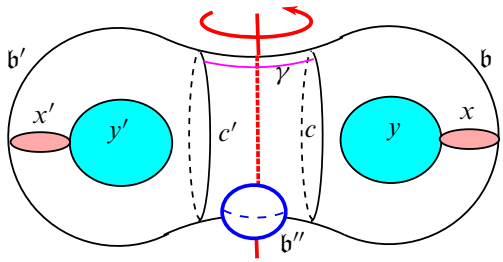


Figure 7: A bubble exchange, via a Goeritz move in $\mathcal{G}_2$.

For example, suppose $\mathfrak{b}$ is *any* genus one bubble for T_g, so $T \cap \mathfrak{b}$ is a once-punctured unknotted torus whose meridian bounds a properly embedded disk $x \in A$ and longitude bounds a properly embedded disk $y \in B$. Let $h\colon (S^3, T_g) \to (S^3, T_g) \in \mathcal{G}_g$ be the homeomorphism which is the identity outside $\mathfrak{b}$ but looks like Figure 2, left, inside $\mathfrak{b}$, with μ_1 replaced by x and λ_1 by y. Then it follows from [10] that h is a generic flip, and from the definition that any generic flip can be described in this way.

Similarly, suppose $\mathfrak{b}$ and $\mathfrak{b}'$ are any two disjoint genus one bubbles for T_g, and $\gamma \subset T$ is a path in their complement between the circles $c = \partial\mathfrak{b} \cap T$ and $c' = \partial\mathfrak{b}' \cap T$. A generic exchange is as shown in Figure 2, right, with $\mathfrak{b}_1$, $\mathfrak{b}_2$, v replaced respectively with $\mathfrak{b}$, $\mathfrak{b}'$, γ. The generic exchange can also be depicted as a generalization of one of Goeritz's original generators, as in Figure 7. In this depiction, the visible genus two splitting is rotated π around the vertical red axis, as Goeritz did. But in our case the blue ball at the base is a genus $g-2$ bubble $\mathfrak{b}''$ and the complement of $\mathfrak{b}''$, the genus two splitting we see, is the tube sum of bubbles $\mathfrak{b}$ and $\mathfrak{b}'$ along the arc γ.

It follows immediately from Lemma 3.1 that, if both $\mathfrak{b}$ and $\mathfrak{b}'$ are standard, then the exchange is a Powell move, even though γ can be arbitrary. More surprising and useful is the following extension:

Proposition 4.2 *If one of the bubbles in a generic exchange is standard, then the exchange is a Powell move.*

We defer the proof in order to describe how the proposition fits in to the main result of this paper.

Proposition 4.3 *For $\mathfrak{b}_{g+1}$ the standard bubble in (S^3, T_{g+1}), suppose the following two assumptions are true*:

(1) *Any generic bubble exchange between $\mathfrak{b}_{g+1}$ and a disjoint genus one bubble in (S^3, T_{g+1}) is a Powell move. In other words, Proposition 4.2 is true.*

(2) *Any eyeglass twist in (S^3, T_{g+1}) in which the eyeglass frame is disjoint from $\mathfrak{b}_{g+1}$ is a Powell move.*

Then the function $\iota^+\colon \mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ defined in Proposition 3.2 is trivial.

Proof We begin with two claims, assuming the two assumptions are true:

Claim 1 *Suppose $\mathfrak{b}$ is a bubble in T_{g+1} that is disjoint from $\mathfrak{b}_{g+1}$. Then the generic flip in $\mathfrak{b}$ is a Powell move.*

Proof Recall from the discussion before Proposition 2.3 that the standard flip ω_{g+1} of $\mathfrak{b}_{g+1}$ is a Powell move. Suppose $\mathfrak{b}$ is a generic bubble disjoint from $\mathfrak{b}_{g+1}$. Choose an embedded path β from $\mathfrak{b}_{g+1}$ to $\mathfrak{b}$ and let ρ be the bubble exchange of $\mathfrak{b}_{g+1}$ and $\mathfrak{b}$ along β, a Powell move by assumption (1). Then $\rho^{-1}\omega_{g+1}\rho$ is the generic flip of $\mathfrak{b}$ and is the composition of Powell moves. This proves the claim. □

Claim 2 *Suppose $\mathfrak{b}$ and $\mathfrak{b}'$ are disjoint bubbles in T_{g+1} that are also disjoint from $\mathfrak{b}_{g+1}$. Suppose $\alpha \subset T_{g+1}$ is an arc between them that is also disjoint from $\mathfrak{b}_{g+1}$. Then the generic exchange of $\mathfrak{b}$ with $\mathfrak{b}'$ along α is a Powell move.*

Proof Consider the 3-punctured genus $g-2$ surface $T_- = T_{g+1} - \mathfrak{b} \cup \mathfrak{b}' \cup \mathfrak{b}_{g+1}$. Let $\beta \subset T_-$ (resp. $\beta' \subset T_-$) be embedded arcs from $\mathfrak{b}$ (resp. $\mathfrak{b}'$) to $\mathfrak{b}_{g+1}$, chosen so that α, β, β' are all disjoint and lie in a disk in T_-.

By assumption (1), the bubble exchanges ρ, ρ' along, respectively, β, β' are Powell moves. Thus the composition $\rho^{-1}\rho'\rho$ is a Powell move. But $\rho^{-1}\rho'\rho\colon T_- \to T_-$ is easily seen to be isotopic to the exchange of $\mathfrak{b}$ and $\mathfrak{b}'$ along α, proving the claim. (The composition $\rho'\rho\rho'^{-1}$ is also isotopic to this exchange. Indeed the consequent isotopy between $\rho^{-1}\rho'\rho$ and $\rho'\rho\rho'^{-1}$ acting on a 3-punctured disk in T_- is the source of the standard braid relation $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$.) □

Following the claims, we exploit the central result of [9]: the Goeritz group $\mathcal{G}_g$ is generated by eyeglass twists and topological conjugates of generators of $\mathcal{P}_g$. We consider the generators of $\mathcal{P}_g$ from Proposition 2.3. These consist of Powell's D_θ, which is itself an eyeglass twist; Powell's D_ω, which is a standard flip; and the collection $\{\phi_i\}$ of standard exchanges. Any topological conjugate of an eyeglass twist is an eyeglass twist, any topological conjugate of a standard flip is a generic flip, and any topological conjugate of a standard exchange ϕ_i is a generic exchange. Assumption (2) above says that ι^+ takes any eyeglass twist in $\mathcal{G}_g$, (its frame disjoint from the point $\star$ by general position) to $\mathcal{P}_{g+1} \subset \mathcal{G}_{g+1}$. Similarly, Claim 1, using assumption (1), says that ι^+ takes any generic flip in $\mathcal{G}_g$ to $\mathcal{P}_{g+1}$, and Claim 2, also using assumption (1), says that ι^+ takes any generic exchange in $\mathcal{G}_g$ to $\mathcal{P}_{g+1}$. Since each generator of $\mathcal{G}_g$ is taken to $\mathcal{P}_{g+1}$, $\iota^+\colon \mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ is trivial, as required. □

We now show that in fact both assumptions in Proposition 4.3 are true. The proofs of both are highly visual. We first prove assumption (2) of Proposition 4.3.

Proposition 4.4 *Suppose $\tau \in \mathcal{G}_{g+1}$ is an eyeglass twist whose eyeglass η is disjoint from the standard bubble $\mathfrak{b}_{g+1}$. Then τ is a Powell move.*

Proof Let $\{\ell_a, \ell_b, v\}$ be the frame of η. Let u be an arc such that

- the ends of u lie on a_{g+1} near the point $a_{g+1} \cap b_{g+1}$ and on ℓ_a near the point $\ell_a \cap v$,
- u crosses the circle $c = \partial\mathfrak{b}_{g+1}$ once, and
- u is otherwise disjoint from $a_{g+1} \cup b_{g+1}$ and η.

See Figure 8.

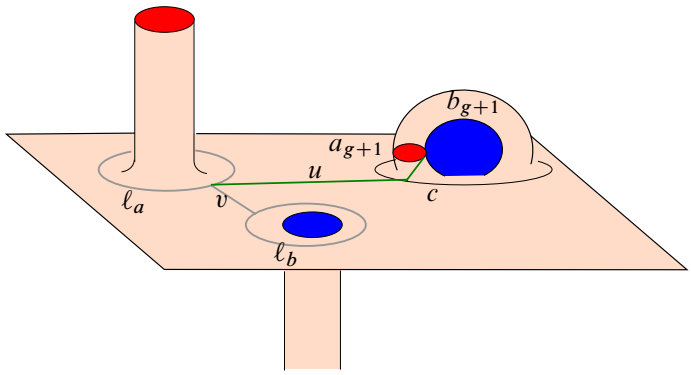


Figure 8

Let η' be the eyeglass given by $\{\mu_{g+1}, \ell_b, u \cup v\}$. We know from Lemma 3.4 that an eyeglass twist τ' along η' is a Powell move. Let ℓ'_a be the band sum of ℓ_a and μ_{g+1} along u and observe, by watching the motion of ℓ_b, that the composition $\tau\tau'$ is an eyeglass twist τ_+ whose eyeglass is η_+ given by $\{\ell'_a, \ell_b, v\}$.

Now let η_1 be the eyeglass given by $\{\ell_a, \lambda_{g+1}, u\}$ and let τ_1 be the eyeglass twist along η_1. See Figure 9. Again, from Lemma 3.4, τ_1 is a Powell move. Further observe that $\tau_1^{-1}(\eta_+)$ is an eyeglass η'_+ with lenses μ_{g+1} and ℓ_b and a bridge that intersects c in a single point, so again an eyeglass twist τ'_+ along η'_+ is a Powell move. See Figure 10.

Since $\tau_1^{-1}(\eta_+) = \eta'_+$ it follows that $\tau_+ = \tau_1\tau'_+\tau_1^{-1}$. As a composition of Powell moves, τ_+ is a Powell move. But we have earlier shown that $\tau = \tau_+\tau'^{-1}$, so, as a composition of Powell moves, τ is a Powell move. □

Proposition 4.5 *Suppose $\tau \in \mathcal{G}_{g+1}$ is a bubble exchange between the standard bubble $\mathfrak{b}_{g+1}$ and a generic bubble $\mathfrak{b}$. Then τ is a Powell move.*

Proof The entire proof is contained in Figure 11. In the top panel, the standard bubble $\mathfrak{b}_{g+1}$ is on the right, with the meridian circle a_{g+1} and the longitude circle b_{g+1} both shown in fuchsia. On the left is

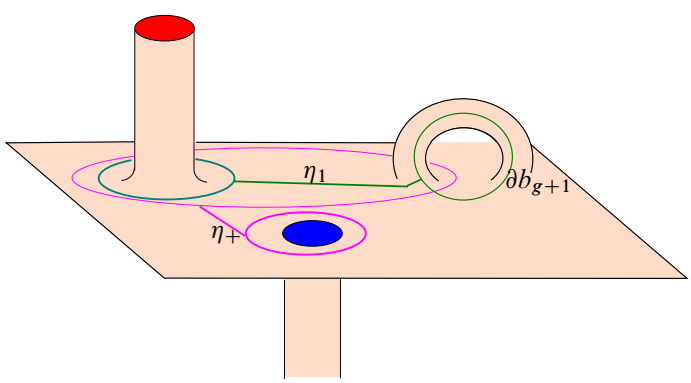


Figure 9

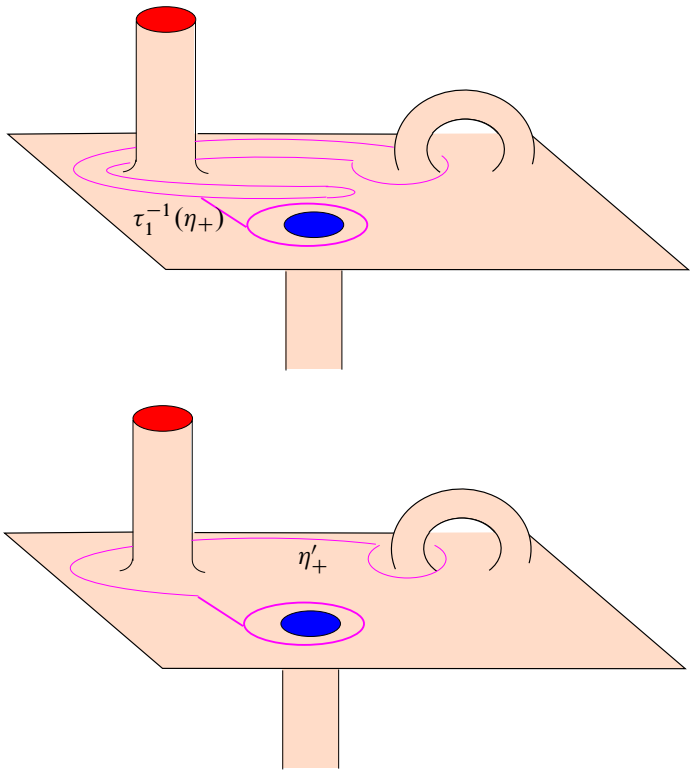


Figure 10

a generic bubble with the boundaries of the meridian $x \subset A$ and longitude $y \subset B$ shown in blue. Also shown there are two eyeglasses η_1 and η_2, in red and green, respectively. Since the bubble on the right is standard, by Lemma 3.4 the eyeglass twists τ_i along η_i are both Powell moves. The direction of the twist along the η_i that we will use is shown by arrows on the longitudinal lenses of η_1 and η_2, parallel respectively to ∂y and b_{g+1}.

The next two panels down show the effect of first τ_1 then τ_2. The following panel shows the effect of a (clockwise π) flip on the standard bubble $\mathfrak{b}_{g+1}$, a Powell move. Then follow two Powell moves: τ_1^{-1} and a final move of the standard bubble around the meridian x of the generic bubble. The final result in the last panel is that the pair μ_{g+1}, λ_{g+1} and the pair x, y have been exchanged.

Note that the orientations of x, y and μ_{g+1}, λ_{g+1} are not an issue: if an exchange as above moves an oriented μ_{g+1} to an unwanted orientation of x (or vice versa), then precomposing (or postcomposing) with a standard flip (a Powell move) will fix the problem. □

Corollary 4.6 *The function $\iota^+ : \mathcal{G}_g \to \mathcal{G}_{g+1}/\mathcal{P}_{g+1}$ defined in Proposition 3.2 is trivial.* □

Two final notes The proof of Proposition 4.5 focuses on the standard bubble $\mathfrak{b}_{g+1}$ and the generic bubble $\mathfrak{b}$, both lying in the neighborhood of a disk D that contains the exchange; the exchange map is the

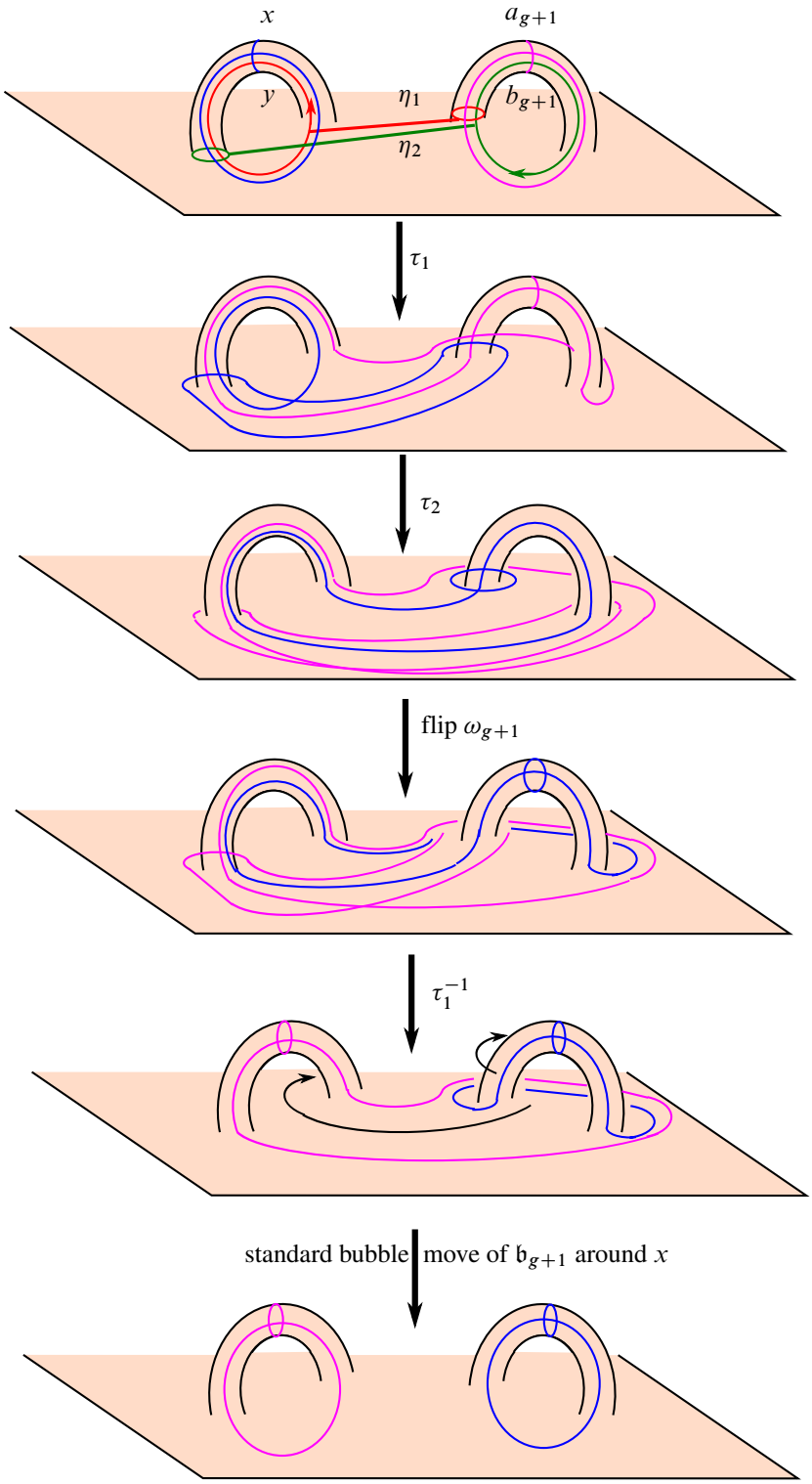


Figure 11: An exchange that is a Powell move.

identity map outside of the neighborhood of D. Left undiscussed is the pair of pants $D-(\mathfrak{b}\cup\mathfrak{b}_{g+1})\subset T_{g+1}$. Observe, though, that any homeomorphism of this pair of pants, fixed on its boundary, is also a Powell move. Here is the argument: Any such homeomorphism can be constructed by Dehn twisting some number of times around each of the boundary components of the pair of pants. A Dehn twist at the boundary of the standard bubble $\mathfrak{b}_{g+1}$ is a Powell move, for it is a composition of two standard flips. Similarly, conjugating this Powell Dehn twist by the exchange just described, also a Powell move, gives a Dehn twist around the bubble $\mathfrak{b}$, so such a Dehn twist is also Powell. Lastly, it is easy to check that the exchange defined above gives a simple (clockwise) half-twist on a collar of ∂D, so a full Dehn twist around that boundary component can be accomplished just by doing the exchange twice.

Because the exchange in Proposition 4.5 moves $\mathfrak{b}_{g+1}$ to another bubble, the exchange itself is not in the image of ι^+ and its construction may seem a bit ad hoc. An online version of this paper [8] contains a lengthy appendix that puts the exchange into a broader context, one involving an order 12 dihedral subgroup of $\mathcal{G}_2$, each of whose elements somewhat naturally defines an element of $\mathcal{G}_{g+1}$ that also lies in $\mathcal{P}_{g+1}$ but not in the image of ι^+. Figure 7 shows the connection between one such element of $\mathcal{G}_2$ and a bubble exchange like that of Proposition 4.5.

References

[1] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Mathematical Series 49, Princeton Univ. Press (2012) MR

[2] **M Freedman**, **M Scharlemann**, *Powell moves and the Goeritz group*, preprint (2018) arXiv 1804.05909

[3] **L Goeritz**, *Die abbildungen der brezelfläche und der vollbrezel vom geschlecht 2*, Abh. Math. Sem. Univ. Hamburg 9 (1933) 244–259 MR

[4] **J Johnson**, **D McCullough**, *The space of Heegaard splittings*, J. Reine Angew. Math. 679 (2013) 155–179 MR

[5] **J W Milnor**, *Topology from the differentiable viewpoint*, Princeton Univ. Press (1997) MR

[6] **J Powell**, *Homeomorphisms of S^3 leaving a Heegaard surface invariant*, Trans. Amer. Math. Soc. 257 (1980) 193–216 MR

[7] **M Scharlemann**, *One Powell generator is redundant*, Proc. Amer. Math. Soc. Ser. B 7 (2020) 138–141 MR

[8] **M Scharlemann**, *Powell's conjecture on the Goeritz group of S^3 is stably true*, preprint (2022) arXiv 2210.13629v2

[9] **M Scharlemann**, *Generating the Goeritz group of S^3*, J. Assoc. Math. Res. 2 (2024) 209–335 MR

[10] **F Waldhausen**, *Heegaard–Zerlegungen der 3-Sphäre*, Topology 7 (1968) 195–203 MR

Department of Mathematics, UC Santa Barbara
Santa Barbara, CA, United States

mgscharl@math.ucsb.edu

Received: 24 January 2023 Revised: 25 October 2023

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 25 Issue 6 (pages 3145–3787) 2025

Holomorphic polygons and the bordered Heegaard Floer homology of link complements 3145

THOMAS HOCKENHULL

Exact Lagrangian tori in symplectic Milnor fibers constructed with fillings 3225

ORSOLA CAPOVILLA-SEARLE

A note on embeddings of 3-manifolds in symplectic 4-manifolds 3251

ANUBHAV MUKHERJEE

A note on knot Floer homology of satellite knots with (1, 1)-patterns 3271

WEIZHE SHEN

A K-theory spectrum for cobordism cut and paste groups 3287

RENEE S HOEKZEMA, CARMEN ROVI and JULIA SEMIKINA

The Curtis–Wellington spectral sequence through cohomology 3315

DANA HUNTER

The slices of quaternionic Eilenberg–Mac Lane spectra 3341

BERTRAND J GUILLOU and CARISSA SLONE

Cocycles of the space of long embeddings and BCR graphs with more than one loop 3385

LEO YOSHIOKA

Asymptotic cones of snowflake groups and the strong shortcut property 3429

CHRISTOPHER H CASHEN, NIMA HODA and DANIEL J WOODHOUSE

Whitney tower concordance and knots in homology spheres 3503

CHRISTOPHER W DAVIS

The asymptotic behaviors of the colored Jones polynomials of the figure-eight knot, and an affine representation 3523

HITOSHI MURAKAMI

The Goldman bracket characterizes homeomorphisms between noncompact surfaces 3585

SUMANTA DAS, SIDDHARTHA GADGIL and AJAY KUMAR NAIR

A geometric computation of cohomotopy groups in codegree one 3603

MICHAEL JUNG and THOMAS O ROT

Calabi–Yau structure on the Chekanov–Eliashberg algebra of a Legendrian sphere 3627

NOÉMIE LEGOUT

On the resolution of kinks of curves on punctured surfaces 3679

CHRISTOF GEISS and DANIEL LABARDINI-FRAGOSO

Weinstein presentations for high-dimensional antisurgery 3707

IPSITA DATTA, OLEG LAZAREV, CHINDU MOHANAKUMAR and ANGELA WU

Singular Legendrian unknot links and relative Ginzburg algebras 3737

JOHAN ASPLUND

An RBG construction of integral surgery homeomorphisms 3755

QIANHE QIN

Powell's conjecture on the Goeritz group of S^3 is stably true 3775

MARTIN SCHARLEMANN