AGT

Algebraic & Geometric Topology

Volume 23 (2023)

Issue 8 (pages 3417–3908)

msp

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

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

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

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Algebraic & Geometric Topology 23:8 (2023) 3417–3496
DOI: 10.2140/agt.2023.23.3417
Published: 5 November 2023

Partial Torelli groups and homological stability

ANDREW PUTMAN

We prove a homological stability theorem for the subgroup of the mapping class group acting as the identity on some fixed portion of the first homology group of the surface. We also prove a similar theorem for the subgroup of the mapping class group preserving a fixed map from the fundamental group to a finite group, which can be viewed as a mapping class group version of a theorem of Ellenberg, Venkatesh and Westerland about braid groups. These results require studying various simplicial complexes formed by subsurfaces of the surface, generalizing work of Hatcher and Vogtmann.

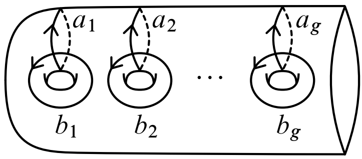
57K20

1 Introduction

Let Σ_g^b be an oriented genus-g surface with b boundary components. The *mapping class group* $\mathrm{Mod}(\Sigma_g^b)$ is the group of isotopy classes of orientation-preserving homeomorphisms of Σ_g^b that fix $\partial\Sigma_g^b$ pointwise. Harer [11] proved that $\mathrm{Mod}(\Sigma_g^b)$ satisfies homological stability. More precisely, an orientation-preserving embedding $\Sigma_g^b \hookrightarrow \Sigma_{g'}^{b'}$ induces a map $\mathrm{Mod}(\Sigma_g^b) \to \mathrm{Mod}(\Sigma_{g'}^{b'})$ that extends mapping classes by the identity, and Harer's theorem says that the induced map $\mathrm{H}_k(\mathrm{Mod}(\Sigma_g^b)) \to \mathrm{H}_k(\mathrm{Mod}(\Sigma_{g'}^{b'}))$ is an isomorphism if $g \gg k$.

Torelli The group $\mathrm{Mod}(\Sigma_g^b)$ acts on $\mathrm{H}_1(\Sigma_g^b)$. For $b \le 1$, the algebraic intersection pairing on $\mathrm{H}_1(\Sigma_g^b)$ is a $\mathrm{Mod}(\Sigma_g^b)$–invariant symplectic form. We thus get a map $\mathrm{Mod}(\Sigma_g^b) \to \mathrm{Sp}_{2g}(\mathbb{Z})$ whose kernel $\mathcal{I}(\Sigma_g^b)$ is the *Torelli group*. The group $\mathcal{I}(\Sigma_g^b)$ is not homologically stable; indeed, Johnson [16] showed that $\mathrm{H}_1(\mathcal{I}(\Sigma_g^b))$ does not stabilize. Church and Farb's work on representation stability [4] connects this to the $\mathrm{Sp}_{2g}(\mathbb{Z})$–action on $\mathrm{H}_k(\mathcal{I}(\Sigma_g^b))$ induced by the conjugation action of $\mathrm{Mod}(\Sigma_g^b)$. Much recent work on $\mathrm{H}_k(\mathcal{I}(\Sigma_g^b))$ focuses on this action; see Boldsen and Hauge Dollerup [2], Kassabov and Putman [18] and Miller, Patzt and Wilson [21].

Partial Torelli We show that homological stability can be restored by enlarging the Torelli group to the group acting trivially on some fixed portion of homology. As an illustration of our results, we begin by describing them in a very special case. Fix a symplectic basis $\{a_1, b_1, \ldots, a_g, b_g\}$ for $\mathrm{H}_1(\Sigma_g^1)$ in the usual way:



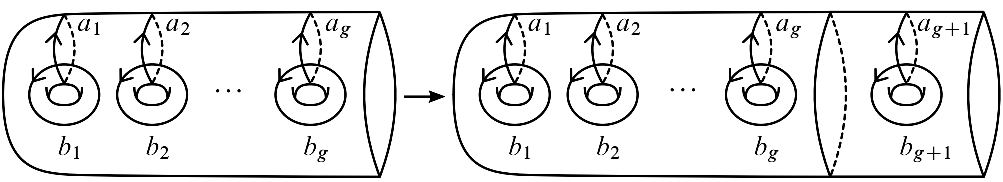
For $0 \le h \le g$, define $\mathcal{I}(\Sigma_g^1, h)$ to be the subgroup of $\mathrm{Mod}(\Sigma_g^1)$ fixing all elements of $\{a_1, b_1, \ldots, a_h, b_h\}$. These groups interpolate between $\mathrm{Mod}(\Sigma_g^1)$ and $\mathcal{I}(\Sigma_g^1)$ in the sense that

$$\mathcal{I}(\Sigma_g^1) = \mathcal{I}(\Sigma_g^1, g) \subset \mathcal{I}(\Sigma_g^1, g-1) \subset \mathcal{I}(\Sigma_g^1, g-2) \subset \cdots \subset \mathcal{I}(\Sigma_g^1, 0) = \mathrm{Mod}(\Sigma_g^1).$$

They were introduced by Bestvina, Bux and Margalit [1]; see especially [1, Conjecture 1.2]. For a fixed $h \ge 1$, we have an increasing chain of groups

$$\mathcal{I}(\Sigma_h^1, h) \subset \mathcal{I}(\Sigma_{h+1}^1, h) \subset \mathcal{I}(\Sigma_{h+2}^1, h) \subset \cdots, \tag{1-1}$$

where $\mathcal{I}(\Sigma_g^1, h)$ is embedded in $\mathcal{I}(\Sigma_{g+1}^1, h)$ via



Our main theorem shows that (1-1) satisfies homological stability: for $h, k \ge 1$, we have

$$\mathrm{H}_k(\mathcal{I}(\Sigma_g^1, h)) \cong \mathrm{H}_k(\mathcal{I}(\Sigma_{g+1}^1, h))$$

for $g \ge (2h+2)k + (4h+2)$.

Homology markings To state our more general result, we need the notion of a homology marking. Let A be a finitely generated abelian group. An *A–homology marking* on Σ_g^1 is a homomorphism $\mu \colon \mathrm{H}_1(\Sigma_g^1) \to A$. Associated to this is a *partial Torelli group*

$$\mathcal{I}(\Sigma_g^1, \mu) = \{f \in \mathrm{Mod}(\Sigma_g^1) \mid \mu(f(x)) = \mu(x) \text{ for all } x \in \mathrm{H}_1(\Sigma_g^1)\}.$$

Example 1.1 If $A = \mathrm{H}_1(\Sigma_g^1)$ and $\mu = \mathrm{id}$, then $\mathcal{I}(\Sigma_g^1, \mu) = \mathcal{I}(\Sigma_g^1)$.

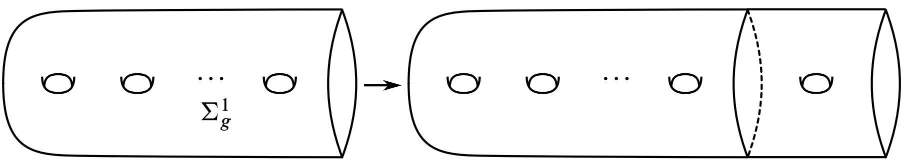
Example 1.2 If $A = \mathrm{H}_1(\Sigma_g^1; \mathbb{Z}/\ell)$ and $\mu\colon \mathrm{H}_1(\Sigma_g^1) \to A$ is the projection, then $\mathcal{I}(\Sigma_g^1, \mu)$ is the *level-ℓ subgroup* of $\mathrm{Mod}(\Sigma_g^1)$, ie the kernel of the action of $\mathrm{Mod}(\Sigma_g^1)$ on $\mathrm{H}_1(\Sigma_g^1; \mathbb{Z}/\ell)$.

Example 1.3 Let A be a *symplectic subspace* of $\mathrm{H}_1(\Sigma_g^1)$, ie a subspace with $\mathrm{H}_1(\Sigma_g^1) = A \oplus A^{\perp}$, where $\perp$ is defined via the intersection form. Such an A is of the form $A \cong \mathbb{Z}^{2h}$ for some $h \geq 0$, called the *genus* of A. If $\mu\colon \mathrm{H}_1(\Sigma_g^1) \to A$ is the projection, then

$$\mathcal{I}(\Sigma_g^1, \mu) = \{f \in \mathrm{Mod}(\Sigma_g^1) \mid f(x) = x \text{ for all } x \in A\}.$$

If A has genus h, then $\mathcal{I}(\Sigma_g^1, \mu) \cong \mathcal{I}(\Sigma_g^1, h)$.

Stability Our first main theorem is a homological stability theorem for the groups $\mathcal{I}(\Sigma_g^1, \mu)$. Define the *stabilization to* Σ_{g+1}^1 of an A–homology marking μ on Σ_g^1 to be the following A–homology marking μ' on Σ_{g+1}^1. Embed Σ_g^1 in Σ_{g+1}^1 just like we did above:



This identifies $\mathrm{H}_1(\Sigma_g^1)$ with a symplectic subspace of $\mathrm{H}_1(\Sigma_{g+1}^1)$, so $\mathrm{H}_1(\Sigma_{g+1}^1) = \mathrm{H}_1(\Sigma_g^1) \oplus \mathrm{H}_1(\Sigma_g^1)^{\perp}$. Let $\mu'\colon \mathrm{H}_1(\Sigma_{g+1}^1) \to A$ be the composition

$$\mathrm{H}_1(\Sigma_{g+1}^1) = \mathrm{H}_1(\Sigma_g^1) \oplus \mathrm{H}_1(\Sigma_g^1)^{\perp} \to \mathrm{H}_1(\Sigma_g^1) \xrightarrow{\mu} A,$$

where the first arrow is the orthogonal projection. The map $\mathrm{Mod}(\Sigma_g^1) \to \mathrm{Mod}(\Sigma_{g+1}^1)$ induced by the above embedding restricts to a map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g+1}^1, \mu')$, called the *stabilization map*. For a finitely generated abelian group A, let $\mathrm{rk}(A)$ denote the minimum size of a generating set[1] for A. Our main theorem is as follows:

Theorem A *Let A be a finitely generated abelian group, μ an A–homology marking on Σ_g^1, and μ' its stabilization to Σ_{g+1}^1. The map $\mathrm{H}_k(\mathcal{I}(\Sigma_g^1, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma_{g+1}^1, \mu'))$ induced by the stabilization map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g+1}^1, \mu')$ is an isomorphism if $g \geq (\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$ and a surjection if $g = (\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+1)$.*

[1]Equivalently, $\mathrm{rk}(A)$ is the maximum $n \geq 0$ such that A is the direct sum of n cyclic subgroups. There are several different commonly used definitions of the rank of an abelian group, and we emphasize that our $\mathrm{rk}(A)$ is *not* the maximum n such that A contains a subgroup isomorphic to $\mathbb{Z}^n$. In particular, $\mathrm{rk}(\mathbb{Z}/\ell) = 1$ for $\ell \geq 2$, and $\mathrm{rk}(A) = 0$ if and only if $A = 0$.

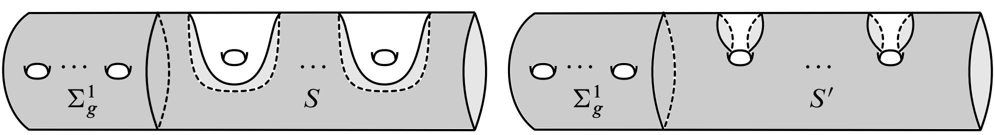
Closed surface trouble Harer's stability theorem implies that the map $\mathrm{Mod}(\Sigma_g^1) \to \mathrm{Mod}(\Sigma_g)$ arising from gluing a disc to $\partial\Sigma_g^1$ induces an isomorphism on H_k for $g \gg k$. One might expect a similar result to hold for the partial Torelli groups. Unfortunately, this is completely false. In Section 7, we will prove that it fails even for H_1 for A–homology markings satisfying a mild nondegeneracy condition called *symplectic nondegeneracy*. One special case of this is the following. For $1 \le h \le g$, define $\mathcal{I}(\Sigma_g, h)$ just like $\mathcal{I}(\Sigma_g^1, h)$, so we have a surjection $\mathcal{I}(\Sigma_g^1, h) \to \mathcal{I}(\Sigma_g, h)$.

Theorem B *For $h \le g$ with $g \ge 3$ and $h \ge 2$, the map* $\mathrm{H}_1(\mathcal{I}(\Sigma_g^1, h)) \to \mathrm{H}_1(\mathcal{I}(\Sigma_g, h))$ *is not an isomorphism.*

The proof uses an extension of the Johnson homomorphism to the partial Torelli groups that was constructed by Broaddus, Farb and Putman [3].

Multiple boundary components In addition to Theorem A, which concerns surfaces with one boundary component, we have a theorem for surfaces with multiple boundary components. The correct statement here is a bit subtle, since the phenomenon underlying Theorem B also obstructs many obvious kinds of generalizations. The purpose of having a generalization like this is to understand how the partial Torelli groups restrict to subsurfaces, which turns out to be fundamental in the author's forthcoming work on the cohomology of the moduli space of curves with level structures [23]. Here is an example of the kind of result we prove; in fact, this is precisely the special case needed in [23].

Example 1.4 Consider an A–homology marking μ on Σ_g^1. For some $h \ge 1$, let μ' be its stabilization to Σ_{g+h}^1. Consider the following subsurfaces $S \cong \Sigma_g^{1+h}$ and $S' \cong \Sigma_g^{1+2h}$ of Σ_{g+h}^1:



Both S and S' include the entire shaded subsurface (including Σ_g^1). The inclusions $S \hookrightarrow \Sigma_{g+h}^1$ and $S' \hookrightarrow \Sigma_{g+h}^1$ induce homomorphisms $\phi\colon \mathrm{Mod}(S) \to \mathrm{Mod}(\Sigma_{g+h}^1)$ and $\psi\colon \mathrm{Mod}(S') \to \mathrm{Mod}(\Sigma_{g+h}^1)$. Then define $\mathcal{I}(S, \mu') = \phi^{-1}(\mathcal{I}(\Sigma_{g+h}^1, \mu'))$ and $\mathcal{I}(S', \mu') = \psi^{-1}(\mathcal{I}(\Sigma_{g+h}^1, \mu'))$. Be warned: while it turns out that $\mathcal{I}(S, \mu')$ can be defined using the action of $\mathrm{Mod}(S)$ on $\mathrm{H}_1(S)$, the group $\mathcal{I}(S', \mu')$ cannot be defined using only $\mathrm{H}_1(S')$. Then our theorem will show that the map

$$\mathrm{H}_k(\mathcal{I}(S, \mu')) \to \mathrm{H}_k(\mathcal{I}(S', \mu'))$$

is an isomorphism if the genus of S (namely g) is at least $(\mathrm{rk}(A)+2)k+(2\,\mathrm{rk}(A)+2)$. However, except in degenerate cases, the maps

$$\mathrm{H}_1(\mathcal{I}(\Sigma_g^1,\mu)) \to \mathrm{H}_1(\mathcal{I}(S,\mu')) \quad \text{and} \quad \mathrm{H}_1(\mathcal{I}(S,\mu')) \to \mathrm{H}_1(\mathcal{I}(\Sigma_{g+h}^1,\mu'))$$

are never isomorphisms, no matter how large g is.

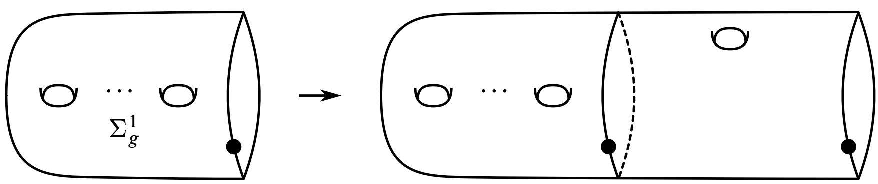
In the above example, we defined the partial Torelli groups on surfaces with multiple boundary components in an ad hoc way. Correctly formulating our theorem requires a more intrinsic definition, and we define a category of "homology-marked surfaces" with multiple boundary components that is inspired by the author's work on the Torelli group on surfaces with multiple boundary components in [22].

Nonabelian markings We also have a theorem for nonabelian markings, whose definition is as follows.[2] Fix a basepoint $* \in \partial\Sigma_g^1$. For a group Λ, a Λ*–marking* on Σ_g^1 is a group homomorphism $\mu\colon \pi_1(\Sigma_g^1,*) \to \Lambda$. If Λ is abelian, then this is equivalent to a Λ–homology marking on Σ_g^1. Given a Λ–marking $\mu\colon \pi_1(\Sigma_g^1,*) \to \Lambda$, define the associated partial Torelli group via the formula

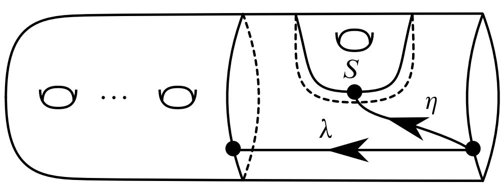
$$\mathcal{I}(\Sigma_g^1,\mu) = \{f \in \mathrm{Mod}(\Sigma_g^1) \mid \mu(f(x)) = \mu(x) \text{ for all } x \in \pi_1(\Sigma_g^1,*)\}.$$

Again, this reduces to our previous definition if Λ is abelian.

Nonabelian stabilization Let μ be a Λ–marking on Σ_g^1. Due to basepoint issues, stabilizing μ to Σ_{g+1}^1 is a little more complicated than the case of homology markings. Let $* \in \partial\Sigma_g^1$ and $*' \in \partial\Sigma_{g+1}^1$ be the basepoints. Embed Σ_g^1 into Σ_{g+1}^1 as in



Let λ, η and $S \cong \Sigma_1^1$ be as in



[2] I am not sure who first defined this concept. Related things appear eg in work of Dunfield and Thurston [6] and Ivanov [13].

Letting $*'' \in \partial S$ be the basepoint of S as above, the paths λ and η induce injective homomorphisms

$$\pi_1(\Sigma_g^1, *) \hookrightarrow \pi_1(\Sigma_{g+1}^1, *') \quad \text{and} \quad \pi_1(S, *'') \hookrightarrow \pi_1(\Sigma_{g+1}^1, *')$$

taking $x \in \pi_1(\Sigma_g^1, *)$ to $\lambda \cdot x \cdot \lambda^{-1} \in \pi_1(\Sigma_{g+1}^1, *')$ and $y \in \pi_1(S, *'')$ to $\eta \cdot y \cdot \eta^{-1} \in \pi_1(\Sigma_{g+1}^1, *')$. Identifying $\pi_1(\Sigma_g^1, *)$ and $\pi_1(S, *'')$ with the corresponding subgroups of $\pi_1(\Sigma_{g+1}^1, *')$, we have a free product decomposition

$$\pi_1(\Sigma_{g+1}^1, *') = \pi_1(\Sigma_g^1, *) \star \pi_1(S, *'').$$

Then define the stabilization $\mu' \colon \pi_1(\Sigma_{g+1}^1, *') \to \Lambda$ of $\mu \colon \pi_1(\Sigma_g^1, *) \to \Lambda$ to be the composition

$$\pi_1(\Sigma_{g+1}^1, *') = \pi_1(\Sigma_g^1, *) \star \pi_1(S, *'') \to \pi_1(\Sigma_g^1, *) \xrightarrow{\mu} \Lambda,$$

where the first arrow quotients out by the normal closure of $\pi_1(S, *'')$. As in the abelian setting, the map $\mathrm{Mod}(\Sigma_g^1) \to \mathrm{Mod}(\Sigma_{g+1}^1)$ induced by our embedding $\Sigma_g^1 \hookrightarrow \Sigma_{g+1}^1$ restricts to a map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g+1}^1, \mu')$, which we will call the *stabilization map*.

Nonabelian stability Our main theorem about this is as follows. It can be viewed as an analogue for the mapping class group of a theorem of Ellenberg, Venkatesh and Westerland [7, Theorem 6.1] concerning braid groups and Hurwitz spaces.

Theorem C *Let Λ be a finite group, μ a Λ–marking on Σ_g^1, and μ' its stabilization to Σ_{g+1}^1. The map $\mathrm{H}_k(\mathcal{I}(\Sigma_g^1, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma_{g+1}^1, \mu'))$ induced by the stabilization map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g+1}^1, \mu')$ is an isomorphism if $g \geq (|\Lambda| + 2)k + (2|\Lambda| + 2)$ and a surjection if $g = (|\Lambda| + 2)k + (2|\Lambda| + 1)$.*

Remark 1.5 Ellenberg, Venkatesh and Westerland's main application in [7] of their stability result concerns point-counting in Hurwitz spaces via the Weil conjectures. Unfortunately, the vast amount of unknown unstable cohomology precludes such applications here.

Remark 1.6 If Λ is a finite abelian group, then Theorems A and C give a similar kind of stability, but the bounds in Theorem A are much stronger.

Remark 1.7 Because of basepoint issues, stating a version of Theorem C on surfaces with multiple boundary components would be rather technical, and unlike for Theorem A we do not know any potential applications of such a result. We thus do not pursue this kind of generalization of Theorem C.

Proof techniques There is an enormous literature on homological stability theorems, starting with unpublished work of Quillen on $\mathrm{GL}_n(\mathbb{F}_p)$. A standard proof technique has emerged that first appeared in its modern formulation in work of van der Kallen [17]. Consider a sequence of groups

$$G_0 \subset G_1 \subset G_2 \subset \cdots \tag{1-2}$$

that we want to prove enjoys homological stability, ie $\mathrm{H}_k(G_{n-1}) \cong \mathrm{H}_k(G_n)$ for $n \gg k$. To compute $\mathrm{H}_k(G_n)$, we would need a contractible simplicial complex on which G_n acts freely. Since we are only interested in the low-degree homology groups, we can weaken contractibility to high connectivity. The key insight for homological stability is that since we only want to compare $\mathrm{H}_k(G_n)$ with the homology of previous groups in (1-2), what we want is not a free action but one whose stabilizer subgroups are related to the previous groups.

Machine There are many variants on the above machine. For proving homological stability for the groups G_n in (1-2), the easiest version requires simplicial complexes X_n upon which G_n acts with the following three properties:

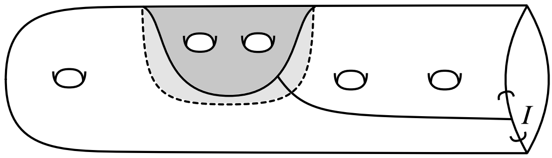
- The connectivity of X_n goes to ∞ as $n \mapsto \infty$.
- For $0 \le k \le n-1$, the G_n–stabilizer of a k–simplex of X_n is conjugate to G_{n-k-1}.
- The group G_n acts transitively on the k–simplices of X_n for all $k \ge 0$.

Some additional technical hypotheses are needed; we will review these in Section 3.1. Hatcher and Vogtmann [12] constructed such X_n for the mapping class group. Our proof of Theorem A is inspired by their work, so we start by describing a variant of it.

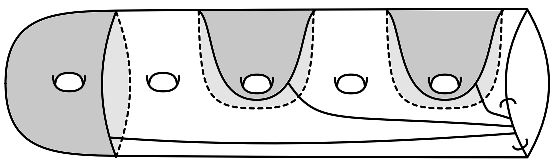
Subsurface complex For $h \ge 1$, the *complex of genus-h subsurfaces* of Σ_g^b, denoted by $\mathcal{S}_h(\Sigma_g^b)$, is the simplicial complex whose k–simplices are sets $\{\iota_0, \ldots, \iota_k\}$ of isotopy classes of orientation-preserving embeddings $\iota_i\colon \Sigma_h^1 \to \Sigma_g^b$ that can be isotoped so that, for $0 \le i < j \le k$, the subsurfaces $\iota_i(\Sigma_h^1)$ and $\iota_j(\Sigma_h^1)$ are disjoint. The group $\mathrm{Mod}(\Sigma_g^b)$ acts on $\mathcal{S}_h(\Sigma_g^b)$. However, it turns out that this is not quite the right complex for homological stability.

Tethered subsurfaces Let $\tau(\Sigma_h^1)$ be the result of gluing the interval $[0,1]$ to Σ_h^1 by identifying $1 \in [0,1]$ with a point of $\partial\Sigma_h^1$. The subset $[0,1] \subset \tau(\Sigma_h^1)$ is the *tether* and $0 \in [0,1] \subset \tau(\Sigma_h^1)$ the *initial point* of the tether. Let $I \subset \partial\Sigma_g^b$ be a finite disjoint union of open intervals. An *I–tethered genus-h subsurface* of Σ_g^b is an embedding

$\iota\colon \tau(\Sigma_h^1) \to \Sigma_g^b$ taking the initial point of the tether to a point of I whose restriction to Σ_h^1 preserves the orientation. For instance, here is an I–tethered genus-2 subsurface:



Tethered subsurface complex The *complex of I–tethered genus-h subsurfaces* of Σ_g^b, denoted by $\mathcal{TS}_h(\Sigma_g^b, I)$, is the simplicial complex whose k–simplices are collections $\{\iota_0, \dots, \iota_k\}$ of isotopy classes of I–tethered genus-h subsurfaces of Σ_g^b that can be realized disjointly. These isotopies are allowed to move the images of the initial points of the tethers within I, so the tethers can be slid past each other and made disjoint. For instance, here is a 2–simplex in $\mathcal{TS}_1(\Sigma_5^1, I)$:



High connectivity The complexes $\mathcal{S}_1(\Sigma_g^b)$ and $\mathcal{TS}_1(\Sigma_g^b, I)$ are closely related to complexes that were introduced by Hatcher and Vogtmann [12], and it follows easily from their work that they are $\frac{1}{2}(g-3)$–connected (see Putman and Sam [24, proof of Theorem 6.25] for details). We generalize this as follows:

Theorem D *Consider $g \geq h \geq 1$ and $b \geq 0$.*

- *The complex $\mathcal{S}_h(\Sigma_g^b)$ is $(g-(2h+1))/(h+1)$–connected.*
- *Assume that $b \geq 1$, and let $I \subset \partial\Sigma_g^b$ be a finite disjoint union of open intervals. The complex $\mathcal{TS}_h(\Sigma_g^b, I)$ is $(g-(2h+1))/(h+1)$–connected.*

Remark 1.8 Our convention is that a space is (-1)–connected if it is nonempty. Using this convention, the genus bounds for (-1)–connectivity and 0–connectivity in Theorem D are sharp. We do not know whether they are sharp for higher connectivity.

Remark 1.9 Hatcher and Vogtmann's proof in [12] that $\mathcal{S}_1(\Sigma_g^b)$ and $\mathcal{TS}_1(\Sigma_g^b, I)$ are $\frac{1}{2}(g-3)$–connected is closely connected to their proof that the separating curve complex is $\frac{1}{2}(g-3)$–connected. Looijenga [20] later showed that the separating curve complex is $(g-3)$–connected. Unfortunately, his techniques do not appear to give an improvement to Theorem D.

Remark 1.10 In applications to homological stability, we will only use complexes made out of genus-1 subsurfaces. However, the more general result of Theorem D will be needed for the proof even of the $h = 1$ case of Theorem E below.

Mod stability Consider the groups

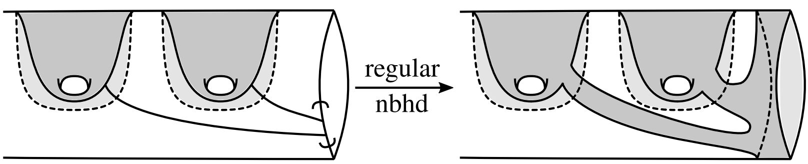
$$\mathrm{Mod}(\Sigma_1^1) \subset \mathrm{Mod}(\Sigma_2^1) \subset \mathrm{Mod}(\Sigma_3^1) \subset \cdots. \tag{1-3}$$

Let $I \subset \partial\Sigma_g^1$ be an open interval. The group $\mathrm{Mod}(\Sigma_g^1)$ acts on $\mathcal{TS}_1(\Sigma_g^1, I)$, and this complex has all three properties needed by the machine to prove homological stability for (1-3):

- As we said above, $\mathcal{TS}_1(\Sigma_g^1, I)$ is $\frac{1}{2}(g-3)$–connected.
- The $\mathrm{Mod}(\Sigma_g^1)$–stabilizer of a k–simplex $\{\iota_0, \ldots, \iota_k\}$ of $\mathcal{TS}_1(\Sigma_g^1, I)$ is the mapping class group of the complement of a regular neighborhood of

$$\partial\Sigma_g^1 \cup \iota_0(\tau(\Sigma_1^1)) \cup \cdots \cup \iota_k(\tau(\Sigma_1^1)).$$

See here:



This complement is homeomorphic to Σ_{g-k-1}^1, so this stabilizer is isomorphic to $\mathrm{Mod}(\Sigma_{g-k-1}^1)$. All such subsurface mapping class groups are conjugate; this follows from the change of coordinates principle of Farb and Margalit [8, Section 1.3.2].

- Another application of the change of coordinates principle shows that $\mathrm{Mod}(\Sigma_g^1)$ acts transitively on the k–simplices of $\mathcal{TS}_1(\Sigma_g^1, I)$.

Partial Torelli problem A first idea for proving homological stability for the partial Torelli groups $\mathcal{I}(\Sigma_g^1, \mu)$ is to consider their actions on $\mathcal{TS}_1(\Sigma_g^1, I)$. Unfortunately, this does not work. The fundamental problem is that $\mathcal{I}(\Sigma_g^1, \mu)$ does not act transitively on the k–simplices of $\mathcal{TS}_1(\Sigma_g^1, I)$; indeed, it does not even act transitively on the vertices. For A–homology markings μ, the issue is that, for an I–tethered torus $\iota\colon \tau(\Sigma_1^1) \to \Sigma_g^1$ and $f \in \mathcal{I}(\Sigma_g^1, \mu)$, the compositions

$$\mathrm{H}_1(\Sigma_1^1) \cong \mathrm{H}_1(\tau(\Sigma_1^1)) \xrightarrow{\iota_*} \mathrm{H}_1(\Sigma_g^1) \xrightarrow{\mu} A, \quad \mathrm{H}_1(\Sigma_1^1) \cong \mathrm{H}_1(\tau(\Sigma_1^1)) \xrightarrow{(f\circ\iota)_*} \mathrm{H}_1(\Sigma_g^1) \xrightarrow{\mu} A$$

will be the same, but the functions $\mu \circ \iota_*\colon \mathrm{H}_1(\tau(\Sigma_1^1)) \to A$ need not be the same for different tethered tori. A similar issue arises in the nonabelian setting. To fix this, we use a subcomplex of $\mathcal{TS}_1(\Sigma_g^1, I)$ that is adapted to μ.

Remark 1.11 The stabilizers are also wrong, but fixing the transitivity will also fix this.

Vanishing surfaces For an A–homology marking μ on Σ_g^1, define $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ to be the full subcomplex of $\mathcal{TS}_h(\Sigma_g^1, I)$ spanned by vertices ι such that the composition

$$\mathrm{H}_1(\tau(\Sigma_h^1)) \xrightarrow{\iota_*} \mathrm{H}_1(\Sigma_g^1) \xrightarrow{\mu} A$$

is the zero map. We will show that $\mathcal{I}(\Sigma_g^1, \mu)$ acts transitively on the k–simplices of $\mathcal{TS}_1(\Sigma_g^1, I, \mu)$, at least for k not too large. However, there is a problem: a priori the subcomplex $\mathcal{TS}_1(\Sigma_g^1, I, \mu)$ of $\mathcal{TS}_1(\Sigma_g^1, I)$ might not be highly connected. Our third main theorem says that in fact it is $\bigl(g-(2\,\mathrm{rk}(A)+3)\bigr)/(\mathrm{rk}(A)+2)$–connected. More generally, we prove the following:

Theorem E *Let A be a finitely generated abelian group, let μ be an A–homology marking on Σ_g^1, and let $I \subset \partial\Sigma_g^1$ be a finite disjoint union of open intervals. Then the complex $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ is $\bigl(g-(2\,\mathrm{rk}(A)+2h+1)\bigr)/(\mathrm{rk}(A)+h+1)$–connected.*

We also prove a similar theorem in the nonabelian setting.

Outline We start in Section 2 by proving Theorem D. We then prove Theorems A, C, and E in Section 3. Next, in Section 4 we define a category of homology-marked surfaces with multiple boundary components. In Section 5 we use our category to state and prove Theorem F, which generalizes Theorem A to surfaces with multiple boundary components. This proof depends on a stabilization result which is proved in Section 6. We close with Section 7, which proves Theorem B.

Conventions Throughout this paper, A denotes a fixed finitely generated abelian group and Λ is a fixed finite group. We also fix a basepoint $* \in \partial\Sigma_g^1$.

Acknowledgments I want to thank Jordan Ellenberg for a useful correspondence and Allen Hatcher for pointing out a confusing typo. I also want to thank the referee for their very careful reading of the paper and many helpful suggestions. This work was supported in part by NSF grant DMS-1811322

2 The complex of subsurfaces

This section is devoted to the proof of Theorem D, which asserts that $\mathcal{S}_h(\Sigma_g^b)$ and $\mathcal{TS}_h(\Sigma_g^b, I)$ are highly connected. There are three parts: Section 2.1 contains a technical result about fibers of maps, Section 2.2 discusses "link arguments", and Section 2.3 proves Theorem D.

2.1 Fibers of maps

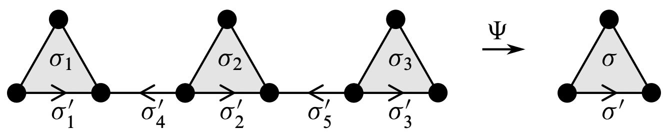
Our proofs will require a technical tool:

Homotopy theory conventions A space X is said to be n–connected if, for $k \leq n$, all maps $S^k \to X$ extend to maps $D^{k+1} \to X$. Since $S^{-1} = \varnothing$ and D^0 is a single point, a space is (-1)–connected precisely when it is nonempty. A map $\psi : X \to Y$ of spaces is an n–homotopy equivalence if, for all $0 \leq k \leq n$, the induced map $[S^k, X] \to [S^k, Y]$ on unbased homotopy classes of maps out of S^k is a bijection. This is equivalent to saying that the induced map on π_k is a bijection for each choice of basepoint.

Relative fibers If $\psi : X \to Y$ is a map of simplicial complexes, σ is a simplex of Y, and σ' is a face of σ, then denote by $\mathrm{Fib}_\psi(\sigma', \sigma)$ the subcomplex of X consisting of all simplices η' of X such that

- $\psi(\eta')$ is a face of σ', and
- there exists a simplex η of X such that η' is a face of η and $\psi(\eta) = \sigma$.

For instance, consider the following map, where ψ takes each 1–simplex σ'_i to σ' (with the specified orientation) and each 2–simplex σ_i to σ:



The relative fiber $\mathrm{Fib}_\psi(\sigma', \sigma)$ then consists of σ'_1 and σ'_2 and σ'_3 (but not σ'_4 or σ'_5).

Fiber lemma With these definitions, we have the following lemma:

Lemma 2.1 *Let $\psi : X \to Y$ be a map of simplicial complexes. For some $n \geq 0$, assume the space $\mathrm{Fib}_\psi(\sigma', \sigma)$ is n–connected for all simplices σ of Y and all faces σ' of σ. Then ψ is an n–homotopy equivalence.*

Proof Replacing Y by its $(n+1)$–skeleton Y_{n+1} and X by $\psi^{-1}(Y_{n+1})$, we can assume that Y is finite dimensional. The proof will be by induction on $m = \dim(Y)$. The base case $m = 0$ is trivial, since in that case Y is a discrete set of points and our assumptions imply that the fiber over each of these points is n–connected. Assume now that $m \geq 1$. The key step in the proof is the following claim:

Claim *Assume that Y is the union of a subcomplex Y' and an m–simplex σ with $\sigma \cap Y' = \partial\sigma$. Define $X' = \psi^{-1}(Y')$, and assume that $\psi : X \to Y$ restricts to an n–homotopy equivalence $\psi' : X' \to Y'$. Then ψ is an n–homotopy equivalence.*

Proof Let $X'' = \mathrm{Fib}_{\psi}(\sigma,\sigma)$. In other words, X'' consists of all simplices of X mapping surjectively onto σ, along with their faces. We thus have $X = X' \cup X''$. By assumption, X'' is n–connected, which implies that ψ restricts to an n–homotopy equivalence $\psi''\colon X'' \to \sigma$. Define $Z = X' \cap X''$. The map ψ restricts to a map $\psi_Z\colon Z \to \partial\sigma$.

We now come to the key observation: the space Z is precisely the subcomplex of X consisting of the union of the subcomplexes $\mathrm{Fib}_{\psi}(\sigma',\sigma)$ as σ' ranges over all simplices of $\partial\sigma$. Moreover, for all simplices σ' of $\partial\sigma$ and all faces σ'' of σ', we have $\mathrm{Fib}_{\psi_Z}(\sigma'',\sigma') = \mathrm{Fib}_{\psi}(\sigma'',\sigma)$, and thus by assumption $\mathrm{Fib}_{\psi_Z}(\sigma'',\sigma')$ is n–connected. We can therefore apply our inductive hypothesis to see that $\psi_Z\colon Z \to \partial\sigma \cong S^{m-1}$ is an n–homotopy equivalence.

Summing up, we have $X = X' \cup X''$ and $Y = Y' \cup \sigma$. The map ψ restricts to n–homotopy equivalences

$$\psi'\colon X' \to Y', \quad \psi''\colon X'' \to \sigma, \quad \text{and} \quad \psi_Z\colon X' \cap X'' = Z \to \partial\sigma = Y' \cap \sigma,$$

and induces a map between the Mayer–Vietoris exact sequences associated to the decompositions $X = X' \cup X''$ and $Y = Y' \cup \sigma$:

$$\begin{tikzcd}
\cdots \arrow[r] & \mathrm{H}_k(X' \cap X'') \arrow[r] \arrow[d] & \mathrm{H}_k(X') \oplus \mathrm{H}_k(X'') \arrow[r] \arrow[d] & \mathrm{H}_k(X) \arrow[r] \arrow[d] & \cdots \\
\cdots \arrow[r] & \mathrm{H}_k(Y' \cap \sigma) \arrow[r] & \mathrm{H}_k(Y') \oplus \mathrm{H}_k(\sigma) \arrow[r] & \mathrm{H}_k(Y) \arrow[r] & \cdots
\end{tikzcd}$$

Other than the maps $\mathrm{H}_k(X) \to \mathrm{H}_k(Y)$, the vertical maps in this commutative diagram are isomorphisms for $k \le n$, so by the five lemma the maps $\mathrm{H}_k(X) \to \mathrm{H}_k(Y)$ are also isomorphisms for $k \le n$. This implies in particular that the map $X \to Y$ is 0–connected, and thus induces a bijection between path-components. If $n \ge 1$, then a similar argument on each path component using the Seifert–van Kampen theorem shows that the map $\psi\colon X \to Y$ induces an isomorphism on π_1 for each choice of basepoint. This allows us to identify local coefficient systems on Y with local coefficient systems on X, and for each local coefficient system A on Y we can run the above Mayer–Vietoris argument on homology with coefficients in A to prove that the map $\psi\colon \mathrm{H}_k(X;A) \to \mathrm{H}_k(Y;A)$ is an isomorphism for $k \le n$. Applying the nonsimply connected version of Whitehead's theorem [5, Theorem 6.71], we deduce that the map $X \to Y$ is an n–homotopy equivalence, as desired. $\square$

Repeatedly applying this claim, we see that the lemma holds for m–dimensional Y with finitely many m–simplices.

The general case reduces to the case where Y has finitely many m–simplices as follows. Consider some $0 \le k \le n$. Our goal is to prove that the map $[S^k, X] \to [S^k, Y]$ induced by ψ is a bijection. The proofs that it is injective and surjective are similar compactness arguments, so we give the details for surjectivity and leave injectivity to the reader.

Consider a map $f\colon S^k \to Y$. By compactness, $f(S^k)$ lies in a subcomplex of Y' of Y containing the $(m-1)$–skeleton and finitely many m–simplices. Letting $X' = \psi^{-1}(Y')$, we know that the map $[S^k, X'] \to [S^k, Y']$ is a bijection, so there exists some $\tilde{f}\colon S^k \to X'$ such that $\psi \circ \tilde{f}\colon S^k \to Y'$ is homotopic to f. It follows that the map $[S^k, X] \to [S^k, Y]$ induced by ψ is surjective, as desired. □

Corollary 2.2 *Let $\psi\colon X \to Y$ be a map of simplicial complexes. For some $n \ge 0$, assume*

- *Y is n–connected, and*
- *all $(n+1)$–simplices of Y are in the image of ψ, and*
- *for all simplices σ of Y whose dimension is at most n and all faces σ' of σ, the space $\mathrm{Fib}_\psi(\sigma', \sigma)$ is n–connected.*

Then X is n–connected.

Proof Let Y' be the n–skeleton of Y and $X' = \psi^{-1}(Y')$, so X' contains the n–skeleton of X. Let $\psi'\colon X' \to Y'$ be the restriction of ψ to X'. Our assumptions allow us to apply Lemma 2.1 to ψ', so ψ' is an n–homotopy equivalence. Since Y is n–connected the space Y' is $(n-1)$–connected, so this implies that X' and thus X are $(n-1)$–connected. We also know that the induced map $\psi'\colon \pi_n(X') \to \pi_n(Y')$ is an isomorphism. Since Y is n–connected, attaching the $(n+1)$–simplices of Y to Y' kills $\pi_n(Y')$. By assumption, for each of these $(n+1)$–simplices σ of Y there is an $(n+1)$–simplex $\tilde{\sigma}$ of X such that $\psi(\tilde{\sigma}) = \sigma$. It follows that the element of $\pi_n(Y')$ represented by $\partial\sigma \to Y'$ lifts to the element of $\pi_n(X')$ represented by $\partial\tilde{\sigma} \to X'$. We conclude that attaching to X' the $(n+1)$–simplices of X that do not already lie in X' kills $\pi_n(X')$, which implies that $\pi_n(X) = 0$, as desired. □

2.2 Link arguments

Let X be a simplicial complex and let $Y \subset X$ be a subcomplex. This section is devoted to a result of Hatcher and Vogtmann [12] that gives conditions under which the pair (X, Y) is n–connected, ie $\pi_k(X, Y) = 0$ for $0 \le k \le n$. The idea is to identify a collection $\mathcal{B}$ of "bad simplices" of X that characterize Y in the sense that a simplex lies in Y precisely when none of its faces lie in $\mathcal{B}$. We then have to understand the local topology of Y around a simplex of $\mathcal{B}$. To that end, if $\mathcal{B}$ is a collection of simplices of X and $\sigma \in \mathcal{B}$, then define $G(X, \sigma, \mathcal{B})$ to be the subcomplex of X consisting of simplices σ' satisfying:

- The join $\sigma * \sigma'$ is a simplex of X, ie σ' is a simplex in the link of σ.
- If σ'' is a face of $\sigma * \sigma'$ such that $\sigma'' \in \mathcal{B}$, then $\sigma'' \subset \sigma$.

Hatcher and Vogtmann's result is then as follows.

Proposition 2.3 [12, Proposition 2.1] *Let Y be a subcomplex of a simplicial complex X and assume that there exists a collection $\mathcal{B}$ of simplices of X satisfying, for some $n \geq 0$:*

(i) *A simplex of X lies in Y if and only if none of its faces lie in $\mathcal{B}$.*

(ii) *If $\sigma_1, \sigma_2 \in \mathcal{B}$ are such that $\sigma_1 \cup \sigma_2$ is a simplex of X, then $\sigma_1 \cup \sigma_2 \in \mathcal{B}$. Here σ_1 and σ_2 might share vertices, so $\sigma_1 \cup \sigma_2$ might not be the join $\sigma_1 * \sigma_2$.*

(iii) *For all k–dimensional $\sigma \in \mathcal{B}$, the complex $G(X, \sigma, \mathcal{B})$ is $(n-k-1)$–connected.*

Then the pair (X, Y) is n–connected.

As an illustration of how Proposition 2.3 might be used, we use it to prove the following result (which will in fact be how we use that proposition in all but two cases).

Corollary 2.4 *Let X be a simplicial complex and let $Y, Y' \subset X$ be disjoint full subcomplexes such that every vertex of X lies in either Y or Y'. For some $n \geq 0$, assume that for all k–dimensional simplices σ of Y' the intersection of Y with the link of σ is $(n-k-1)$–connected. Then the pair (X, Y) is n–connected.*

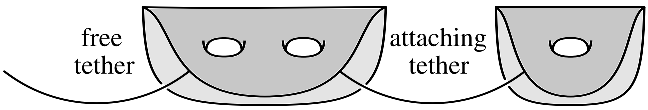
Proof We will verify the hypotheses of Proposition 2.3 for the set $\mathcal{B}$ of all simplices of Y'. Since Y is a full subcomplex of X and all vertices of X lie in either Y or Y', a simplex of X lies in Y if and only if none of its vertices lie in Y'. Hypothesis (i) follows. Hypothesis (ii) is immediate from the fact that Y' is a full subcomplex of X. As for hypothesis (iii), it is immediate from the definitions that, for a simplex $\sigma \in \mathcal{B}$, the complex $G(X, \sigma, \mathcal{B})$ is precisely the intersection of the link of σ with Y. □

2.3 Subsurface complexes

Proof of Theorem D The proofs for $\mathcal{S}_h(\Sigma_g^b)$ and $\mathcal{TS}_h(\Sigma_g^b, I)$ are similar. Keeping track of the tethers introduces a few complications, so we will give the details for $\mathcal{TS}_h(\Sigma_g^b, I)$ and leave $\mathcal{S}_h(\Sigma_g^b)$ to the reader.

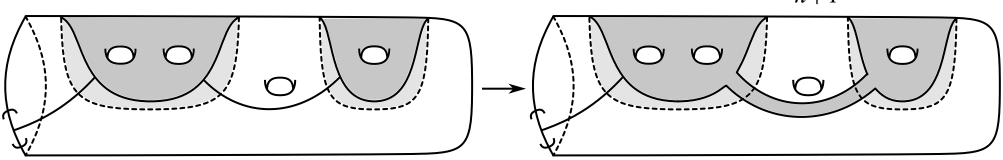
The proof that $\mathcal{TS}_h(\Sigma_g^b, I)$ is $(g-(2h+1))/(h+1)$–connected will be by induction on h. The base case $h=1$ is [24, Theorem 6.25], which we remark shows how to derive it from a closely related result of Hatcher and Vogtmann [12]. For the inductive step, assume that $\mathcal{TS}_h(\Sigma_g^b, I)$ is $(g-(2h+1))/(h+1)$–connected. We will prove that $\mathcal{TS}_{h+1}(\Sigma_g^b, I)$ is $(g-(2h+3))/(h+2)$–connected.

Let $\tau(\Sigma_h^1, \Sigma_1^1)$ be the space obtained from $\tau(\Sigma_h^1) \sqcup \Sigma_1^1$ by gluing in an interval $[0,1]$ with 0 being attached to a point of $\partial\Sigma_h^1$ different from the attaching point of the tether in $\tau(\Sigma_h^1)$ and 1 being attached to a point of $\partial\Sigma_1^1$:



The tether in $\tau(\Sigma_h^1)$ will be called the *free tether* and the interval connecting $\tau(\Sigma_h^1)$ to Σ_1^1 will be called the *attaching tether*. The points 0 of the two tethers will be called their *initial points* and the points 1 will be called their *endpoints*.

Given an embedding $\tau(\Sigma_h^1, \Sigma_1^1) \to \Sigma_g^b$ taking the initial point of the free tether to a point of I, thickening up the attaching tether gives an I–tethered Σ_{h+1}^1:



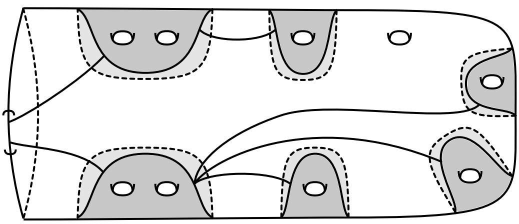
In fact, there is a bijection between isotopy classes of orientation-preserving I–tethered Σ_{h+1}^1 in Σ_g^b and isotopy classes of embeddings $\tau(\Sigma_h^1, \Sigma_1^1) \to \Sigma_g^b$ whose restrictions to Σ_h^1 and Σ_1^1 preserve the orientation and which take the initial point of the free tether to a point of I. For short, we will call these *orientation-preserving I–tethered* $\tau(\Sigma_h^1, \Sigma_1^1)$ in Σ_g^b. We remark that this is slightly awkward terminology, since the free tether is part of $\tau(\Sigma_h^1, \Sigma_1^1)$, while on the other hand we previously talked about I–tethered Σ_{h+1}^1 with the tether implicit. By the above, we can regard $\mathcal{TS}_{h+1}(\Sigma_g^b, I)$ as being the simplicial complex whose k–simplices are collections $\{\iota_0, \dots, \iota_k\}$ of isotopy classes of orientation-preserving I–tethered $\tau(\Sigma_h^1, \Sigma_1^1)$ in Σ_g^b that can be realized so that their images are disjoint.

We now define an auxiliary space. Let X be the simplicial complex whose k–simplices are collections $\{\iota_0, \dots, \iota_k\}$ of isotopy classes of orientation-preserving I–tethered $\tau(\Sigma_h^1, \Sigma_1^1)$ in Σ_g^b that can be realized so that, for all distinct $0 \le i, j \le k$:

- Either $\iota_i|_{\tau(\Sigma_h^1)} = \iota_j|_{\tau(\Sigma_h^1)}$, or the images under ι_i and ι_j of $\tau(\Sigma_h^1)$ are disjoint.

- If $\iota_i|_{\tau(\Sigma_h^1)} = \iota_j|_{\tau(\Sigma_h^1)}$, then the images under ι_i and ι_j of Σ_1^1 together with the attaching tether are disjoint, except for the initial point of the attaching tether.
- If the images under ι_i and ι_j of $\tau(\Sigma_h^1)$ are disjoint, then the images under ι_i and ι_j of $\tau(\Sigma_h^1, \Sigma_1^1)$ are disjoint.

For instance, here is a 3–simplex of X for $g = 9$, $b = 1$ and $h = 2$:



We have $\mathcal{TS}_{h+1}(\Sigma_g^b, I) \subset X$. The next claim says that X enjoys the connectivity property we are trying to prove for $\mathcal{TS}_{h+1}(\Sigma_g^b, I)$:

Claim *X is $(g-(2h+3))/(h+2)$–connected.*

Proof Let $\psi\colon X \to \mathcal{TS}_h(\Sigma_g^b, I)$ be the map that takes a vertex $\iota\colon \tau(\Sigma_h^1, \Sigma_1^1) \to \Sigma_g^b$ of X to the vertex $\iota|_{\tau(\Sigma_h^1)}\colon \tau(\Sigma_h^1) \to \Sigma_g^b$ of $\mathcal{TS}_h(\Sigma_g^b, I)$. We prove that $\psi\colon X \to \mathcal{TS}_h(\Sigma_g^b, I)$ satisfies the conditions of Corollary 2.2 for $n = (g-(2h+3))/(h+2)$. Once we have done this, Corollary 2.2 will show that X is n–connected, as desired.

The first condition is that $\mathcal{TS}_h(\Sigma_g^b, I)$ is n–connected. In fact, our inductive hypothesis says that it is $(g-(2h+1))/(h+1)$–connected, which is even stronger.

The second condition says that all $(n+1)$–simplices of $\mathcal{TS}_h(\Sigma_g^b, I)$ are in the image of ψ. The map ψ is $\mathrm{Mod}(\Sigma_g^b)$–equivariant, and by the change of coordinates principle from [8, Section 1.3.2] the actions of $\mathrm{Mod}(\Sigma_g^b)$ on $\mathcal{TS}_{h+1}(\Sigma_g^b, I)$ and $\mathcal{TS}_h(\Sigma_g^b, I)$ are transitive on k–simplices for all k. To prove the second condition, therefore, it is enough to show that $\mathcal{TS}_{h+1}(\Sigma_g^b, I) \subset X$ contains an $(n+1)$–simplex. Such a simplex contains $n+2$ disjoint copies of $\tau(\Sigma_h^1, \Sigma_1^1)$. Since

$$\begin{aligned}(n+2)(h+1) &= \left(\frac{g-(2h+3)}{h+2}+2\right)(h+1) = \left(\frac{g-(2h+3)}{h+2}\right)(h+1)+2(h+1)\\ &< (g-(2h+3))+2(h+1) = g-1 < g,\end{aligned}$$

there is enough room on Σ_g^b to find these $n+2$ disjoint copies of $\tau(\Sigma_h^1, \Sigma_1^1)$.

The final condition says that, for all simplices σ of $\mathcal{TS}_h(\Sigma_g^b, I)$ whose dimension is at most n and all faces σ' of σ, the space $\mathrm{Fib}_\psi(\sigma', \sigma)$, defined right before Lemma 2.1, is n–connected. Recall that $\mathrm{Fib}_\psi(\sigma', \sigma)$ is the subcomplex of X consisting of all simplices η' of X such that

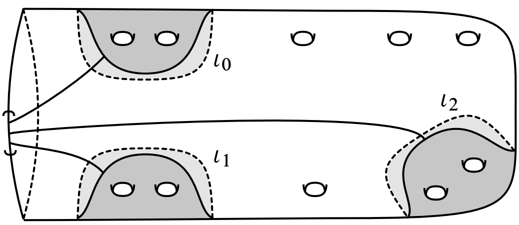
(i) $\psi(\eta')$ is a face of σ', and

(ii) there exists a simplex η of X such that η' is a face of η and $\psi(\eta) = \sigma$.

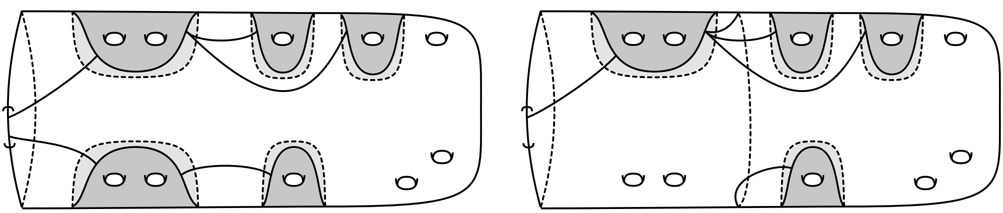
Write

$$\sigma' = \{\iota_0, \dots, \iota_{m'}\} \quad \text{and} \quad \sigma = \{\iota_0, \dots, \iota_{m'}, \dots, \iota_m\},$$

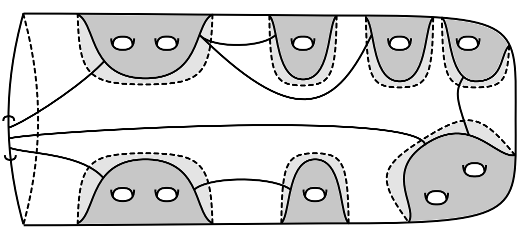
so $0 \le m' \le m \le n$. We will illustrate all our constructions here with the following running example, where $\sigma' = \{\iota_0, \iota_1\}$ and $\sigma = \{\iota_0, \iota_1, \iota_2\}$:



The left side of the following depicts a 2–simplex of X lying in $\mathrm{Fib}_\psi(\sigma', \sigma)$ and the right side depicts a 2–simplex of X that does not lie in $\mathrm{Fib}_\psi(\sigma', \sigma)$:



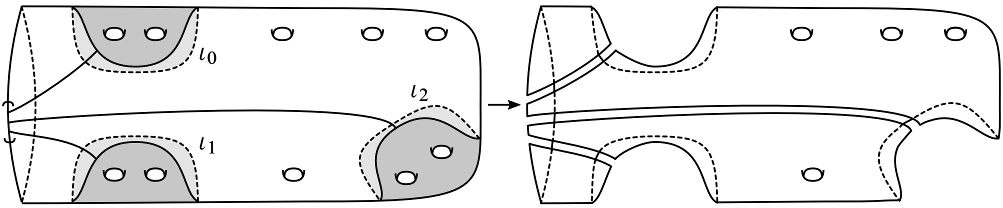
The issue with the simplex η' on the right is that there is not enough genus remaining on the surface to find a simplex η of X satisfying condition (ii) above. For the simplex η' on the left, the desired simplex η is as follows:



To understand the connectivity of $\mathrm{Fib}_\psi(\sigma', \sigma)$, we must relate it to a complex we already understand. Let Σ be the surface obtained by first removing the interior of

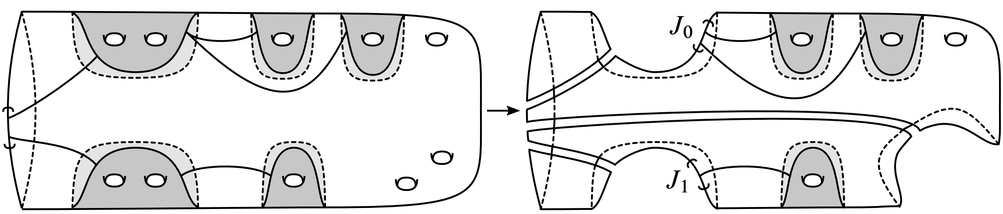
$$\iota_0(\Sigma_h^1) \cup \dots \cup \iota_m(\Sigma_h^1)$$

from Σ_g^b and then cutting open the result along the images of the tethers. In our running example, Σ is obtained as follows:



We thus have $\Sigma \cong \Sigma_{g-(m+1)h}^b$. For $0 \le i \le m'$, let $J_i \subset \partial\Sigma$ be an open interval in $\iota_i(\partial\Sigma_h^1)$ containing the image of the point on $\partial\Sigma_h^1$ to which the attaching tether is attached when forming $\tau(\Sigma_h^1, \Sigma_1^1)$. Set $J = J_1 \cup \cdots \cup J_{m'}$.

The complex $\mathrm{Fib}_\psi(\sigma', \sigma)$ is isomorphic to a subcomplex $\mathcal{TS}_1'(\Sigma, J)$ of $\mathcal{TS}_1(\Sigma, J)$. In our running example, the simplex of $\mathrm{Fib}_\psi(\sigma', \sigma)$ on the left-hand side corresponds to the simplex of $\mathcal{TS}_1(\Sigma, J)$ on the right-hand side:



The different tethers in a simplex of $\mathrm{Fib}_\psi(\sigma', \sigma) \subset X$ that meet at a point of $\iota_i(\partial\Sigma_h^1)$ are "spread out" in J_i so as to be disjoint. The reason that $\mathrm{Fib}_\psi(\sigma', \sigma)$ is only isomorphic to a subcomplex $\mathcal{TS}_1'(\Sigma, J)$ of $\mathcal{TS}_1(\Sigma, J)$ and not the whole thing is that it only corresponds to simplices where there is enough genus remaining to ensure condition (ii) above holds.

Recall that m' and m satisfy $0 \le m' \le m \le n$. As we noted in the first paragraph, the connectivity of $\mathcal{TS}_1(\Sigma, J)$ is at least

$$\tfrac{1}{2}(g - (m+1)h - 3) \ge \tfrac{1}{2}(g - (n+1)h - 3).$$

To prove that connectivity of $\mathcal{TS}_1'(\Sigma, J)$ is at least $n = (g - (2h+3))/(h+2)$, it is enough to prove that

- $\frac{1}{2}(g - (n+1)h - 3) \ge n$, and
- $\mathcal{TS}_1'(\Sigma, J)$ contains the $(n{+}1)$–skeleton of $\mathcal{TS}_1(\Sigma, J)$.

For the first fact, we calculate as follows:

$$\frac{1}{2}(g-(n+1)h-3)=\frac{1}{2}\left(g-\left(\frac{g-(2h+3)}{h+2}+1\right)h-3\right)=\frac{g+\frac{1}{2}h^2-h-3}{h+2}\geq\frac{g-2h-3}{h+2}.$$

Here the final inequality follows from $\frac{1}{2}h^2-h\geq -2h$, which holds for $h\geq 0$.

For the second, consider a simplex $\{j_0,\ldots,j_\ell\}$ of the $(n+1)$–skeleton of $\mathcal{TS}_1(\Sigma,J)$. Let Σ' be the surface obtained by first removing the interior of

$$j_0(\Sigma_1^1)\cup\cdots\cup j_\ell(\Sigma_1^1)$$

from $\Sigma\cong\Sigma^b_{g-(m+1)h}$ and then cutting open the result along the images of the tethers. It follows that

$$\Sigma'\cong\Sigma^b_{g-(m+1)h-(\ell+1)}.$$

In the worst case, where the corresponding simplex η' of X maps to a vertex of σ', we need at least m genus remaining to complete η' to a simplex mapping to σ. In other words, what we must prove is that

$$g-(m+1)h-(\ell+1)\geq m.$$

Since our simplex lies in the $(n+1)$–skeleton, we have $\ell\leq n+1$. Also, $m\leq n$. It follows that it is enough to prove that

$$g-(n+1)h-(n+2)\geq n.$$

Rearranging this, we get

$$\frac{g-h-2}{h+2}\geq n.$$

This follows from the fact that $n=(g-(2h+3))/(h+2)$. □

We now use this to prove the desired connectivity property for $\mathcal{TS}_{h+1}(\Sigma^b_g,I)$.

Claim *$\mathcal{TS}_{h+1}(\Sigma^b_g,I)$ is $(g-(2h+3))/(h+2)$–connected.*

Proof We prove that $\mathcal{TS}_{h+1}(\Sigma^b_g,I)$ is n–connected for $-1\leq n\leq(g-(2h+3))/(h+2)$ by induction on n. The base case $n=-1$ simply asserts that $\mathcal{TS}_{h+1}(\Sigma^b_g,I)$ is nonempty when $(g-(2h+3))/(h+2)\geq -1$. This condition is equivalent to $g\geq h+1$, in which case $\mathcal{TS}_{h+1}(\Sigma^b_g,I)\neq\varnothing$ is obvious.

Assume now that $0\leq n\leq(g-(2h+3))/(h+2)$ and that, for all surfaces $\Sigma^{b'}_{g'}$ and all finite disjoint unions of open intervals $I'\subset\partial\Sigma^{b'}_{g'}$, the space $\mathcal{TS}_{h+1}(\Sigma^{b'}_{g'},I')$ is n'–connected for $n'=\min\{n-1,(g'-(2h+3))/(h+2)\}$. We must prove that $Y:=\mathcal{TS}_{h+1}(\Sigma^b_g,I)$ is n–connected.

We know that X is n–connected, so to prove that its subcomplex Y is n–connected it is enough to prove that the pair (X, Y) is $(n+1)$–connected. We do this using Proposition 2.3. For this, we must identify a set $\mathcal{B}$ of "bad simplices" of X and verify the three hypotheses of the proposition. Define $\mathcal{B}$ to be the set of all simplices σ of X such that, for all vertices v of σ, there exists another vertex v' of σ such that the edge $\{v, v'\}$ of σ does not lie in $Y = \mathcal{TS}_{h+1}(\Sigma_g^b, I)$.

We now verify the hypotheses of Proposition 2.3. The first two are easy:

(i) A simplex of X lies in $Y = \mathcal{TS}_{h+1}(\Sigma_g^b, I)$ if and only if none of its faces lie in $\mathcal{B}$, which is obvious.

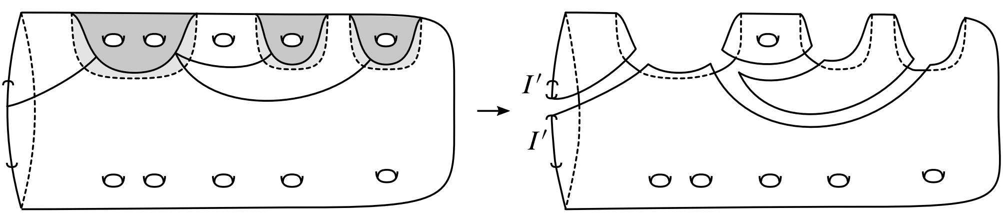
(ii) If $\sigma_1, \sigma_2 \in \mathcal{B}$ are such that $\sigma_1 \cup \sigma_2$ is a simplex of X, then $\sigma_1 \cup \sigma_2 \in \mathcal{B}$, which again is obvious.

The only thing left to check is (iii), which says that, for all k–dimensional $\sigma \in \mathcal{B}$, the complex $G(X, \sigma, \mathcal{B})$ has connectivity at least $(n+1) - k - 1 = n - k$.

Write $\sigma = \{\iota_0, \dots, \iota_k\}$. Let Σ' be the surface obtained by first removing the interiors of

$$\iota_0(\Sigma_h^1 \sqcup \Sigma_1^1) \cup \dots \cup \iota_k(\Sigma_h^1 \sqcup \Sigma_1^1)$$

from Σ_g^b and then cutting open the result along the images of the free and attaching tethers:



The surface Σ' is connected, and when cutting along the free and attaching tethers the open set $I \subset \partial \Sigma_g^b$ is divided into a finer collection I' of open segments (as in the above example). Examining its definition in Section 2.2, we see that

$$G(X, \sigma, \mathcal{B}) \cong \mathcal{TS}_{h+1}(\Sigma', I').$$

We must prove that $\mathcal{TS}_{h+1}(\Sigma', I')$ is $(n-k)$–connected. Let g' be the genus of Σ'. Since $k \geq 1$, we have $n-k < n$, so our inductive hypothesis will say that $\mathcal{TS}_{h+1}(\Sigma', I')$ is $(n-k)$–connected if we can prove that $n - k \leq (g' - (2h+3))/(h+2)$.

This requires estimating g'. The most naive such estimate of g' is

$$g' \geq g - (k+1)(h+1).$$

This is a poor estimate since it does not use the fact that $\sigma \in \mathcal{B}$, which implies that every genus-h surface contributing to this estimate is at least double-counted. Taking this into account, we see that in fact

$$g' \geq g - \tfrac{1}{2}(k+1)h - (k+1) = g - \tfrac{1}{2}(k+1)(h+2).$$

This implies that

$$\frac{g' - (2h+3)}{h+2} \geq \frac{g - (2h+3)}{h+2} - \frac{\frac{1}{2}(k+1)(h+2)}{h+2} \geq n - \tfrac{1}{2}(k+1) \geq n - k,$$

where the final inequality uses the fact that $k \geq 1$. □

This completes the proof of Theorem D. □

3 Stability for surfaces with one boundary component

In this section we prove Theorems A and C. The outline is as follows. In Section 3.1, we discuss the homological stability machine. In Sections 3.2 and 3.3 we prove a number of preliminary results needed to apply this machine. Our proof of Theorem E (and its nonabelian analogue) is in Section 3.3.2. Finally, in Section 3.4 we prove Theorems A and C.

3.1 The stability machine

We now introduce the standard homological stability machine. This is discussed in many places, but the account in [12, Section 1] is particularly convenient for our purposes. We remark that our results could also be proved using the framework of [19] (which generalizes [25]), but since it would not simplify our proofs we decided not to use that framework.

Semisimplicial sets The natural setting for the machine is that of semisimplicial sets, whose definition we now briefly recall. For more details see [9], which calls them Δ–sets. Let Δ be the category with objects the sets $[k] = \{0, \dots, k\}$ for $k \geq 0$ and whose morphisms $[k] \to [\ell]$ are the strictly increasing functions. A *semisimplicial set* is a contravariant functor X from Δ to the category of sets. The *k–simplices* of X are the image X_k of $[k] \in \Delta$. The maps $X_\ell \to X_k$ corresponding to the Δ–morphisms $[k] \to [\ell]$ are called the *face maps*.

Geometric properties A semisimplicial set X has a geometric realization $|X|$ obtained by taking standard k–simplices for each element of X_k and then gluing these simplices together using the face maps. Whenever we talk about topological properties of a semisimplicial set, we are referring to its geometric realization. An action of a group G on a semisimplicial set X consists of actions of G on each X_n that commute with the face maps. This induces an action of G on $|X|$.

The machine The version of the homological stability machine we need is as follows. In it, the indexing is chosen so that the complex X_1 upon which G_1 acts is connected.

Theorem 3.1 *Let*

$$G_0 \subset G_1 \subset G_2 \subset \cdots$$

be an increasing sequence of groups. For each $n \geq 1$, let X_n be a semisimplicial set upon which G_n acts. Assume, for some $c \geq 2$, that the following hold:

(1) *The space X_n is $(n-1)/c$–connected.*

(2) *For all $0 \leq i < n$, the group G_{n-i-1} is the G_n–stabilizer of some i–simplex of X_n.*

(3) *For all $0 \leq i < n$, the group G_n acts transitively on the i–simplices of X_n.*

(4) *For all $n \geq c+1$ and all 1–simplices e of X_n whose boundary consists of vertices v and v', there exists some $\lambda \in G_n$ such that $\lambda(v) = v'$ and such that λ commutes with all elements of $(G_n)_e$.*

Then, for $k \geq 1$, the map $\mathrm{H}_k(G_{n-1}) \to \mathrm{H}_k(G_n)$ is an isomorphism for $n \geq ck+1$ and a surjection for $n = ck$.

Proof This is proved exactly like [12, Theorem 1.1]. □

3.2 Stabilizing and destabilizing markings

We next discuss the process of stabilizing and destabilizing markings. Recall that A is a fixed finitely generated abelian group and Λ is a fixed finite group.

Stabilizing and destabilizing, abelian If μ is an A–homology marking on Σ_g^1 and $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ is an embedding, then we can define the stabilization to $\Sigma_{g'}^1$ of μ just like we did in the introduction. Namely, $\mathrm{H}_1(\Sigma_g)$ can be identified with a symplectic subspace of $\mathrm{H}_1(\Sigma_{g'})$, so

$$\mathrm{H}_1(\Sigma_{g'}) = \mathrm{H}_1(\Sigma_g^1) \oplus \mathrm{H}_1(\Sigma_g^1)^{\perp},$$

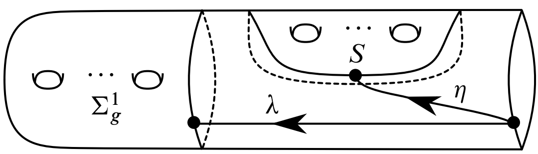
where the $\perp$ is with respect to the algebraic intersection pairing. Define the *stabilization* $\mu'\colon \mathrm{H}_1(\Sigma^1_{g'}) \to A$ of μ to be the composition

$$\mathrm{H}_1(\Sigma^1_{g'}) = \mathrm{H}_1(\Sigma^1_g) \oplus \mathrm{H}_1(\Sigma^1_g)^{\perp} \to \mathrm{H}_1(\Sigma^1_g) \xrightarrow{\mu} A,$$

where the first arrow is the orthogonal projection. We also say μ is a *destabilization* of μ'.

Stabilizing and destabilizing, nonabelian Now let μ be a Λ–marking on Σ^1_g and $\Sigma^1_g \hookrightarrow \Sigma^1_{g'}$ be an embedding. Defining the stabilization of μ to $\Sigma^1_{g'}$ is subtle since there is not a canonical[3] way to stabilize. We thus need to make some auxiliary choices.

Let $* \in \partial\Sigma^1_g$ and $*' \in \partial\Sigma^1_{g'}$ be the basepoints. Let S be a subsurface of $\Sigma^1_{g'} \setminus \mathrm{Int}(\Sigma^1_g)$ with $S \cong \Sigma^1_{g'-g}$. Choose a basepoint $*'' \in \partial S$, and let λ and η be embedded paths in $\Sigma^1_{g'} \setminus \mathrm{Int}(\Sigma^1_g \cup S)$ connecting $*$ to $*'$ and $*''$, respectively. Assume that λ and η are disjoint aside from their initial points:



The paths λ and η induce injective homomorphisms

$$\pi_1(\Sigma^1_g, *) \hookrightarrow \pi_1(\Sigma^1_{g'}, *') \quad \text{and} \quad \pi_1(S, *'') \hookrightarrow \pi_1(\Sigma^1_{g'}, *')$$

taking $x \in \pi_1(\Sigma^1_g, *)$ to $\lambda \cdot x \cdot \lambda^{-1} \in \pi_1(\Sigma^1_{g'}, *')$ and $y \in \pi_1(S, *'')$ to $\eta \cdot y \cdot \eta^{-1} \in \pi_1(\Sigma^1_{g'}, *')$. Identifying $\pi_1(\Sigma^1_g, *)$ and $\pi_1(S, *'')$ with the corresponding subgroups of $\pi_1(\Sigma^1_{g'}, *')$, we have a free product decomposition

$$\pi_1(\Sigma^1_{g'}, *') = \pi_1(\Sigma^1_g, *) \star \pi_1(S, *'').$$

Define $\mu'\colon \pi_1(\Sigma^1_{g'}, *') \to \Lambda$ to be the composition

$$\pi_1(\Sigma^1_{g'}, *') = \pi_1(\Sigma^1_g, *) \star \pi_1(S, *'') \to \pi_1(\Sigma^1_g, *) \xrightarrow{\mu} \Lambda,$$

where the first arrow quotients out by the normal closure of $\pi_1(S, *'')$.

A different choice of η would change the subgroup $\pi_1(S, *'')$ of $\pi_1(\Sigma^1_{g'}, *')$ to a conjugate subgroup, so would not change μ'. It follows that μ' only depends on the pair (S, λ), and we will call μ' the (S, λ)*–stabilization* of μ to $\Sigma^1_{g'}$. If we do not want to specify (S, λ) we will just say that μ' is a stabilization of μ, but be warned that different choices of (S, λ) will lead to different stabilizations. We will also say that μ is a *destabilization* of μ' with *destabilization data* (S, λ).

[3]In the introduction, we made a very specific choice when we stabilized a Λ–marking on Σ^1_g to Σ^1_{g+1}.

Remark 3.2 The choice of S is more important than the choice of λ. Indeed, changing λ would have the effect of conjugating μ' by an element of Λ. This would not affect the associated partial Torelli group $\mathcal{I}(\Sigma_{g'}^1, \mu')$.

Maps between partial Torelli groups Let μ be either an A–homology marking or a Λ–marking on Σ_g^1, let $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ be an embedding, and let μ' be a stabilization of μ to $\Sigma_{g'}^1$. The embedding $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ induces an injective map $\mathrm{Mod}(\Sigma_g^1) \to \mathrm{Mod}(\Sigma_{g'}^1)$ on mapping class groups, and from our definitions it is clear that this restricts to a map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g'}^1, \mu')$ between the associated partial Torelli groups. In fact:

Lemma 3.3 *Suppose that μ is either an A–homology marking or a Λ–marking on Σ_g^1, that $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ is an embedding, and that μ' is a stabilization of μ to $\Sigma_{g'}^1$. Let $\iota\colon \mathrm{Mod}(\Sigma_g^1) \to \mathrm{Mod}(\Sigma_{g'}^1)$ be the map induced by $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$. Then*

$$\mathcal{I}(\Sigma_g^1, \mu) = \{\phi \in \mathrm{Mod}(\Sigma_g^1) \mid \iota(\phi) \in \mathcal{I}(\Sigma_{g'}^1, \mu')\}.$$

Proof This is immediate. □

Vanishing surfaces Recall that the rank $\mathrm{rk}(A)$ of the finitely generated abelian group A is the minimum size of a generating set for A. Consider a subsurface S of Σ_g^1. For an A–homology marking μ on Σ_g^1, we say that μ *vanishes on S* if μ vanishes on the image of $\mathrm{H}_1(S)$ in $\mathrm{H}_1(\Sigma_g^1)$. Similarly, for a Λ–marking μ, we say that μ *vanishes on S* if $\mu(x) = 1$ for all $x \in \pi_1(\Sigma_g^1, *)$ that are freely homotopic to a loop in S. Here $* \in \partial\Sigma_g^1$ is our fixed basepoint.

Proposition 3.4 *Consider some $g, h \geq 1$.*

- *Let μ be an A–homology marking on Σ_g^1, and assume that $g \geq \mathrm{rk}(A) + h$. Then there exists an embedding $S \hookrightarrow \Sigma_g^1$ with $S \cong \Sigma_h^1$ such that μ vanishes on S.*
- *Let μ be a Λ–marking on Σ_g^1, and assume that $g \geq |\Lambda| + h$. Then there exists an embedding $S \hookrightarrow \Sigma_g^1$ with $S \cong \Sigma_h^1$ such that μ vanishes on S.*

Proof of Proposition 3.4 for A–homology markings Consider a symplectic subspace U of $\mathrm{H}_1(\Sigma_g^1)$, ie a subgroup such that $\mathrm{H}_1(\Sigma_g^1) = U \oplus U^{\perp}$, where the $\perp$ is with respect to the algebraic intersection pairing. Such a U is of the form $U \cong \mathbb{Z}^{2k}$ for an integer $k \geq 0$, called the *genus* of U. Every genus-h symplectic subspace U of $\mathrm{H}_1(\Sigma_g^1)$ can be written as $U = \mathrm{H}_1(S)$ for some subsurface S of Σ_g^1 satisfying $S \cong \Sigma_h^1$; see eg [15, Lemma 9]. The proposition is thus equivalent to the purely algebraic Lemma 3.5. □

Lemma 3.5 *Let $V \cong \mathbb{Z}^{2g}$ be a free abelian group equipped with a symplectic form $\omega(-,-)$ and let $\mu\colon V \to A$ be a group homomorphism. Assume that $g \geq \mathrm{rk}(A)+h$ for some $h \geq 1$. There then exists a genus-h symplectic subspace U of V such that $\mu|_U = 0$.*

Proof Without loss of generality, μ is surjective and $A \neq 0$. Also, increasing h if necessary, we can assume that $g = \mathrm{rk}(A)+h$. We will prove the "dual" statement that there exists a genus-$\mathrm{rk}(A)$ symplectic subspace W of V such that $\mu|_{W^\perp} = 0$. The desired U is then $U = W^\perp$. The proof will be by induction on $\mathrm{rk}(A)$. The base case is $\mathrm{rk}(A) = 1$, so A is cyclic. We can factor μ as

$$V \xrightarrow{\tilde{\mu}} \mathbb{Z} \to A.$$

By definition, $\omega(-,-)$ identifies V with its dual, $\mathrm{Hom}(V,\mathbb{Z})$. There thus exists some $a \in V$ such that $\tilde{\mu}(x) = \omega(a,x)$ for all $x \in V$. Pick $b \in V$ with $\omega(a,b) = 1$ and let $W = \langle a,b\rangle$. Then W is a genus-1 symplectic subspace and

$$W^\perp \subset \ker(\omega(a,-)) = \ker(\tilde{\mu}) \subset \ker(\mu),$$

as desired.

Now assume that $\mathrm{rk}(A) > 1$ and that the lemma is true for all smaller ranks. We can then find a short exact sequence

$$0 \to A' \to A \xrightarrow{\phi} A'' \to 0$$

such that $0 < \mathrm{rk}(A') < \mathrm{rk}(A)$ and $\mathrm{rk}(A'')+\mathrm{rk}(A') = \mathrm{rk}(A)$. By our inductive hypothesis, there exists a genus-$\mathrm{rk}(A'')$ symplectic subspace W'' of V such that $(\phi\circ\mu)|_{(W'')^\perp} = 0$. Set $V' = (W'')^\perp$, so V' is a symplectic subspace of V and the image of $\mu' := \mu|_{V'}$ lies in A'. Our inductive hypothesis implies that there is a genus-$\mathrm{rk}(A')$ symplectic subspace W' of V' such that $\mu'|_{(W')^\perp} = 0$. Setting $W = W' \oplus W''$, we have that W is a genus $\mathrm{rk}(A')+\mathrm{rk}(A'') = \mathrm{rk}(A)$ symplectic subspace of V such that $\mu|_{W^\perp} = 0$, as desired. □

Proof of Proposition 3.4 for Λ–markings The proposition is a small variant of a result of Dunfield and Thurston [6, Proposition 6.16] — the only difference is that their result is for closed surfaces, while we need to deal with Σ_g^1. However, the exact same proof works, so we omit the details. □

Deeply destabilizing Proposition 3.4 has the following corollary:

Corollary 3.6 *Consider some $g' \geq 1$.*

- *Let μ' be an A–homology marking on $\Sigma_{g'}^1$. Assume that $g' > \mathrm{rk}(A)$, and let $g = \mathrm{rk}(A)$. Then there exists an embedding $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ and an A–homology marking μ on Σ_g^1 such that μ is a destabilization of μ'.*

- *Let μ' be a Λ–marking on $\Sigma_{g'}^1$. Assume that $g' > |\Lambda|$, and let $g = |\Lambda|$. Then there exists an embedding $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ and a Λ–marking μ on Σ_g^1 such that μ is a destabilization of μ'.*

Proof The proofs for A–homology markings and Λ–markings are similar, so we give the details for Λ–markings (which are slightly more complicated). Let $*' \in \partial\Sigma_{g'}^1$ be the basepoint. By Proposition 3.4, we can find a subsurface $S \hookrightarrow \Sigma_{g'}^1$ with $S \cong \Sigma_{g'-g}^1$ such that μ' vanishes on S. Pick

- an embedding $\Sigma_g^1 \hookrightarrow \Sigma_{g'}^1$ that is disjoint from S, as well as a basepoint $* \in \partial\Sigma_g^1$,
- an embedded path λ in $\Sigma_{g'}^1 \setminus \operatorname{Int}(\Sigma_g^1 \cup S)$ connecting $*'$ to $*$.

Define $\mu\colon \pi_1(\Sigma_g^1, *) \to \Lambda$ via the formula

$$\mu(x) = \mu'(\lambda \cdot x \cdot \lambda^{-1}) \quad \text{for } x \in \pi_1(\Sigma_g^1, *).$$

It is immediate from the definitions that μ' is the (S, λ)–stabilization of μ to $\Sigma_{g'}^1$. $\square$

3.3 Vanishing surfaces

This section constructs the semisimplicial sets we need to apply Theorem 3.1 to the partial Torelli groups.

3.3.1 Vanishing surfaces: definition and basic properties We define the complexes separately for A–homology markings and Λ–markings.

Vanishing subsurfaces, abelian We start by recalling the definition of the complex of vanishing subsurfaces for a homology marking from the introduction. Let μ be an A–homology marking on Σ_g^1. Then define $\mathcal{S}_h(\Sigma_g^1, \mu)$ to be the full subcomplex of $\mathcal{S}_h(\Sigma_g^1)$ spanned by vertices $\iota\colon \Sigma_h^1 \to \Sigma_g^1$ such that μ vanishes on Σ_h^1 in the sense of Section 3.2. The group $\mathcal{I}(\Sigma_g^1, \mu)$ acts on $\mathcal{S}_h(\Sigma_g^1, \mu)$. Similarly, if $I \subset \partial\Sigma_g^1$ is a finite disjoint union of open intervals, then define $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ to be the full subcomplex of $\mathcal{TS}_h(\Sigma_g^1, I)$ spanned by vertices $\iota\colon \tau(\Sigma_h^1) \to \Sigma_g^1$ whose restriction to Σ_h^1 is a vertex of $\mathcal{S}_h(\Sigma_g^1, \mu)$. Again, the group $\mathcal{I}(\Sigma_g^1, \mu)$ acts on $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$.

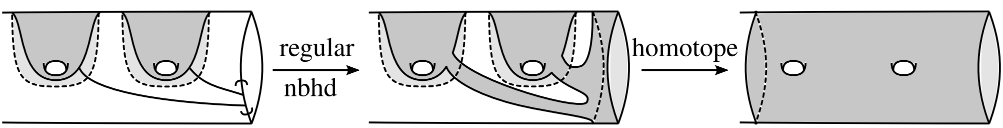
Vanishing subsurfaces, nonabelian Let μ be a Λ–marking on Σ_g^1. Then define $\mathcal{S}_h(\Sigma_g^1, \mu)$ to be the full subcomplex of $\mathcal{S}_h(\Sigma_g^1)$ spanned by vertices $\iota\colon \Sigma_h^1 \to \Sigma_g^1$ such that μ vanishes on Σ_h^1 in the sense of Section 3.2. The group $\mathcal{I}(\Sigma_g^1, \mu)$ acts on $\mathcal{S}_h(\Sigma_g^1, \mu)$. Similarly, if $I \subset \partial\Sigma_g^1$ is a finite disjoint union of open intervals, then define $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ to be the full subcomplex of $\mathcal{TS}_h(\Sigma_g^1, I)$ spanned by vertices $\iota\colon \tau(\Sigma_h^1) \to \Sigma_g^1$ whose restriction to Σ_h^1 is a vertex of $\mathcal{S}_h(\Sigma_g^1, \mu)$. Again, the group $\mathcal{I}(\Sigma_g^1, \mu)$ acts on $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$.

Semisimplicial In the rest of this section, let μ be either an A–homology marking or a Λ–marking on Σ_g^1 and let $I \subset \partial\Sigma_g^1$ be a single interval. We claim then that $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ is naturally a semisimplicial set. The key point here is that its simplices $\{\iota_0, \dots, \iota_k\}$ possess a natural ordering based on the order their tethers leave I.

Stabilizers The $\mathrm{Mod}(\Sigma_g^1)$–stabilizers of simplices of $\mathcal{S}_h(\Sigma_g^1)$ are poorly behaved. The issue is that mapping classes can permute their vertices arbitrarily (which is not possible for $\mathcal{TS}_h(\Sigma_g^1, I)$ since mapping classes must preserve the order in which the tethers leave I). This prevents their stabilizers from being mapping class groups of subsurfaces. For $\mathcal{TS}_h(\Sigma_g^1, I)$, however, this issue does not occur, and the $\mathrm{Mod}(\Sigma_g^1)$–stabilizer of a simplex $\{\iota_0, \dots, \iota_k\}$ of $\mathcal{TS}_h(\Sigma_g^1, I)$ equals $\mathrm{Mod}(\Sigma)$, where Σ is the complement of an open regular neighborhood of

$$\partial\Sigma_g^1 \cup \iota_0(\tau(\Sigma_h^1)) \cup \cdots \cup \iota_k(\tau(\Sigma_h^1)).$$

We will call the complement of this open neighborhood the *stabilizer subsurface* of the simplex. See here, where the stabilizer subsurface is the complement of the shaded region:

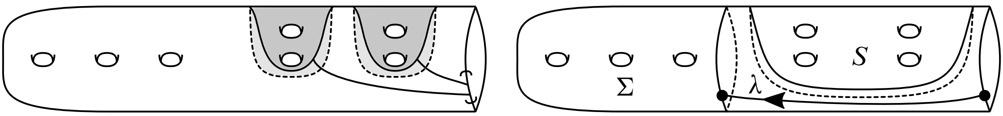


The $\mathcal{I}(\Sigma_g^1, \mu)$ version of this is the following lemma:

Lemma 3.7 *Let μ be either an A–homology marking or a Λ–marking on Σ_g^1, let $I \subset \partial\Sigma_g^1$ be an open interval, and let σ be a k–simplex of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$. Let $\Sigma \cong \Sigma_{g-(k+1)h}^1$ be the stabilizer subsurface of σ. Then there exists a marking μ_0 of the same type as μ (either an A–homology marking or a Λ–marking) on Σ such that μ_0 is a destabilization of μ and such that the $\mathcal{I}(\Sigma_g^1, \mu)$–stabilizer of σ is $\mathcal{I}(\Sigma, \mu_0)$.*

Proof The proofs for A–homology markings and Λ–markings are similar, so we will give the details for Λ–markings. Let $* \in \partial\Sigma_g^1$ and $*_0 \in \partial\Sigma$ be basepoints. Write $\sigma = \{\iota_0, \dots, \iota_k\}$. For $0 \le i \le k$, let $S_i = \iota_i(\Sigma_h^1)$. Let S be a subsurface of $\Sigma_g^1 \setminus \mathrm{Int}(\Sigma)$ such that S contains each S_i and $S \cong \Sigma_{(k+1)h}^1$, and let λ be an embedded path in

$\Sigma_g^1 \setminus \operatorname{Int}(\Sigma \cup S)$ connecting $*$ to $*_0$:



Define $\mu_0 \colon \pi_1(\Sigma, *_0) \to \Lambda$ via the formula

$$\mu_0(x) = \mu(\lambda \cdot x \cdot \lambda^{-1}) \quad \text{for } x \in \pi_1(\Sigma, *_0).$$

It follows from the definitions that μ is the (S,λ)–stabilization of μ_0. Since the $\operatorname{Mod}(\Sigma_g^1)$–stabilizer of σ is $\operatorname{Mod}(\Sigma)$, it follows that the $\mathcal{I}(\Sigma_g^1,\mu)$–stabilizer of σ is $\operatorname{Mod}(\Sigma) \cap \mathcal{I}(\Sigma_g^1,\mu)$, which by Lemma 3.3 is $\mathcal{I}(\Sigma,\mu_0)$. $\square$

3.3.2 Vanishing surfaces: high connectivity The following theorem subsumes Theorem E:

Theorem 3.8 *Fix $g \ge h \ge 1$ and let $I \subset \partial\Sigma_g^1$ be a finite disjoint union of open intervals.*

- *Let μ be an A–homology marking on Σ_g^1. The complexes $\mathcal{S}_h(\Sigma_g^1,\mu)$ and $\mathcal{TS}_h(\Sigma_g^1, I,\mu)$ are both $\big(g-(2\operatorname{rk}(A)+2h+1)\big)/(\operatorname{rk}(A)+h+1)$–connected.*
- *Let μ be a Λ–marking on Σ_g^1. The complexes $\mathcal{S}_h(\Sigma_g^1,\mu)$ and $\mathcal{TS}_h(\Sigma_g^1, I,\mu)$ are both $(g-(2|\Lambda|+2h+1))/(|\Lambda|+h+1)$–connected.*

Proof The proofs for A–homology markings and Λ–markings are identical, so we will give the details for Λ–markings. Also, the proofs that $\mathcal{S}_h(\Sigma_g^b,\mu)$ and $\mathcal{TS}_h(\Sigma_g^b, I,\mu)$ are $(g-(2|\Lambda|+2h+1))/(|\Lambda|+h+1)$–connected are similar. Keeping track of the tethers introduces a few complications, so we will give the details for $\mathcal{TS}_h(\Sigma_g^b, I,\mu)$ and leave $\mathcal{S}_h(\Sigma_g^b,\mu)$ to the reader.

We start by defining an auxiliary space. Let X be the simplicial complex whose vertices are the union of the vertices of the complexes $\mathcal{TS}_h(\Sigma_g^1, I,\mu)$ and $\mathcal{TS}_{|\Lambda|+h}(\Sigma_g^1, I)$ and whose simplices are collections $\{\iota_0,\dots,\iota_k\}$ of vertices that can be isotoped so that their images are disjoint. Both $\mathcal{TS}_h(\Sigma_g^1, I,\mu)$ and $\mathcal{TS}_{|\Lambda|+h}(\Sigma_g^1, I)$ are thus full subcomplexes of X.

We now prove that X enjoys the connectivity property we are trying to prove for $\mathcal{TS}_h(\Sigma_g^b, I,\mu)$:

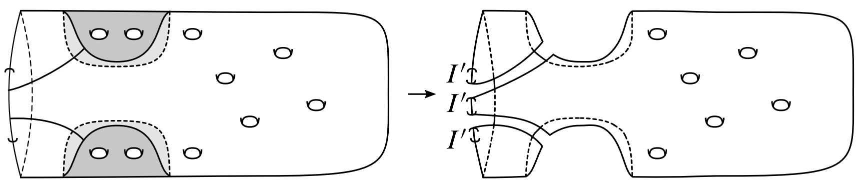
Claim *The space X is $(g-(2|\Lambda|+2h+1))/(|\Lambda|+h+1)$–connected.*

Proof Set $n = (g-(2|\Lambda|+2h+1))/(|\Lambda|+h+1)$, $Y = \mathcal{TS}_{|\Lambda|+h}(\Sigma_g^1, I)$ and $Y' = \mathcal{TS}_h(\Sigma_g^1, I, \mu)$. Theorem D says that Y is n–connected, so it is enough to prove that the pair (X, Y) is n–connected. To do this, we apply Corollary 2.4. This requires, letting σ be a k–dimensional simplex of $Y' = \mathcal{TS}_h(\Sigma_g^1, I, \mu)$ and L be the link of σ in X, showing that $L \cap Y$ is $(n-k-1)$–connected.

Write $\sigma = \{\iota_0, \ldots, \iota_k\}$. Let Σ' be the surface obtained by first removing the interiors of

$$\iota_0(\Sigma_h^1) \cup \cdots \cup \iota_k(\Sigma_h^1)$$

from Σ_g^1 and then cutting open the result along the images of the tethers:



The surface Σ' is connected, and when cutting along the tethers the open set $I \subset \partial\Sigma_g^1$ is divided into a finer collection I' of open segments (as in the above example). Then

$$L \cap Y \cong \mathcal{TS}_{|\Lambda|+h}(\Sigma', I'),$$

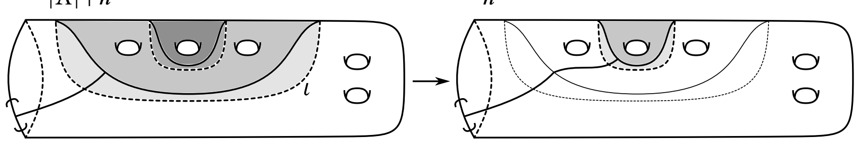
so we must prove that $\mathcal{TS}_{|\Lambda|+h}(\Sigma', I')$ is $(n-k-1)$–connected. Letting g' be the genus of Σ', Theorem D says that $\mathcal{TS}_{|\Lambda|+h}(\Sigma', I')$ is $(g'-(2|\Lambda|+2h+1))/(|\Lambda|+h+1)$–connected, so what we must prove is that

$$n-k-1 \leq \frac{g'-(2|\Lambda|+2h+1)}{|\Lambda|+h+1}.$$

Examining the construction of Σ', we see that $g' = g-(k+1)h$. We now calculate that

$$\frac{g'-(2|\Lambda|+2h+1)}{|\Lambda|+h+1} = \frac{g-(2|\Lambda|+2h+1)}{|\Lambda|+h+1} - \frac{(k+1)h}{|\Lambda|+h+1} \geq n-(k+1). \qquad \square$$

To complete the proof, it is enough to construct a retraction $r\colon X \to \mathcal{TS}_h(\Sigma_g^1, I, \mu)$. For a vertex ι of X, we define $r(\iota)$ as follows. If ι is a vertex of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$, then $r(\iota) = \iota$. If instead ι is a vertex of $\mathcal{TS}_{|\Lambda|+h}(\Sigma_g^1, I)$, then Proposition 3.4 implies that we can find a subsurface $\Sigma_h^1 \hookrightarrow \iota(\Sigma_{|\Lambda|+h})$ such that μ vanishes on Σ_h^1. Define $r(\iota)$ to be the vertex of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ obtained by adjoining the tether of ι and an arbitrary arc in $\iota(\Sigma_{|\Lambda|+h}^1)$ connecting the boundary to Σ_h^1:



Of course, $r(\iota)$ depends on various choices, but we simply make an arbitrary choice. It is clear that this extends over the simplices of X to give a retract $r: X \to \mathcal{TS}_h(\Sigma_g^1, I, \mu)$. $\square$

3.3.3 Vanishing surfaces: transitivity The last fact about the complex of vanishing surfaces we will need is as follows:

Lemma 3.9 *Fix $g \geq h \geq 1$ and let $I \subset \partial \Sigma_g^1$ be an open interval.*

- *Let μ be an A–homology marking on Σ_g^1. The group $\mathcal{I}(\Sigma_g^1, \mu)$ acts transitively on the k–simplices of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ if $g \geq 2h + 2\,\mathrm{rk}(A) + 1 + kh$.*
- *Let μ be a Λ–marking on Σ_g^1. The group $\mathcal{I}(\Sigma_g^1, \mu)$ acts transitively on the k–simplices of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ if $g \geq 2h + 2|\Lambda| + 1 + kh$.*

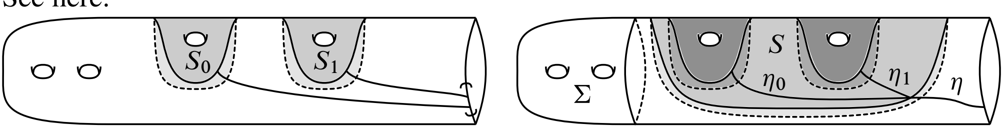
Proof The proofs for A–homology markings and Λ–markings are identical, so we will give the details for Λ–markings. The proof will be by induction on k. We start with the base case $k = 0$.

Claim *If $g \geq 2h + 2|\Lambda| + 1$, then $\mathcal{I}(\Sigma_g^1, \mu)$ acts transitively on the 0–simplices of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$.*

Proof In this case, Theorem 3.8 says that $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ is connected, so it is enough to prove that if ι_0 and ι_1 are vertices of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ that are connected by an edge, then there exists some $f \in \mathcal{I}(\Sigma_g^1, \mu)$ taking ι_0 to ι_1. Let Σ be the stabilizer subsurface of the edge $\{\iota_0, \iota_1\}$, and let $S_0 = \iota_0(\Sigma_h^1)$ and $S_1 = \iota_1(\Sigma_h^1)$. Let S, η, η_0 and η_1 be as follows:

- S is a subsurface of $\Sigma_g^1 \setminus \mathrm{Int}(\Sigma)$ containing S_0 and S_1 and satisfying $S \cong \Sigma_{2h}^1$.
- η is an embedded path in $\Sigma_g^1 \setminus \mathrm{Int}(\Sigma \cup S)$ connecting a point of I to a basepoint of S lying in ∂S.
- For $i = 0, 1$, we have that η_i is an embedded arc in $S \setminus \mathrm{Int}(S_0 \cup S_1)$ connecting the basepoint in ∂S to a basepoint in S_i lying in ∂S_i.
- For $i = 0, 1$, the path $\eta \cdot \eta_i$ is isotopic to the tether of ι_i while keeping its initial point in I and its terminal point fixed.

See here:[4]



[4] In this figure η_0 and η_1 are disjoint, aside from their initial points. This can always be achieved, but is not needed for our proof.

It follows that there exists a Λ–marking μ_0 on Σ and some λ such that μ is the (S,λ)–stabilization of μ_0.

Using the change of coordinates principle from [8, Section 1.3.2], we find $F \in \mathrm{Mod}(S)$ taking $S_0 \cup \eta_0$ to something isotopic to $S_1 \cup \eta_1$. This isotopy will fix the common initial point of η_0 and η_1. Let $f \in \mathrm{Mod}(\Sigma_g^1)$ be the image of F under the map $\mathrm{Mod}(S) \to \mathrm{Mod}(\Sigma_g^1)$. Since f is supported on S, we have $f \in \mathcal{I}(\Sigma_g^1, \mu)$, and by construction we have $f(\iota_0) = \iota_1$. $\square$

Now assume that $k > 0$ and that the theorem is true for simplices of dimension $k-1$. For some $g \geq 2h + 2|\Lambda| + 1 + kh$, let μ be a Λ–marking on Σ_g^1 and $I \subset \partial\Sigma_g^1$ be an open interval. Consider k–simplices σ and σ' of $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$. Enumerate these simplices using the natural ordering discussed above:

$$\sigma = \{\iota_0, \ldots, \iota_k\} \quad \text{and} \quad \sigma' = \{\iota'_0, \ldots, \iota'_k\}. \tag{3-1}$$

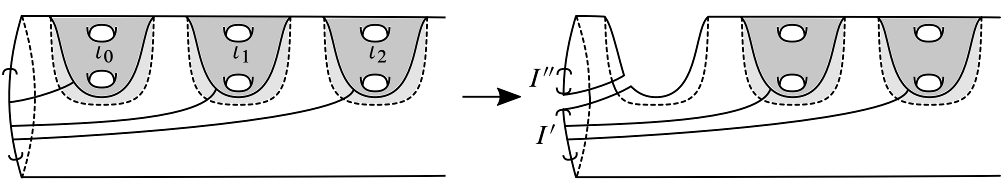
We want to find some $f \in \mathcal{I}(\Sigma_g^1, \mu)$ such that $f(\sigma) = \sigma'$. By the base case $k = 0$, there exists some $f_0 \in \mathcal{I}(\Sigma_g^1, \mu)$ such that $f(\iota_0) = \iota'_0$. Replacing σ by $f(\sigma)$, we can assume that $\iota_0 = \iota'_0$.

Define

$$\sigma_1 = \{\iota_1, \ldots, \iota_k\} \quad \text{and} \quad \sigma'_1 = \{\iota'_1, \ldots, \iota'_k\}.$$

Both σ_1 and σ'_1 are $(k-1)$–simplices in the link of the vertex ι_0, and our goal is to find an element f_1 in the $\mathcal{I}(\Sigma_g^1, \mu)$–stabilizer of ι_0 such that $f_1(\sigma_1) = \sigma'_1$.

Let Σ' be the stabilizer subsurface of ι_0 and let μ' be the Λ–marking on Σ' given by Lemma 3.7, so the $\mathcal{I}(\Sigma_g^1, \mu)$–stabilizer of ι_0 is $\mathcal{I}(\Sigma', \mu')$. The surface Σ' can be constructed by removing the interior of $\iota_0(\Sigma_h^1)$ and then cutting open the result along the tether:



We thus have $\Sigma' \cong \Sigma_{g-h}^1$. Cutting along the tether divides the interval $I \subset \partial\Sigma_g^1$ into two disjoint intervals $I', I'' \subset \partial\Sigma'$, and the link of ι_0 in $\mathcal{TS}_h(\Sigma_g^1, I, \mu)$ can be identified with $\mathcal{TS}_h(\Sigma', I' \sqcup I'', \mu')$. Identifying σ_1 and σ'_1 with simplices in $\mathcal{TS}_h(\Sigma', I' \sqcup I'', \mu')$, the key observation is that, since we enumerated the simplices in (3-1) using the order coming from I, we have (possibly flipping I' and I'') that $\sigma_1, \sigma_1' \subset \mathcal{TS}_h(\Sigma', I', \mu')$. Since $\Sigma' \cong \Sigma_{g-h}^1$ and

$$g-h \geq (2h+2|\Lambda|+1+kh)-h = 2h+2|\Lambda|+1+(k-1)h,$$

we can apply our inductive hypothesis and find some $f_1 \in \mathcal{I}(\Sigma', \mu')$ with $f_1(\sigma_1) = \sigma_1'$, as desired. □

3.4 Proof of stability for surfaces with one boundary component

Proof of Theorems A and C The proofs of the two theorems are identical, so we will give the details for Theorem C. We start by recalling the statement and introducing some notation. Let Λ be a nontrivial finite group, let μ be a Λ–marking on Σ_g^1, and let μ' be the stabilization of μ to Σ_{g+1}^1 in the sense of the introduction.[5] Setting

$$c = |\Lambda|+2 \quad \text{and} \quad d = 2|\Lambda|+2,$$

we want to prove that the map $\mathrm{H}_k(\mathcal{I}(\Sigma_g^1, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma_{g+1}^1, \mu'))$ induced by the stabilization map $\mathcal{I}(\Sigma_g^1, \mu) \to \mathcal{I}(\Sigma_{g+1}^1, \mu')$ is an isomorphism if $g \geq ck+d$ and a surjection if $g = ck+d-1$. We will prove this using Theorem 3.1. This requires fitting $\mathcal{I}(\Sigma_g^1, \mu) \hookrightarrow \mathcal{I}(\Sigma_{g+1}^1, \mu')$ into an increasing sequence of group $\{G_n\}$ and constructing appropriate simplicial complexes.

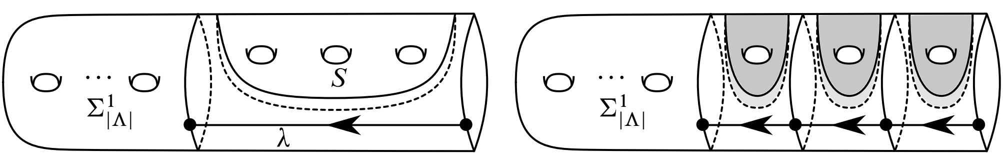
Corollary 3.6 says that there exists an embedding $\Sigma_{|\Lambda|}^1 \hookrightarrow \Sigma_g^1$ and a Λ–marking $\mu_{|\Lambda|}$ on $\Sigma_{|\Lambda|}^1$ such that $\mu_{|\Lambda|}$ is a destabilization of μ. The embedding $\Sigma_{|\Lambda|}^1 \hookrightarrow \Sigma_g^1$ can be factored into a sequence of embeddings

$$\Sigma_{|\Lambda|}^1 \hookrightarrow \Sigma_{|\Lambda|+1}^1 \hookrightarrow \cdots \hookrightarrow \Sigma_g^1,$$

which can then be continued to

$$\Sigma_{|\Lambda|}^1 \hookrightarrow \Sigma_{|\Lambda|+1}^1 \hookrightarrow \cdots \hookrightarrow \Sigma_g^1 \hookrightarrow \Sigma_{g+1}^1 \hookrightarrow \Sigma_{g+2}^1 \hookrightarrow \cdots.$$

As in the following figure, we can break up the destabilization data (S, λ) for the destabilization $\mu_{|\Lambda|}$ of μ into stabilization data (S_h, λ_h) for $|\Lambda|+1 \leq h \leq g$, where (S_h, λ_h) allows us to stabilize from Σ_{h-1}^1 to Σ_h^1:



[5]This uses a specific choice of stabilization data (S, λ).

Starting with $\mu_{|\Lambda|}$, for $h+1 \leq |\Lambda| \leq g$ inductively let μ_h be the (S_h, λ_h)–stabilization of μ_{h-1} on Σ^1_{h-1} to μ_h on Σ^1_h. By construction, $\mu_g = \mu$. Continue stabilizing (now using the choice of stabilization data from the introduction) to define μ_h on Σ^1_h for $h \geq g+1$, so $\mu_{g+1} = \mu'$.

We have thus fit our partial Torelli groups into an increasing sequence of groups

$$\mathcal{I}(\Sigma^1_{|\Lambda|}, \mu_{|\Lambda|}) \subset \mathcal{I}(\Sigma^1_{|\Lambda|+1}, \mu_{|\Lambda|+1}) \subset \mathcal{I}(\Sigma^1_{|\Lambda|+2}, \mu_{|\Lambda|+2}) \subset \cdots.$$

For $h \geq |\Lambda|$, let $I_h \subset \partial \Sigma^1_h$ be an open interval. Theorem 3.8 says that $\mathcal{TS}_1(\Sigma^1_h, I_h, \mu_h)$ is $(h-(d+1))/c$–connected, where c and d are as defined in the first paragraph.

For $n \geq 0$, let

$$G_n = \mathcal{I}(\Sigma_{d+n}, \mu_{d+n}) \quad \text{and} \quad X_n = \mathcal{TS}_1(\Sigma_{d+n}, I_{d+n}, \mu_{d+n}).$$

For this to make sense, we must have $d+n \geq |\Lambda|$, which follows from

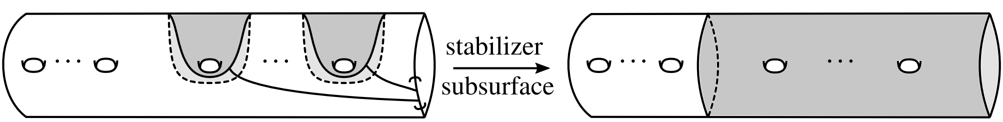
$$d+n = 2|\Lambda| + 2 + n \geq |\Lambda|.$$

We thus have an increasing sequence of groups

$$G_0 \subset G_1 \subset G_2 \subset \cdots$$

with G_n acting on X_n. The indexing convention here is chosen so that X_1 is 0–connected, and more generally so that X_n is $(n-1)/c$–connected, as in Theorem 3.1. Our goal is to prove that the map $\mathrm{H}_k(G_{n-1}) \to \mathrm{H}_k(G_n)$ is an isomorphism for $n \geq ck+1$ and a surjection for $n = ck$, which will follow from Theorem 3.1 once we check its conditions.

- The first is that X_n is $(n-1)/c$–connected, which follows from Theorem 3.8.
- The second is that, for $0 \leq i < n$, the group G_{n-i-1} is the G_n–stabilizer of some i–simplex of X_n, which follows from Lemma 3.7 via the following picture:



- The third is that, for all $0 \leq i < n$, the group G_n acts transitively on the i–simplices of X_n, which follows from Lemma 3.9. For transitivity on the i–simplices this lemma requires that the genus $g = d+n$ used to define $G_n = \mathcal{I}(\Sigma_{d+n}, \mu_{d+n})$ satisfies $g \geq 3 + 2|\Lambda| + i$, which follows from the fact that

$$d+n = (2|\Lambda|+2) + n \geq (2|\Lambda|+2) + (i+1) = 3 + 2|\Lambda| + i.$$

- The fourth is that, for all $n \geq c+1$ and all 1–simplices e of X_n whose boundary consists of vertices v and v', there exists some $\lambda \in G_n$ such that $\lambda(v) = v'$ and such that λ commutes with all elements of $(G_n)_e$. This actually does not require the condition $n \geq c+1$. Let Σ be the stabilizer subsurface of e, so by Lemma 3.7 the stabilizer $(G_n)_e$ consists of mapping classes supported on Σ. The surface $\Sigma^1_{d+n} \setminus \mathrm{Int}(\Sigma)$ is diffeomorphic to Σ^2_2 (as in the picture above), and in particular is connected. The change of coordinates principle from [8, Section 1.3.2] implies that we can find a mapping class λ supported on $\Sigma^1_{d+n} \setminus \mathrm{Int}(\Sigma)$ taking the tethered torus v to v'. This λ clearly lies in G_n and commutes with $(G_n)_e$. □

4 Homology-marked partitioned surfaces

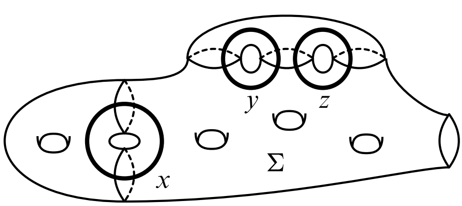
We now turn to partial Torelli groups on surfaces with multiple boundary components. Unfortunately, this introduces genuine difficulties in the proofs, so quite a bit more technical setup is needed. This section contains the categorical framework we will need to even state our result.

Let `Surf` be the category whose objects are compact connected oriented surfaces with boundary and whose morphisms are orientation-preserving embeddings. There is a functor from `Surf` to groups taking $\Sigma \in$ `Surf` to $\mathrm{Mod}(\Sigma)$ and a morphism $\Sigma \hookrightarrow \Sigma'$ to the map $\mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\Sigma')$ that extends mapping classes by the identity. In this section, we augment `Surf` to construct a new category `PSurf` on which we can define partial Torelli groups. This is done in two steps: in Section 4.1 we define the category `PSurf` along with a "partitioned homology functor", and in Section 4.2 we discuss homology markings and construct their associated partial Torelli groups.

4.1 The category `PSurf`

We start with the partitioned surface category, which was introduced in [22].

Motivation This category captures aspects of the homology of a larger surface in which our surface is embedded. For instance, consider the following embedding of a genus-3 surface Σ with six boundary components into Σ^1_7:



For $f \in \text{Mod}(\Sigma)$, the action of f on $\text{H}_1(\Sigma)$ does not determine the action of f on $\text{H}_1(\Sigma_7^1)$. The issue is that we also need to know the action of f on $[x], [y], [z] \in \text{H}_1(\Sigma_7^1)$. The portions of these homology classes that live on Σ are arcs connecting boundary components, so we must consider relative homology groups that incorporate such arcs. However, we do not want to allow all arcs connecting boundary components, since some of these cannot be completed to loops in the larger ambient surface.

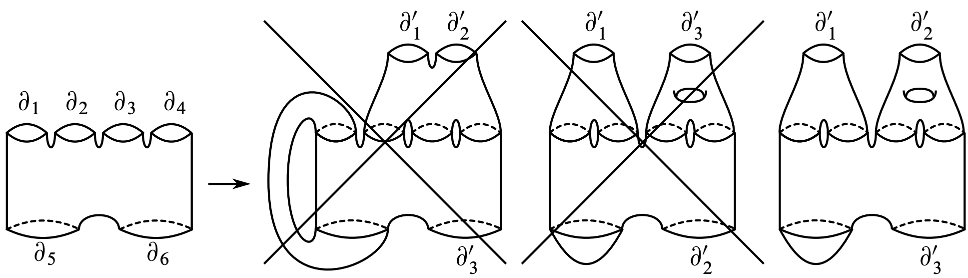
Category To that end, we define a category `PSurf` whose objects are pairs $(\Sigma, \mathcal{P})$:

- Σ is a compact connected oriented surface with boundary.
- $\mathcal{P}$ is a partition of the components of $\partial\Sigma$.

The partition $\mathcal{P}$ tells us which boundary components are allowed to be connected by arcs. The morphisms in `PSurf` from $(\Sigma, \mathcal{P})$ to $(\Sigma', \mathcal{P}')$ are orientation-preserving embeddings $\Sigma \hookrightarrow \Sigma'$ compatible with the partitions $\mathcal{P}$ and $\mathcal{P}'$ in the following sense. For a component S of $\Sigma' \setminus \text{Int}(\Sigma)$, let B_S (resp. B'_S) denote the set of components of $\partial\Sigma$ (resp. $\partial\Sigma'$) that lie in S. In the degenerate case where $S \cong S^1$ (so S is a component of $\partial\Sigma$ and $\partial\Sigma'$), we have $B_S = B'_S$. Our compatibility requirements are then that

- each B_S is a subset of some $p \in \mathcal{P}$, and
- for all $p' \in \mathcal{P}'$ and all $\partial'_1, \partial'_2 \in p'$ such that $\partial'_1 \in B'_{S_1}$ and $\partial'_2 \in B'_{S_2}$ with $S_1 \neq S_2$, there exists some $p \in \mathcal{P}$ such that $B_{S_1} \cup B_{S_2} \subset p$.

Example 4.1 Let $\Sigma = \Sigma_0^6$, $\mathcal{P} = \{\{\partial_1, \partial_2, \partial_3, \partial_4\}, \{\partial_5, \partial_6\}\}$, $\Sigma' = \Sigma_3^3$ and $\mathcal{P}' = \{\{\partial'_1, \partial'_2\}, \{\partial'_3\}\}$. Here are two embeddings $(\Sigma, \mathcal{P}) \hookrightarrow (\Sigma', \mathcal{P}')$ that are not `PSurf`–morphisms and one that is:



We remark that the difference between the second and third embedding is the labeling of the boundary components.

Partitioned homology Consider some $(\Sigma, \mathcal{P}) \in$ `PSurf`. Say that components ∂_1 and ∂_2 of $\partial\Sigma$ are *$\mathcal{P}$–adjacent* if there exists some $p \in \mathcal{P}$ with $\partial_1, \partial_2 \in p$. Define $\text{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ to be the subgroup of the relative homology group $\text{H}_1(\Sigma, \partial\Sigma)$ spanned by the homology classes of oriented closed curves and arcs connecting $\mathcal{P}$–adjacent boundary components. The group $\mathrm{Mod}(\Sigma)$ acts on $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$.

Remark 4.2 This is slightly different from the partitioned homology group defined in [22], which was *not* functorial. The Torelli groups defined via the above homology groups are thus different from those in [22].

Functoriality The assignment

$$(\Sigma, \mathcal{P}) \mapsto \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$$

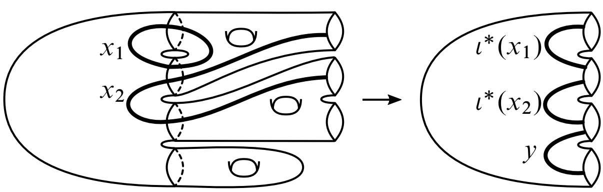
is a contravariant functor from `PSurf` to abelian groups. To see this, consider a `PSurf`–morphism $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$. Identify Σ with its image under ι. We then have maps

$$\mathrm{H}_1(\Sigma', \partial\Sigma') \to \mathrm{H}_1(\Sigma', \Sigma' \setminus \mathrm{Int}(\Sigma)) \xrightarrow{\cong} \mathrm{H}_1(\Sigma, \partial\Sigma),$$

where the second map is the excision isomorphism. From the definition of a `PSurf`–morphism, it follows immediately that this composition restricts to a map

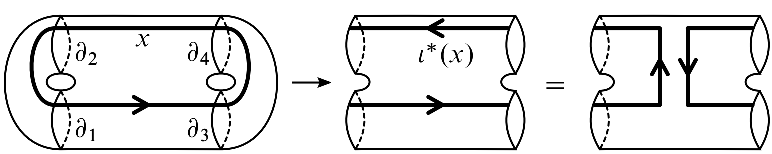
$$\iota^*\colon \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma).$$

Example 4.3 Let $\Sigma = \Sigma_0^4$ and $\Sigma' = \Sigma_4^3$. Let $\mathcal{P}$ (resp. $\mathcal{P}'$) be the partition of the components of $\partial\Sigma$ (resp. $\partial\Sigma'$) consisting of a single partition element containing all the boundary components. The following picture shows a `PSurf`–morphism $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ along with $x_1, x_2 \in \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma')$ and $\iota^*(x_1), \iota^*(x_2) \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$:



To simplify the picture we do not indicate the orientations of the curves/arcs. The above picture also shows an element $y \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ that is not in the image of ι^*.

Example 4.4 Let $\Sigma = \Sigma_0^4$ and $\Sigma' = \Sigma_2$. Let $\{\partial_1, \dots, \partial_4\}$ be the boundary components of Σ, and let $\mathcal{P} = \{\{\partial_1, \partial_2\}, \{\partial_3, \partial_4\}\}$ and $\mathcal{P}' = \varnothing$. The following picture shows a `PSurf`–morphism $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ along with $x \in \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma')$ and $\iota^*(x) \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$:



As it is initially drawn, $\iota^*(x)$ does not appear to be in $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ since it is a pair of arcs connecting boundary components that are not $\mathcal{P}$–adjacent; however, as the figure shows, this pair of arcs is homologous to a pair of arcs connecting boundary components that are $\mathcal{P}$–adjacent.

Action on partitioned homology The mapping class group is a covariant functor from `Surf` to groups, while the partitioned homology group is a contravariant functor from `PSurf` to abelian groups. They are related by the following "push-pull" formula:

Lemma 4.5 *Let $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a `PSurf`–morphism, $\iota_*\colon \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\Sigma')$ be the induced map on mapping class groups and $\iota^*\colon \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ be the induced map on partitioned homology groups. Then*

$$\iota^*(\iota_*(f)(x')) = f(\iota^*(x')) \quad \text{for } f \in \mathrm{Mod}(\Sigma) \text{ and } x' \in \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma').$$

Proof This is obvious. □

4.2 Homology markings on `PSurf`

Recall that A is a fixed finitely generated abelian group.

Markings and partial Torelli groups Consider $(\Sigma, \mathcal{P}) \in$ `PSurf`. An *A–homology marking* on $(\Sigma, \mathcal{P})$ is a homomorphism $\mu\colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to A$. The associated *partial Torelli group* is

$$\mathcal{I}(\Sigma, \mathcal{P}, \mu) = \{f \in \mathrm{Mod}(\Sigma) \mid \mu(f(x)) = \mu(x) \text{ for all } x \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)\}.$$

Stabilizations If $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ is a `PSurf`–morphism and μ is an A–homology marking on $(\Sigma, \mathcal{P})$, then the *stabilization* of μ to $(\Sigma', \mathcal{P}')$ is the composition

$$\mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \xrightarrow{\iota^*} \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \xrightarrow{\mu} A.$$

Lemma 4.6 *Let $\iota\colon (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a `PSurf`–morphism, μ be an A–homology marking on $(\Sigma, \mathcal{P})$, μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$, and $\iota_*\colon \mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\Sigma')$ be the induced map. Then $\iota_*(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \subset \mathcal{I}(\Sigma', \mathcal{P}', \mu')$.*

Proof Let $\iota^*\colon \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ be the induced map. For $f \in \mathcal{I}(\Sigma, \mathcal{P}, \mu)$ and $x' \in \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma')$, we have

$$\mu'(\iota_*(f)(x')) = \mu\big(\iota^*(\iota_*(f)(x'))\big) = \mu\big(f(\iota^*(x'))\big) = \mu(\iota^*(x')) = \mu'(x').$$

Here the second equality follows from Lemma 4.5 and the third from the fact that $f \in \mathcal{I}(\Sigma, \mathcal{P}, \mu)$. The lemma follows. □

5 Stability for surfaces with multiple boundary components

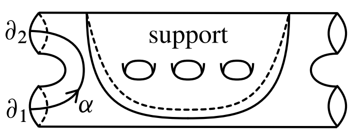
In this section, we state our stability theorem for the partial Torelli groups on surfaces with multiple boundary components and reduce this theorem to a result that will be proved in the next section using the homological stability machine. The statement of our result is in Section 5.1 and the reductions are in Sections 5.2, 5.3 and 5.4.

5.1 Statement of result

To get around the issues with closed surfaces underlying Theorem B from Section 1, we will need to impose some conditions on our stabilization maps.

Support If μ is an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$, we say that μ is *supported on a genus-h symplectic subsurface* if there exists a $\mathtt{PSurf}$–morphism $(\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ with $\Sigma' \cong \Sigma_h^1$ and an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ such that μ is the stabilization of μ' to $(\Sigma, \mathcal{P})$. If there exists some $h \geq 1$ such that μ is supported on a genus-h symplectic subsurface, then we will simply say that μ is *supported on a symplectic subsurface*.

Remark 5.1 Not all A–homology markings are supported on a symplectic subsurface. Indeed, letting ∂_1 and ∂_2 be $\mathcal{P}$–adjacent boundary components of Σ, this condition implies that we can find an arc α connecting ∂_1 to ∂_2 such that $\mu([\alpha]) = 0$; see here:



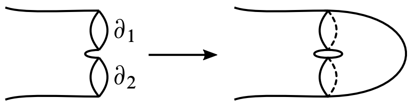
It is easy to construct A–homology markings not satisfying this property; for instance, let $A = \widetilde{\mathrm{H}}_0(\partial\Sigma)$ and let $\mu\colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to A$ be the restriction to $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ of the boundary map $\mathrm{H}_1(\Sigma, \partial\Sigma) \to \widetilde{\mathrm{H}}_0(\partial\Sigma)$. We will later show that this is the only obstruction; see Lemma 6.2 below.

Partition bijectivity Consider a $\mathtt{PSurf}$–morphism $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$. Identify Σ with its image in Σ'. We will call this morphism *partition-bijective* if the following holds for all $p \in \mathcal{P}$:

- Let S be the union of the components of $\Sigma' \setminus \mathrm{Int}(\Sigma)$ that contain a boundary component in p. Then there exists a unique $p' \in \mathcal{P}'$ such that p' consists of the components of $S \cap \partial\Sigma'$.

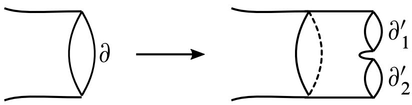
This condition implies in particular that S contains components of $\partial\Sigma'$. It rules out two kinds of morphisms:

- The first is morphisms where, for some $p \in \mathcal{P}$, the union of the components of $\Sigma' \setminus \mathrm{Int}(\Sigma)$ that contain a boundary component in p contains no components of $\partial\Sigma'$. See here:



Here $p = \{\partial_1, \partial_2\}$.

- The second is morphisms where a single $p \in \mathcal{P}$ "splits" into multiple elements of $\mathcal{P}'$ like this:



Here $p = \{\partial\}$ and $\mathcal{P}'$ contains both $\{\partial_1'\}$ and $\{\partial_2'\}$.

Main theorem With this definition, we have the following theorem:

Theorem F *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf` that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a partition-bijective `PSurf`–morphism and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Then the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism if the genus of Σ is at least $(\mathrm{rk}(A) + 2)k + (2\,\mathrm{rk}(A) + 2)$.*

Remark 5.2 Theorem F does not assert that the map is a surjection if the genus of Σ is at least $(\mathrm{rk}(A) + 2)k + (2\,\mathrm{rk}(A) + 1)$. We do not know if this is true—while an appropriate surjectivity statement will follow from our invocation of the homological stability machine, this will only cover certain special kinds of morphisms $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$, the "double boundary stabilizations", and the general case will be reduced to these special morphisms in a fairly involved way.

Counterexamples We do not know whether or not the condition in Theorem F that μ be supported on a symplectic subsurface is necessary. However, the condition that the morphism be partition-bijective is necessary. Indeed, in Section 7 we will prove the following theorem. The condition of being *symplectically nondegenerate* in it will be defined in that section; it is satisfied by most interesting homology markings.

Theorem 5.3 *Let μ be a symplectically nondegenerate A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf` that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$*

be a non-partition-bijective PSurf*–morphism and let* μ' *be the stabilization of* μ *to* $(\Sigma', \mathcal{P}')$*. Assume that the genus of* Σ *is at least* $3\,\mathrm{rk}(A)+4$*. Then the induced map* $\mathrm{H}_1(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma',\mathcal{P}',\mu'))$ *is not an isomorphism.*

Remark 5.4 The map is frequently not an isomorphism, even when the genus of Σ is smaller. We use the genus assumption in Theorem 5.3 so we can apply Theorem F to change Σ and Σ' so as to put ourselves in a situation where the phenomenon underlying Theorem B occurs.

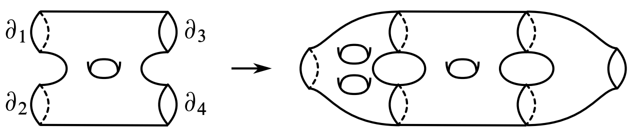
5.2 Reduction I: open cappings

In this section, we reduce Theorem F to certain kinds of PSurf–morphisms called open cappings, whose definition is below.

Open cappings An *open capping* is a PSurf–morphism $(\Sigma,\mathcal{P}) \to (\Sigma',\mathcal{P}')$ such that the following holds for all $p \in \mathcal{P}$:

- Let S be the union of the components of $\Sigma' \setminus \mathrm{Int}(\Sigma)$ that contain a boundary component in p. Then S is connected and $S \cap \partial\Sigma'$ consists of a single component.

Unraveling the definition of a PSurf–morphism, this implies that $\mathcal{P}'$ is the *discrete partition*, that is, the partition $\mathcal{P}' = \{\{\partial'\} \mid \partial' \text{ is a component of } \partial\Sigma'\}$. See the following example, where $\mathcal{P} = \{\{\partial_1,\partial_2\},\{\partial_3,\partial_4\}\}$:



By definition, an open capping is partition-bijective.

Remark 5.5 In [22], a capping is defined similarly to an open capping, but where Σ' is closed and ∂S is simply an element of $\mathcal{P}$.

Reduction The following is a special case of Theorem F:

Proposition 5.6 *Let* μ *be an* A*–homology marking on* $(\Sigma,\mathcal{P}) \in$ PSurf *that is supported on a symplectic subsurface. Let* $(\Sigma,\mathcal{P}) \to (\Sigma',\mathcal{P}')$ *be an open capping and let* μ' *be the stabilization of* μ *to* $(\Sigma',\mathcal{P}')$*. Then the induced map*

$$\mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu'))$$

is an isomorphism if the genus of Σ *is at least* $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$*.*

The proof of Proposition 5.6 begins in Section 5.3. First, we use it to deduce Theorem F:

Proof of Theorem F, assuming Proposition 5.6 We start by recalling the statement of the theorem. Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a partition-bijective $\mathtt{PSurf}$–morphism and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Assume that the genus of Σ is at least $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$. Our goal is to prove that the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism.

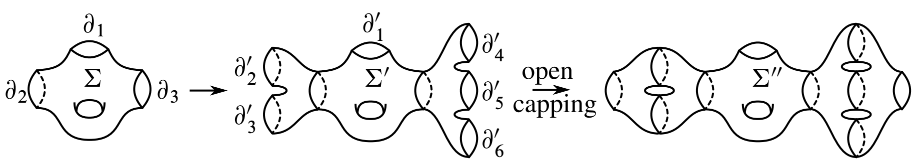
Identify Σ with its image in Σ'. The proof has two cases. Recall that the discrete partition of the boundary components of a surface S is $\{\{\partial\} \mid \partial \text{ is a component of } \partial S\}$.

Case 1 *$\mathcal{P}$ is the discrete partition of Σ.*

Let $S_1, \dots, S_b$ be the components of $\Sigma' \setminus \mathrm{Int}(\Sigma)$. For each $1 \le i \le b$, let B_{S_i} be the components of $\partial\Sigma$ that are contained in S_i and let B'_{S_i} be the components of $\partial\Sigma'$ that are contained in S_i. Then:

- Since $\mathcal{P}$ is the discrete partition, each B_{S_i} is a one-element set containing a single boundary component of Σ, and $\mathcal{P} = \{B_{S_1}, \dots, B_{S_b}\}$.
- Since the morphism $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ is partition-bijective, each B'_{S_i} is a nonempty set of boundary components of Σ', and $\mathcal{P}' = \{B'_{S_1}, \dots, B'_{S_b}\}$.

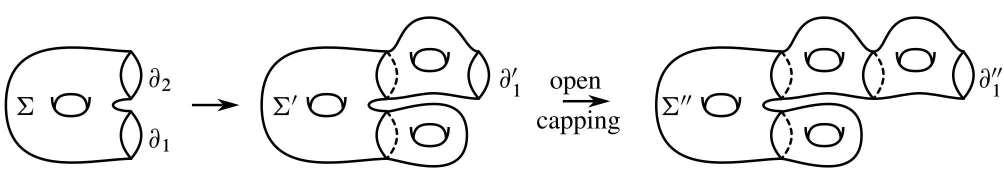
See the following figure, where $\Sigma \cong \Sigma_1^3$ with the discrete partition $\mathcal{P} = \{\{\partial_1\}, \{\partial_2\}, \{\partial_3\}\}$ and $\Sigma' \cong \Sigma_1^6$ with the partition $\mathcal{P}' = \{\{\partial'_1\}, \{\partial'_2, \partial'_3\}, \{\partial'_4, \partial'_5, \partial'_6\}\}$:



As in that figure, let $(\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$ be an open capping and let μ'' be the stabilization of μ' to $(\Sigma'', \mathcal{P}'')$. It follows from the above that the composition

$$(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$$

is also an open capping. We remark that this can fail if $\mathcal{P}$ is not the discrete partition. For instance, consider the morphisms $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ and $(\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$ in the following figure, where $\mathcal{P} = \{\{\partial_1, \partial_2\}\}$, $\mathcal{P}' = \{\{\partial'_1\}\}$ and $\mathcal{P}'' = \{\{\partial''_1\}\}$:



We have maps

$$\mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu')) \to \mathrm{H}_k(\mathcal{I}(\Sigma'',\mathcal{P}'',\mu'')).$$

Proposition 5.6 implies that

$$\mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma'',\mathcal{P}'',\mu'')), \quad \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu')) \to \mathrm{H}_k(\mathcal{I}(\Sigma'',\mathcal{P}'',\mu''))$$

are isomorphisms. We conclude that the map

$$\mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu'))$$

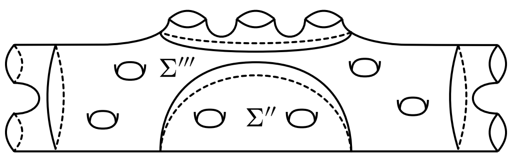
is an isomorphism, as desired.

Case 2 *$\mathcal{P}$ is not the discrete partition of $\partial\Sigma$.*

Since μ is supported on a symplectic subsurface, we can find a PSurf–morphism $(\Sigma'',\mathcal{P}'') \to (\Sigma,\mathcal{P})$ with $\Sigma'' \cong \Sigma_h^1$ and an A–homology marking μ'' on $(\Sigma'',\mathcal{P}'')$ such that μ is the stabilization of μ'' to $(\Sigma,\mathcal{P})$. We can factor $(\Sigma'',\mathcal{P}'') \to (\Sigma,\mathcal{P})$ as

$$(\Sigma'',\mathcal{P}'') \to (\Sigma''',\mathcal{P}''') \to (\Sigma,\mathcal{P})$$

so that Σ''' has the same genus as Σ, $\mathcal{P}'''$ is the discrete partition of $\partial\Sigma'''$, and $(\Sigma''',\mathcal{P}''') \to (\Sigma,\mathcal{P})$ is partition-bijective; see here:



In this example, $\mathcal{P}$ consists of three sets of boundary components (the ones on the left, right, and top). Let μ''' be the stabilization of μ'' to $(\Sigma''',\mathcal{P}''')$. We have maps

$$\mathrm{H}_k(\mathcal{I}(\Sigma''',\mathcal{P}''',\mu''')) \to \mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu')).$$

Case 1 implies that the maps

$$\mathrm{H}_k(\mathcal{I}(\Sigma''',\mathcal{P}''',\mu''')) \to \mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)), \quad \mathrm{H}_k(\mathcal{I}(\Sigma''',\mathcal{P}''',\mu''')) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu'))$$

are isomorphisms. We conclude that the map

$$\mathrm{H}_k(\mathcal{I}(\Sigma,\mathcal{P},\mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma',\mathcal{P}',\mu'))$$

is an isomorphism, as desired. □

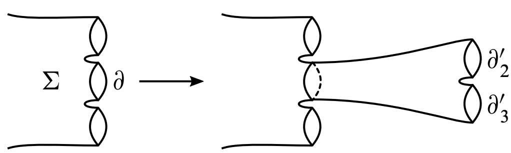
5.3 Reduction II: boundary stabilizations

In this section we reduce Proposition 5.6 to showing that certain kinds of PSurf–morphisms, called increasing boundary stabilizations and decreasing boundary stabilizations, induce isomorphisms on homology.

Increasing boundary stabilization Let $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. An *increasing boundary stabilization* of $(\Sigma, \mathcal{P})$ is a PSurf–morphism $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ constructed as follows. Let ∂ be a component of $\partial\Sigma$ and let $p \in \mathcal{P}$ be the partition element with $\partial \in p$. Also, let $\partial\Sigma_0^3 = \{\partial_1', \partial_2', \partial_3'\}$.

- Σ' is obtained by attaching Σ_0^3 to Σ by gluing $\partial_1' \subset \Sigma_0^3$ to $\partial \subset \Sigma$.
- $\mathcal{P}'$ is obtained from $\mathcal{P}$ by replacing p with $p' = (p \setminus \{\partial\}) \cup \{\partial_2', \partial_3'\}$.

See here:



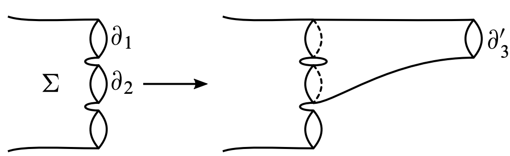
In Section 5.4, we will prove the following.

Proposition 5.7 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be an increasing boundary stabilization and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Then the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism if the genus of Σ is at least $(\mathrm{rk}(A) + 2)k + (2\,\mathrm{rk}(A) + 2)$.*

Decreasing boundary stabilization Let $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. A *decreasing boundary stabilization* of $(\Sigma, \mathcal{P})$ is a PSurf–morphism $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ constructed as follows. Let ∂_1 and ∂_2 be distinct components of $\partial\Sigma$ that both lie in some $p \in \mathcal{P}$, and let $\partial\Sigma_0^3 = \{\partial_1', \partial_2', \partial_3'\}$.

- Σ' is obtained by attaching Σ_0^3 to Σ by gluing ∂_1' and ∂_2' to ∂_1 and ∂_2, respectively.
- $\mathcal{P}'$ is obtained from $\mathcal{P}$ by replacing p with $p' = (p \setminus \{\partial_1, \partial_2\}) \cup \{\partial_3'\}$.

See here:

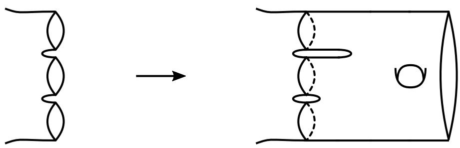


In Section 5.4, we will prove the following:

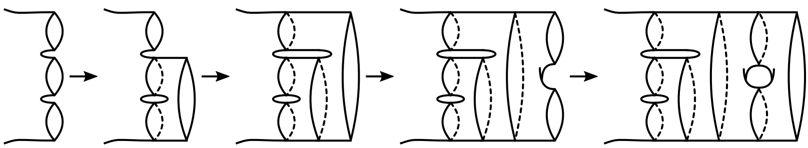
Proposition 5.8 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a decreasing boundary stabilization and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Then the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism if the genus of Σ is at least $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$.*

Deriving Proposition 5.6 As we said above, we will prove Propositions 5.7 and 5.8 in Section 5.4. Here we will explain how to use them to prove Proposition 5.6.

Proof of Proposition 5.6, assuming Propositions 5.7 and 5.8 It is geometrically clear that an open capping $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ can be factored as a composition of increasing boundary stabilizations and decreasing boundary stabilizations. For instance,



can be factored as



The proposition follows. □

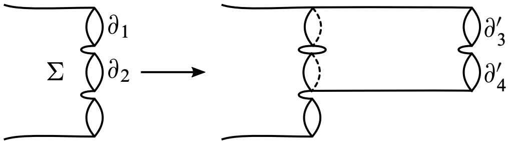
5.4 Reduction III: double boundary stabilizations

In this section, we adapt a beautiful idea of Hatcher and Vogtmann [12] to show how to reduce our two different boundary stabilizations (increasing and decreasing) to a single kind of stabilization called a double boundary stabilization.

Double boundary stabilization Let $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. A *double boundary stabilization* of $(\Sigma, \mathcal{P})$ is a $\mathtt{PSurf}$–morphism $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ constructed as follows. Let ∂_1 and ∂_2 be components of $\partial\Sigma$ that lie in a single element $p \in \mathcal{P}$. Also, let $\partial\Sigma_0^4 = \{\partial_1', \partial_2', \partial_3', \partial_4'\}$.

- Σ' is obtained by attaching $\Sigma_{0,4}$ to Σ by gluing ∂_1' and ∂_2' to ∂_1 and ∂_2, respectively.
- $\mathcal{P}'$ is obtained from $\mathcal{P}$ by replacing p with $p' = (p \setminus \{\partial_1, \partial_2\}) \cup \{\partial_3', \partial_4'\}$.

See here:



In Section 6, we will use the homological stability machine to prove the following:

Proposition 5.9 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf` that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a double boundary stabilization and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Then the induced map*

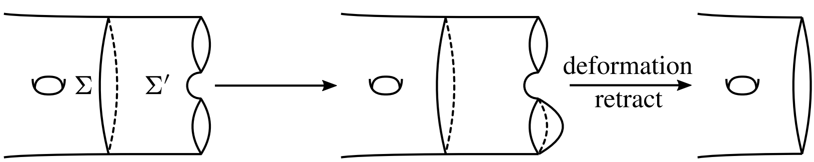
$$\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$$

is an isomorphism if the genus of Σ is at least $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$ and a surjection if the genus of Σ is $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+1)$.

Deriving Propositions 5.7 and 5.8 As we said above, we will prove Proposition 5.9 in Section 6. Here we will explain how to use it to prove Propositions 5.7 and 5.8.

Proof of Proposition 5.7, assuming Proposition 5.9 We start by recalling the statement. Consider an increasing boundary stabilization $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$. Let μ be an A–homology marking on $(\Sigma, \mathcal{P})$ that is supported on a symplectic subsurface and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Assume that the genus of Σ is at least $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$. We must prove that the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism.

The first observation is that the map $\mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathcal{I}(\Sigma', \mathcal{P}', \mu')$ is split injective via a splitting map $\mathcal{I}(\Sigma', \mathcal{P}', \mu') \to \mathcal{I}(\Sigma, \mathcal{P}, \mu)$ induced by gluing a disc to one of the two components of $\partial\Sigma' \setminus \partial\Sigma$:

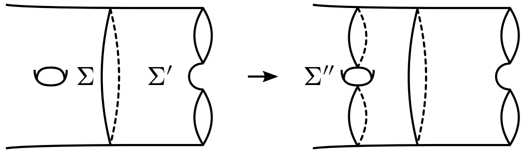


The map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is thus injective, so it is enough to prove that it is surjective.

Combining the fact that μ is supported on a symplectic subsurface with Corollary 3.6, we see that μ is in fact supported on a symplectic subsurface of genus at most $\mathrm{rk}(A)$. Since the genus of Σ is greater than $\mathrm{rk}(A)$, this implies that we can find a decreasing boundary stabilization $(\Sigma'', \mathcal{P}'') \to (\Sigma, \mathcal{P})$ and an A–homology marking μ'' on $(\Sigma'', \mathcal{P}'')$ that is supported on a symplectic subsurface such that μ is the stabilization of μ'' to $(\Sigma, \mathcal{P})$ and such that the composition

$$(\Sigma'', \mathcal{P}'') \to (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$$

is a double boundary stabilization; see here:



The genus of Σ'' is one less than that of Σ, and so is at least $(\mathrm{rk}(A)+2)k+(2\,\mathrm{rk}(A)+1)$. We can thus apply Proposition 5.9 to deduce that the composition

$$\mathrm{H}_k(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu'')) \to \mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$$

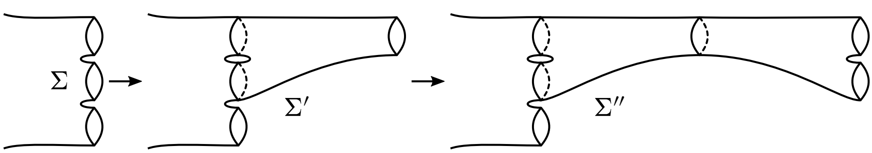
is surjective, and thus that the map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is surjective, as desired. □

Proof of Proposition 5.8, assuming Proposition 5.9 We start by recalling the statement. Consider a decreasing boundary stabilization $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$. Let μ be an A–homology marking on $(\Sigma, \mathcal{P})$ that is supported on a symplectic subsurface and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Assume that the genus of Σ is at least $(\mathrm{rk}(A)+2)k + (2\,\mathrm{rk}(A)+2)$. We must prove that the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism.

Let ∂' be the component of $\partial\Sigma'$ that is not a component of $\partial\Sigma$. As in the following picture, we can construct an increasing boundary stabilization $(\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$ that attaches a 3–holed torus to ∂' such that the composition

$$(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$$

is a double boundary stabilization:



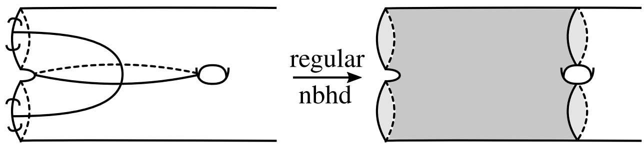
Let μ'' be the stabilization of μ' to $(\Sigma'', \mathcal{P}'')$. We then have maps

$$\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu')) \to \mathrm{H}_k(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu'')). \tag{5-1}$$

Proposition 5.9 implies that the composition (5-1) is an isomorphism, and Proposition 5.7 implies that $\mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu')) \to \mathrm{H}_k(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu''))$ is an isomorphism. We conclude that $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism, as desired. □

6 Double boundary stabilization

Adapting an argument due to Hatcher and Vogtmann [12], we will prove Proposition 5.9 by studying a complex of "order-preserving double-tethered loops" whose vertex-stabilizers yield double boundary stabilizations:



We will require that the homology classes of both the loop and the "double-tether" arc vanish under the homology marking. Getting the loop to vanish will be an easy variant on the argument we used for vanishing surfaces in Section 3.3.2, but getting the double-tether to vanish is harder and will require new ideas. We will build up the complex in three stages (tethered vanishing loops, then double-tethered vanishing loops, and then finally order-preserving double-tethered vanishing loops) in Section 6.3–6.7. These five sections are preceded by two technical sections: Section 6.1 gives a necessary and sufficient condition for an A–homology marking to be supported on a symplectic subsurface, and Section 6.2 is about destabilizing A–homology markings. After all this is complete, we prove Proposition 5.9 in Section 6.8.

6.1 Identifying markings supported on a symplectic subsurface

Consider some $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. In this section, we give a necessary and sufficient condition for an A–homology marking on $(\Sigma, \mathcal{P})$ to be supported on a symplectic subsurface. This requires some preliminary definitions (which will also be used later).

Intersection map Let q be a finite set of oriented simple closed curves on Σ and let $\mathbb{Z}[q]$ be the set of formal $\mathbb{Z}$–linear combinations of elements of q. Define the *q–intersection map* to be the map $\mathrm{i}_q \colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to \mathbb{Z}[q]$ defined as follows. Let

$$\omega_\Sigma \colon \mathrm{H}_1(\Sigma, \partial\Sigma) \times \mathrm{H}_1(\Sigma) \to \mathbb{Z}$$

be the algebraic intersection pairing. For $x \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$, we then set

$$\mathrm{i}_q(x) = \sum_{\gamma \in q} \omega_\Sigma(x, [\gamma]) \cdot \gamma.$$

Total boundary map For a set q as above, define

$$\widetilde{\mathbb{Z}}[q] = \left\{ \sum_{\gamma \in q} c_\gamma \cdot \gamma \in \mathbb{Z}[q] \,\middle|\, \sum_{\gamma \in q} c_\gamma = 0 \right\}.$$

Consider $p \in \mathcal{P}$. Each boundary component $\partial \in p$ is a simple closed curve on Σ, and the orientation on Σ induces an orientation on ∂ such that $\mathrm{Int}(\Sigma)$ lies to the left of ∂. We thus have the map $\mathfrak{i}_p \colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to \mathbb{Z}[p]$. Since $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ is generated by the homology classes of oriented loops and arcs connecting $\mathcal{P}$–adjacent boundary components, the image of $\mathfrak{i}_p$ is $\widetilde{\mathbb{Z}}[p]$. Define

$$\widetilde{\mathbb{Z}}_{\mathcal{P}} = \bigoplus_{p \in \mathcal{P}} \widetilde{\mathbb{Z}}[p].$$

The *total boundary map* of $(\Sigma, \mathcal{P})$ is the map $\mathfrak{i}_{\mathcal{P}} \colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to \widetilde{\mathbb{Z}}_{\mathcal{P}}$ obtained by taking the direct sum of all the $\mathfrak{i}_p$ for $p \in \mathcal{P}$.

Remark 6.1 Each $\widetilde{\mathbb{Z}}[p]$ naturally lies in $\widetilde{\mathrm{H}}_0(\partial\Sigma)$, and the total boundary map can be identified with the restriction to $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ of the usual boundary map $\mathrm{H}_1(\Sigma, \partial\Sigma) \to \widetilde{\mathrm{H}}_0(\partial\Sigma)$.

Symplectic support Now consider an A–homology marking μ on $(\Sigma, \mathcal{P})$. Back in Remark 5.1, we observed that a necessary condition for μ to be supported on a symplectic subsurface is that $\mathfrak{i}_{\mathcal{P}}(\ker(\mu)) = \widetilde{\mathbb{Z}}_{\mathcal{P}}$. The following lemma says that this condition is also sufficient:

Lemma 6.2 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf`. Then μ is supported on a symplectic subsurface if and only if $\mathfrak{i}_{\mathcal{P}}(\ker(\mu)) = \widetilde{\mathbb{Z}}_{\mathcal{P}}$.*

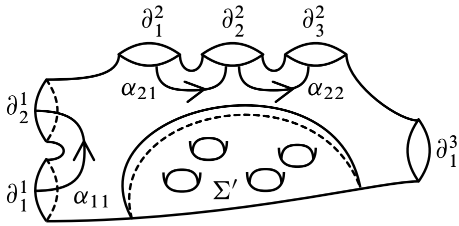
Proof The nontrivial direction is that if $\mathfrak{i}_{\mathcal{P}}(\ker(\mu)) = \widetilde{\mathbb{Z}}_{\mathcal{P}}$, then μ is supported on a symplectic subsurface, so that is what we prove. Write

$$\mathcal{P} = \{\{\partial_1^1, \dots, \partial_{k_1}^1\}, \{\partial_1^2, \dots, \partial_{k_2}^2\}, \dots, \{\partial_1^n, \dots, \partial_{k_n}^n\}\}.$$

Below we will prove that, for all $1 \le i \le n$ and $1 \le j < k_i$, we can find embedded arcs α_{ij} satisfying

- α_{ij} connects ∂_j^i to ∂_{j+1}^i,
- the α_{ij} are pairwise disjoint, and
- $\mu([\alpha_{ij}]) = 0$ for all i and j.

Letting g be the genus of Σ, we can then find a subsurface Σ' of Σ that is homeomorphic to Σ_g^1 such that Σ' is disjoint from $\partial\Sigma$ and the α_{ij}; see here:



Let $\mathcal{P}' = \{\partial\Sigma'\}$, so $(\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ is a $\mathtt{PSurf}$–morphism. It is easy to see that we can find an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ such that μ is the stabilization of μ' to $(\Sigma, \mathcal{P})$; see Lemma 6.3 for a more general result that implies this. The lemma follows.

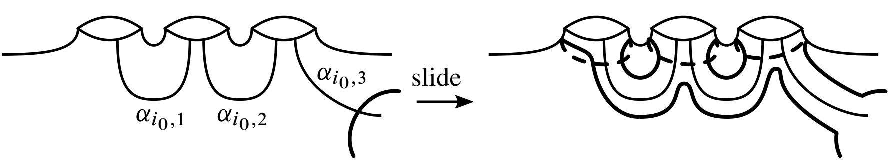
It remains to find the α_{ij}. The assumptions in the lemma imply that, for $1 \le i \le n$ and $1 \le j < k_i$, we can find arcs α_{ij} (not necessarily embedded or pairwise disjoint) satisfying

- α_{ij} connects ∂_j^i to ∂_{j+1}^i, and
- $\mu([\alpha_{ij}]) = 0$ for all i and j.

Homotoping the α_{ij}, we can assume that their endpoints are disjoint from each other, their interiors lie in the interior of Σ, and all intersections and self-intersections are transverse. Choose these α_{ij} so as to minimize the number of intersections and self-intersections. We claim that the α_{ij} are then all embedded and pairwise disjoint from each other. Assume otherwise. Let α_{i_0,j_0} be the first element of the ordered list

$$\alpha_{11}, \alpha_{12}, \dots, \alpha_{1,k_1-1}, \alpha_{21}, \dots, \alpha_{2,k_2-1}, \alpha_{31}, \dots, \alpha_{n,k_n-1}$$

that intersects either itself or one of the other α_{ij}. As in the following picture, we can then "slide" the first intersection of α_{i_0,j_0} off of the union of the $\partial_j^{i_0}$ and $\alpha_{i_0,j}$ with $j \le j_0$:



Since the homology classes of all the ∂_j^i are in the kernel of μ, this does not change the value of any of the $\mu([\alpha_{ij}])$, but it does eliminate one of the intersections, contradicting the minimality of this number. □

6.2 Destabilizing homology-marked partitioned surfaces

Consider $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. This section is devoted to "destabilizing" A–homology markings on $(\Sigma, \mathcal{P})$ to subsurfaces.

Existence Let μ be an A–homology marking on $(\Sigma, \mathcal{P})$ and let $(\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ be a $\mathtt{PSurf}$–morphism. One obvious necessary condition for there to exist an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ whose stabilization to $(\Sigma, \mathcal{P})$ is μ is that μ must vanish on elements of $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ supported on $\Sigma \setminus \Sigma'$. This condition is also sufficient:

Lemma 6.3 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let $(\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ be a $\mathtt{PSurf}$–morphism. Then there exists an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ whose stabilization to $(\Sigma, \mathcal{P})$ is μ if and only if $\mu(x) = 0$ for all $x \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ supported on $\Sigma \setminus \Sigma'$.*

Proof The nontrivial assertion here is that, if $\mu(x) = 0$ for all $x \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ supported on $\Sigma \setminus \Sigma'$, then μ' exists, so this is what we prove. Let $\iota\colon (\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ be the inclusion. We want to show that $\mu\colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to A$ factors through

$$\iota^*\colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma').$$

The cokernel of ι^* is obviously free abelian. It is thus enough to prove that μ vanishes on $\ker(\iota^*)$. To do this, we will show that $\ker(\iota^*)$ is generated by elements supported on $\Sigma \setminus \Sigma'$. The map ι^* is the restriction to $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ of the composition

$$\mathrm{H}_1(\Sigma, \partial\Sigma) \xrightarrow{f} \mathrm{H}_1(\Sigma, \Sigma \setminus \mathrm{Int}(\Sigma')) \xrightarrow{\cong} \mathrm{H}_1(\Sigma', \partial\Sigma').$$

It is thus enough to show that all elements of $\ker(f)$ are supported on $\Sigma \setminus \Sigma'$. The long exact sequence in homology for the triple $(\Sigma, \Sigma \setminus \mathrm{Int}(\Sigma'), \partial\Sigma)$ implies that $\ker(f)$ is generated by the image of

$$\mathrm{H}_1(\Sigma \setminus \mathrm{Int}(\Sigma'), \partial\Sigma) \to \mathrm{H}_1(\Sigma, \partial\Sigma).$$

The desired result follows. □

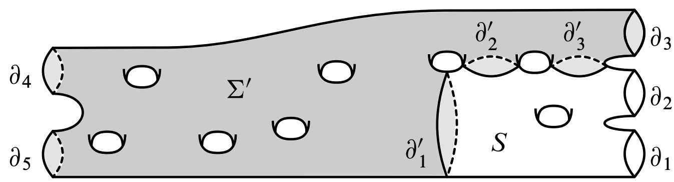
$\mathcal{P}$–simple subsurfaces We now study when destabilizations of markings supported on symplectic subsurfaces are supported on symplectic subsurfaces. Rather than prove the most general result possible, we will focus on the case of *$\mathcal{P}$–simple subsurfaces* of Σ, which are subsurfaces Σ' satisfying (see Example 6.4 below):

- Σ' is connected.
- The closure S of $\Sigma \setminus \Sigma'$ is connected.

- The set of components of ∂S can be partitioned into two disjoint nonempty subsets q and q' as follows:
  - The elements of q' all lie in the interior of Σ, and are thus components of $\partial\Sigma'$. These will be called the *interior boundary components*.
  - The elements of q are components of $\partial\Sigma \setminus \partial\Sigma'$ lying in a single $p \in \mathcal{P}$. These will be called the *exterior boundary components*.

Given a $\mathcal{P}$–simple subsurface Σ' of Σ, the *induced partition* $\mathcal{P}'$ of the components of $\partial\Sigma'$ is obtained from $\mathcal{P}$ by replacing p with $(p \setminus q) \cup q'$, where p, q and q' are as above. The map $(\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ is clearly a `PSurf`–morphism.

Example 6.4 Let $\Sigma = \Sigma_8^5$ and $\mathcal{P} = \{\{\partial_1, \partial_2, \partial_3\}, \{\partial_4, \partial_5\}\}$. Consider the following shaded subsurface Σ' of Σ:



The subsurface Σ' is a $\mathcal{P}$–simple subsurface with interior boundary components $q' = \{\partial_1', \partial_2', \partial_3'\}$, exterior boundary components $q = \{\partial_1, \partial_2\}$, and induced partition $\mathcal{P}' = \{\{\partial_1', \partial_2', \partial_3', \partial_3\}, \{\partial_4, \partial_5\}\}$.

Closed markings and intersection maps We now introduce some notation needed to state our result. Let $(\Sigma, \mathcal{P}) \in$ `PSurf` and let μ be an A–homology marking on $(\Sigma, \mathcal{P})$. The associated *closed marking* on $(\Sigma, \mathcal{P})$ is the map $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ defined via the composition

$$\mathrm{H}_1(\Sigma) \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \xrightarrow{\mu} A.$$

Also, for a finite set q of oriented simple closed curves on Σ, define the *closed q–intersection map* to be the map $\hat{\mathfrak{i}}_q\colon \mathrm{H}_1(\Sigma) \to \mathbb{Z}[q]$ defined via the composition

$$\mathrm{H}_1(\Sigma) \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \xrightarrow{\mathfrak{i}_q} \mathbb{Z}[q].$$

If the elements of q are disjoint and their union bounds a subsurface on one side (with respect to the orientations on the curves of q), then the image of $\hat{\mathfrak{i}}_q$ lies in $\widetilde{\mathbb{Z}}[q]$.

Destabilizing and symplectic support With the above notation, we have the following lemma:

Lemma 6.5 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P})$ that is supported on a symplectic subsurface. Let Σ' be a $\mathcal{P}$–simple subsurface of Σ with induced partition $\mathcal{P}'$ and let μ' be an A–homology marking on $(\Sigma', \mathcal{P}')$ whose stabilization to $(\Sigma, \mathcal{P})$ is μ. Assume the following:*

- *Let q' be the interior boundary components of Σ' and let $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ. Orient each $\partial' \in q'$ so that Σ' lies to its left. Then $\hat{\mathfrak{i}}_{q'}(\ker(\hat{\mu})) = \widetilde{\mathbb{Z}}[q']$.*

Then μ' is supported on a symplectic subsurface.

Proof By Lemma 6.2, we must prove that the map

$$\mathfrak{i}_{\mathcal{P}'}\colon \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \to \widetilde{\mathbb{Z}}_{\mathcal{P}'}$$

takes $\ker(\mu')$ onto $\widetilde{\mathbb{Z}}_{\mathcal{P}'}$. Below we will prove two facts:

- $\widetilde{\mathbb{Z}}[q'] \subset \mathfrak{i}_{\mathcal{P}'}(\ker(\mu'))$.
- Letting $\iota\colon (\Sigma', \mathcal{P}') \to (\Sigma, \mathcal{P})$ be the inclusion and $\iota^*\colon \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \to \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma')$ be the induced map, there exists a surjection $\beta\colon \widetilde{\mathbb{Z}}_{\mathcal{P}} \twoheadrightarrow \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q']$ such that the diagram

(6-1)
$$\begin{tikzcd}
\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \arrow[rr, "\mathfrak{i}_{\mathcal{P}}"] \arrow[d, "\iota^*"] & & \widetilde{\mathbb{Z}}_{\mathcal{P}} \arrow[d, two heads, "\beta"] \\
\mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \arrow[r, "\mathfrak{i}_{\mathcal{P}'}"] & \widetilde{\mathbb{Z}}_{\mathcal{P}'} \arrow[r, two heads, "\pi"] & \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q']
\end{tikzcd}$$

  commutes.

Assume these, for the moment. Since $\widetilde{\mathbb{Z}}[q'] \subset \mathfrak{i}_{\mathcal{P}'}(\ker(\mu'))$, to prove $\mathfrak{i}_{\mathcal{P}'}(\ker(\mu')) = \widetilde{\mathbb{Z}}_{\mathcal{P}'}$ it is enough to prove that $\pi\big(\mathfrak{i}_{\mathcal{P}'}(\ker(\mu'))\big) = \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q']$. Since μ is supported on a symplectic subsurface, Lemma 6.2 says that $\mathfrak{i}_{\mathcal{P}}(\ker(\mu)) = \widetilde{\mathbb{Z}}_{\mathcal{P}}$, so

(6-2)
$$\pi\big(\mathfrak{i}_{\mathcal{P}'}\big(\iota^*(\ker(\mu))\big)\big) = \beta\big(\mathfrak{i}_{\mathcal{P}}(\ker(\mu))\big) = \beta(\widetilde{\mathbb{Z}}_{\mathcal{P}}) = \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q'].$$

Since μ is the stabilization of μ' to $(\Sigma, \mathcal{P})$, by definition we have $\mu = \mu' \circ \iota^*$, so $\iota^*(\ker(\mu)) \subset \ker(\mu')$. Plugging this into (6-2), we get that

$$\pi\big(\mathfrak{i}_{\mathcal{P}'}(\ker(\mu'))\big) = \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q'],$$

as desired.

It remains to prove the above two facts. We start with the first. Since elements of $\mathrm{H}_1(\Sigma)$ can be represented by cycles that are disjoint from all components of $\partial\Sigma$, the image of the composition

$$\mathrm{H}_1(\Sigma) \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma) \xrightarrow{\iota^*} \mathrm{H}_1^{\mathcal{P}'}(\Sigma', \partial\Sigma') \xrightarrow{\mathfrak{i}_{\mathcal{P}'}} \widetilde{\mathbb{Z}}_{\mathcal{P}'}$$

must lie in $\widetilde{\mathbb{Z}}[q'] \subset \widetilde{\mathbb{Z}}_{\mathcal{P}'}$. From its definition, it is clear that this composition in fact equals $\hat{\mathrm{i}}_{q'}$. Our hypothesis about $\hat{\mathrm{i}}_{q'}$ thus implies that

$$\widetilde{\mathbb{Z}}[q'] \subset \mathrm{i}_{\mathcal{P}'}\big(\iota^*(\ker(\mu))\big) \subset \mathrm{i}_{\mathcal{P}'}(\ker(\mu')),$$

as desired. Here we are using the fact (already observed in the previous paragraph) that $\iota^*(\ker(\mu)) \subset \ker(\mu')$.

We now construct $\beta\colon \widetilde{\mathbb{Z}}_{\mathcal{P}} \twoheadrightarrow \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q']$. Let q be the exterior boundary components of Σ'. Write $\mathcal{P} = \{p_1, \dots, p_k\}$ with $q \subset p_1$. Setting $p_1' = (p_1 \setminus q) \cup q'$, we then have $\mathcal{P}' = \{p_1', p_2, \dots, p_k\}$. Thus

$$\widetilde{\mathbb{Z}}_{\mathcal{P}} = \widetilde{\mathbb{Z}}[p_1] \oplus \bigoplus_{i=2}^{k} \widetilde{\mathbb{Z}}[p_i] \quad \text{and} \quad \widetilde{\mathbb{Z}}_{\mathcal{P}'}/\widetilde{\mathbb{Z}}[q'] = \widetilde{\mathbb{Z}}[p_1']/\widetilde{\mathbb{Z}}[q'] \oplus \bigoplus_{i=2}^{k} \widetilde{\mathbb{Z}}[p_i].$$

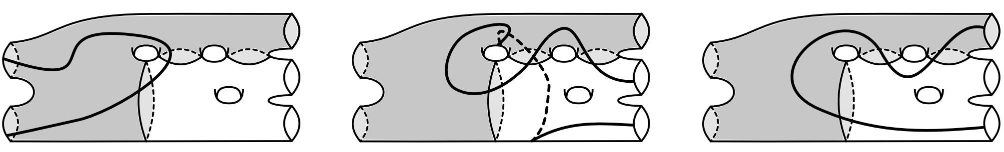
On the $\widetilde{\mathbb{Z}}[p_i]$ summand for $2 \le i \le k$, the map β is the identity. On the $\widetilde{\mathbb{Z}}[p_1]$ summand, the map β is the restriction to $\widetilde{\mathbb{Z}}[p_1]$ of the map

$$\mathbb{Z}[p_1] = \mathbb{Z}[p_1 \setminus q] \oplus \mathbb{Z}[q] \to \mathbb{Z}[p_1 \setminus q] \oplus \mathbb{Z}[q']/\widetilde{\mathbb{Z}}[q'] = \widetilde{\mathbb{Z}}[p_1']/\widetilde{\mathbb{Z}}[q']$$

that is the identity on $\mathbb{Z}[p_1 \setminus q]$ and takes every element of q to the generator of $\mathbb{Z}[q']/\widetilde{\mathbb{Z}}[q'] \cong \mathbb{Z}$. This map β is clearly a surjection. The fact that (6-1) commutes follows from the fact that an arc in Σ from a component ∂_1 of $\partial\Sigma$ to a component ∂_2 of $\partial\Sigma$ with ∂_1 and ∂_2 both lying in some p_i has the following algebraic intersection number with the union of the components of q:

- 0 if $i \ge 2$, if $i = 1$ and $\partial_1, \partial_2 \in p_1 \setminus q$, or if $i = 1$ and $\partial_1, \partial_2 \in q$,
- 1 if $i = 1$, $\partial_1 \in p_1 \setminus q$ and $\partial_2 \in q$,
- -1 if $i = 1$, $\partial_2 \in q$ and $\partial_1 \in p_1 \setminus q$.

The reason for this is that each time the arc crosses from Σ' to S it adds $+1$ to its total intersection with q, while each time it crosses from S to Σ' it adds -1 to its total intersection with q. See the following figure, where Σ is shaded and S is unshaded:



The lemma follows. $\square$

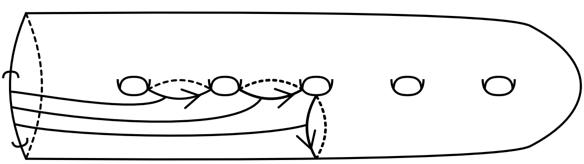
6.3 The complex of tethered vanishing loops

We now begin our long trek to the complex of order-preserving double-tethered vanishing loops, starting with the complex of tethered vanishing loops. The definition takes several steps.

Tethered loops Define $\tau(S^1)$ to be the result of gluing $1 \in [0,1]$ to a point of S^1. The subset $[0,1] \in \tau(S^1)$ is the *tether* and $0 \in [0,1] \subset \tau(S^1)$ is the *initial point* of the tether. For a surface $\Sigma \in \mathtt{Surf}$ and a finite disjoint union of open intervals $I \subset \partial\Sigma$, an *I–tethered loop* in Σ is an embedding $\iota\colon \tau(S^1) \to \Sigma$ such that

- ι takes the initial point of the tether to a point of I, and
- orienting $\iota(S^1)$ using the natural orientation of S^1, the image $\iota([0,1])$ of the tether approaches $\iota(S^1)$ from its right.

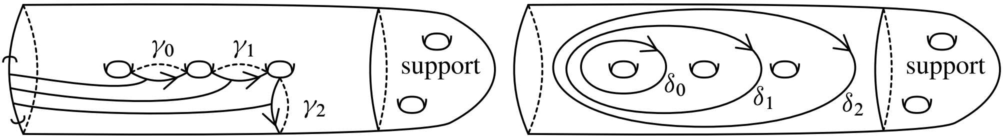
Complex of tethered loops For a surface $\Sigma \in \mathtt{Surf}$ and a finite disjoint union of open intervals $I \subset \partial\Sigma$, the *complex of I–tethered loops* on Σ, denoted by $\mathcal{TL}(\Sigma, I)$, is the simplicial complex whose k–simplices are collections $\{\iota_0, \dots, \iota_k\}$ of isotopy classes of I–tethered loops on Σ that can be realized so as to be disjoint and not separate Σ:



This complex was introduced by Hatcher and Vogtmann [12], who proved that if Σ has genus g then $\mathcal{TL}(\Sigma, I)$ is $\frac{1}{2}(g-3)$–connected; see [12, Proposition 5.1].

Complex of tethered vanishing loops Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let $I \subset \partial\Sigma$ be a finite disjoint union of open intervals. Define $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ to be the subcomplex of $\mathcal{TL}(\Sigma, I)$ consisting of k–simplices $\{\iota_0, \dots, \iota_k\}$ satisfying the following conditions. For $0 \le i \le k$, let γ_i be the oriented loop $(\iota_i)|_{S^1}$. Set $\Gamma = \{\gamma_0, \dots, \gamma_k\}$. As in Section 6.2, let $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ and let $\hat{\mathfrak{i}}_\Gamma\colon \mathrm{H}_1(\Sigma) \to \mathbb{Z}[\Gamma]$ be the closed Γ–intersection map. We then require that $\hat{\mu}([\gamma_i]) = 0$ for all $0 \le i \le k$ and that $\hat{\mathfrak{i}}_\Gamma(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma]$.

Remark 6.6 This last condition might seem a little unmotivated, but is needed to ensure that the stabilizer of our simplex is supported on a symplectic subsurface (at least in favorable situations). It clearly always holds when μ is supported on a symplectic subsurface that is disjoint from the images of all the ι_i. This is best illustrated by an example:



If $\Gamma = \{\gamma_0, \gamma_1, \gamma_2\}$ and $\delta_0, \delta_1, \delta_2$ are as shown, then $\hat{\mu}([\delta_i]) = 0$ and $\hat{\mathfrak{i}}_\Gamma([\delta_i]) = \gamma_i$ for $0 \le i \le 2$, which implies that $\hat{\mathfrak{i}}_\Gamma(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma]$.

High connectivity Our main topological theorem about $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ is as follows:

Theorem 6.7 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$, $I \subset \partial\Sigma$ be a finite disjoint union of open intervals, and g be the genus of Σ. Then $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ is $\big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2)$–connected.*

Proof The proof is very similar to that of Theorem 3.8. We start by defining an auxiliary space. Let X be the simplicial complex whose vertices are the union of the vertices of the spaces $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ and $\mathcal{TS}_{\operatorname{rk}(A)+1}(\Sigma, I)$ and whose simplices are collections σ of vertices such that:

- The vertices in σ (which are embeddings of either $\tau(S^1)$ or $\tau(\Sigma^1_{\operatorname{rk}(A)+1})$ into Σ) can be homotoped so that their images are disjoint and do not separate Σ.
- Let $\sigma' \subset \sigma$ be the subset consisting of vertices of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$. Then σ' is a simplex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$.

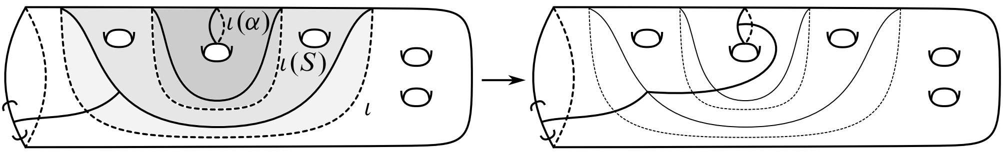
Both $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ and $\mathcal{TS}_{\operatorname{rk}(A)+1}(\Sigma, I)$ are subcomplexes of X.

The subcomplex $\mathcal{TS}_{\operatorname{rk}(A)+1}(\Sigma, I)$ of X is $\big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2)$–connected by Theorem D. An argument using Corollary 2.4 identical to the one in the proof of Theorem 3.8 shows that this implies that X is $\big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2)$–connected. As in the proof of Theorem 3.8, this implies that it is enough to construct a retraction $r\colon X \to \mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$.

For a vertex ι of X, we define $r(\iota)$ as follows. If ι is a vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, then $r(\iota) = \iota$. If instead ι is a vertex of $\mathcal{TS}_{\operatorname{rk}(A)+1}(\Sigma, I)$, then we do the following. Let $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ. Define $\mu'\colon \mathrm{H}_1(\Sigma^1_{\operatorname{rk}(A)+1}) \to A$ to be the composition

$$\mathrm{H}_1(\Sigma^1_{\operatorname{rk}(A)+1}) \cong \mathrm{H}_1(\tau(\Sigma^1_{\operatorname{rk}(A)+1})) \xrightarrow{\iota_*} \mathrm{H}_1(\Sigma) \xrightarrow{\hat{\mu}} A.$$

Proposition 3.4 implies that there exists a subsurface $S \subset \Sigma^1_{\mathrm{rk}(A)+1}$ with $S \cong \Sigma^1_1$ and $\mu'|_{\mathrm{H}_1(S)} = 0$. Let α be a nonseparating oriented simple closed curve in S. Define $r(\iota)$ to be the vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ obtained by adjoining the tether of ι and an arbitrary arc in $\iota(\Sigma^1_{\mathrm{rk}(A)+1})$ to $\iota(\alpha)$; see here:



To see that this is actually a vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, observe that by construction

$$\hat{\mu}([\iota(\alpha)]) = 0, \quad \iota_*(\mathrm{H}_1(S)) \subset \ker(\hat{\mu}) \quad \text{and} \quad \hat{\mathrm{i}}_{\{\iota(\alpha)\}}\big(\iota_*(\mathrm{H}_1(S))\big) = \iota(\alpha).$$

Of course, $r(\iota)$ depends on various choices, but we simply make an arbitrary choice.

To complete the proof, we must show that r extends over the simplices of X. Let σ be a simplex of X. Enumerate the vertices of σ as $\{\iota_0, \dots, \iota_k, \iota'_0, \dots, \iota'_\ell\}$, where the ι_i are vertices of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ and the ι'_j are vertices of $\mathcal{TS}_{\mathrm{rk}(A)+1}(\Sigma, I)$. We must prove that

$$r(\sigma) = \{\iota_0, \dots, \iota_k, r(\iota'_0), \dots, r(\iota'_\ell)\}$$

is a simplex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$. The images of the vertices in $r(\sigma)$ can clearly be homotoped so as to be disjoint, so the only thing we must prove is the following. For $0 \le i \le k$ and $0 \le j \le \ell$, let $\gamma_i = \iota_i|_{S^1}$ and $\gamma'_j = \iota'_j|_{S^1}$. Setting $\Gamma = \{\gamma_0, \dots, \gamma_k, \gamma'_0, \dots, \gamma'_\ell\}$, we have to show that $\hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma]$. Setting $\Gamma_1 = \{\gamma_0, \dots, \gamma_k\}$ and $\Gamma_2 = \{\gamma'_0, \dots, \gamma'_\ell\}$, we will show that Γ_1 and Γ_2 are both contained in $\hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu}))$.

We start with Γ_2. By construction, for $0 \le j \le \ell$ there exists a subsurface S_j of Σ with $S_j \cong \Sigma^1_1$ such that

- $\gamma'_j \subset S_j$, and
- the S_j are disjoint from each other and from all the γ_i, and
- regarding $\mathrm{H}_1(S_j)$ as a subgroup of $\mathrm{H}_1(\Sigma)$, we have $\mathrm{H}_1(S_j) \subset \ker(\hat{\mu})$.

Since $\hat{\mathrm{i}}_\Gamma(\mathrm{H}_1(S_j)) = \gamma'_j$, we have $\gamma'_j \in \hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu}))$, as desired.

It remains to show that $\Gamma_1 \subset \hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu}))$. Since $\{\iota_0, \dots, \iota_k\}$ is a simplex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, by definition we have $\hat{\mathrm{i}}_{\Gamma_1}(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma_1]$. For some $0 \le i \le k$, let $x \in \ker(\hat{\mu})$ be such that $\hat{\mathrm{i}}_{\Gamma_1}(x) = \gamma_i$. We then have $\hat{\mathrm{i}}_\Gamma(x) = \gamma_i + z$ with $z \in \mathbb{Z}[\Gamma_2]$. Since we already showed that $\mathbb{Z}[\Gamma_2] \subset \hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu}))$, we conclude that $\gamma_i \in \hat{\mathrm{i}}_\Gamma(\ker(\hat{\mu}))$, as desired. □

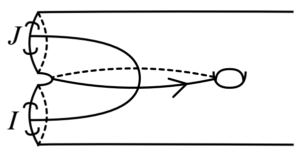
6.4 The complex of double-tethered vanishing loops

The definition of the complex of double-tethered vanishing loops takes several steps.

Double-tethered loops Define $\tau^2(S^1)$ to be the result of gluing $1 \in [0,2]$ to S^1. We will call $[0,2] \subset \tau^2(S^1)$ the *double tether*; the point $0 \in [0,2]$ is the double tether's *initial point* and $2 \in [0,2]$ is its *terminal point*. For a surface $\Sigma \in \mathtt{Surf}$ and finite disjoint unions of open intervals $I, J \subset \partial\Sigma$ with $I \cap J = \varnothing$, an (I,J)*–double-tethered loop* in Σ is an embedding $\iota\colon \tau^2(S^1) \to \Sigma$ where

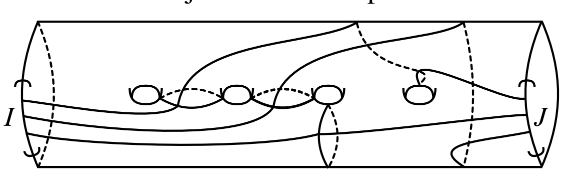
- ι takes the initial point of the double tether to a point of I and the terminal point of the double tether to a point of J, and
- orienting $\iota(S^1)$ using the natural orientation of S^1, the image $\iota([0,1])$ approaches $\iota(S^1)$ from its right and the image $\iota([1,2])$ leaves $\iota(S^1)$ from its left.

See here:



We remark that right now we allow boundary components that contain components of both I and J, but later when we discuss double-tethered vanishing loops our hypotheses will exclude this possibility.

Complex of double-tethered loops For a surface $\Sigma \in \mathtt{Surf}$ and finite disjoint unions of open intervals $I, J \subset \partial\Sigma$ with $I \cap J = \varnothing$, the *complex of* (I,J)*–double-tethered loops* on Σ, denoted by $\mathcal{DTL}(\Sigma, I, J)$, is the simplicial complex whose k–simplices are collections $\{\iota_0, \ldots, \iota_k\}$ of isotopy classes of (I,J)–double-tethered loops on Σ that can be realized so as to be disjoint and not separate Σ. See here:



This complex was introduced by Hatcher and Vogtmann [12], who proved that if Σ has genus g then, like $\mathcal{TL}(\Sigma, I)$, it is $\frac{1}{2}(g-3)$–connected; see [12, Proposition 5.2].

$\mathcal{P}$–adjacency Consider $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let $I, J \subset \partial\Sigma$ be finite disjoint unions of open intervals with $I \cap J = \varnothing$. Recall that components ∂ and ∂' of $\partial\Sigma$ are said to be $\mathcal{P}$–adjacent if there exists some $p \in \mathcal{P}$ such that $\partial, \partial' \in p$. We will say that I and J are

$\mathcal{P}$*–adjacent* if, for all components ∂_I and ∂_J of $\partial\Sigma$ such that ∂_I contains a component of I and ∂_J contains a component of J, the components ∂_I and ∂_J are distinct and $\mathcal{P}$–adjacent.

Complex of double-tethered vanishing loops Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent disjoint unions of open intervals with $I \cap J = \varnothing$. In particular, there are no boundary components of Σ containing components of both I and J. Define $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ to be the subcomplex of $\mathcal{DTL}(\Sigma, I, J)$ consisting of k–simplices $\{\iota_0, \dots, \iota_k\}$ satisfying the following conditions. Let $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ.

- For $0 \le i \le k$, let γ_i be the oriented loop $(\iota_i)|_{S^1}$ and α_i be the oriented arc $(\iota_i)|_{[0,2]}$. We then require that $\hat{\mu}([\iota_i]) = 0$ and $\mu([\alpha_i]) = 0$. This second condition makes sense since I and J are $\mathcal{P}$–adjacent.
- Set $\Gamma = \{\gamma_0, \dots, \gamma_k\}$. We then require that $\hat{\mathfrak{i}}_\Gamma(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma]$.

Identifying $\tau(S^1)$ with the union of $[0,1]$ and S^1 in $\tau^2(S^1)$, these conditions imply that $\{(\iota_0)|_{\tau(S^1)}, \dots, (\iota_k)|_{\tau(S^1)}\}$ is a simplex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$.

6.5 The complex of mixed-tethered vanishing loops

Our main theorem about the complex of double-tethered vanishing loops says that it is highly connected. We will prove this in Section 6.6 below. This section is devoted to an intermediate complex that will play a technical role in that proof.

Complex of mixed-tethered vanishing loops Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent disjoint unions of open intervals with $I \cap J = \varnothing$. Let $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ. Define $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$ to be the simplicial complex whose k–simplices are sets $\{\iota_0, \dots, \iota_k\}$, where each ι_i is the isotopy class of either an I–tethered loop or an (I, J)–double-tethered loop and where the following conditions are satisfied:

- The ι_i can be realized so that their images are disjoint and do not separate Σ.
- For $0 \le i \le k$, let γ_i be the oriented loop $(\iota_i)|_{S^1}$. We then require that $\hat{\mu}([\gamma_i]) = 0$.
- For $0 \le i \le k$ such that ι_i is an (I, J)–double-tethered loop, let α_i be the oriented arc $(\iota_i)|_{[0,2]}$. We then require that $\mu([\alpha_i]) = 0$.
- Set $\Gamma = \{\gamma_0, \dots, \gamma_k\}$. We then require that $\hat{\mathfrak{i}}_\Gamma(\ker(\hat{\mu})) = \mathbb{Z}[\Gamma]$.

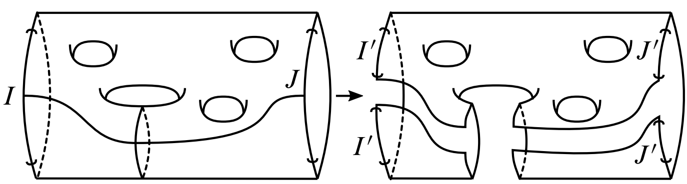
These conditions ensure that both $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ and $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ are full subcomplexes of $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$.

Links Our first task will be to identify links in $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$.

Lemma 6.8 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \texttt{PSurf}$. Let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent finite disjoint unions of open intervals with $I \cap J = \varnothing$. Finally, let σ be a k–simplex of $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$. Then there exists some $(\Sigma', \mathcal{P}') \in \texttt{PSurf}$, an A–homology marking μ' on $(\Sigma', \mathcal{P}')$, and $\mathcal{P}'$–adjacent finite disjoint unions of open intervals $I', J' \subset \partial\Sigma'$ with $I' \cap J' = \varnothing$ such that*:

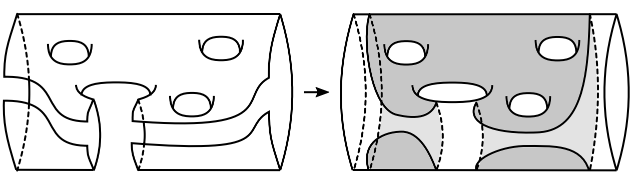
- *The link of σ is isomorphic to $\mathcal{MTL}(\Sigma', I', J', \mathcal{P}', \mu')$. Moreover, the intersections of the link of σ with $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$ and $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ are $\mathcal{TL}(\Sigma', I', \mathcal{P}', \mu')$ and $\mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$, respectively.*
- *If Σ is a genus-g surface, then Σ' is a genus-$(g-k-1)$ surface.*
- *If μ is supported on a symplectic subsurface, then so is μ'.*

Proof It is enough to deal with the case where σ has dimension 0; the general case will then follow by applying the dimension 0 case repeatedly. We thus can assume that $\sigma = \{\iota\}$, where ι is either an I–tethered loop or an (I, J)–double-tethered loop. The two cases are similar, so we will give the details for when ι is an (I, J)–double-tethered loop. Let Σ' be the result of cutting Σ open along the image of ι:



We remark that the fact that I and J are $\mathcal{P}$–adjacent implies that the initial and terminal points of the double tether are on distinct boundary components.

By isotoping Σ' into the interior of Σ, we can regard Σ' as a $\mathcal{P}$–simple subsurface of Σ:



Let $\mathcal{P}'$ be the induced partition of the components of $\partial\Sigma'$. By Lemma 6.3, there exists an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ such that μ is the stabilization of μ' to $(\Sigma, \mathcal{P})$.

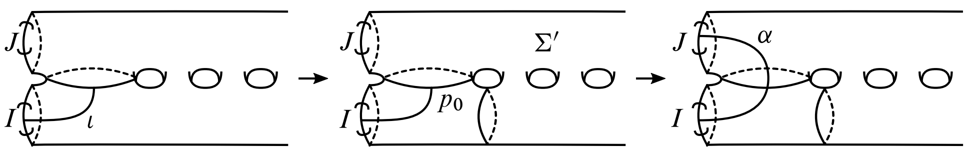
As is clear from the above figure, when forming Σ' the sets I and J are divided into finer collections I' and J' of open intervals in $\partial\Sigma'$ such that the link of σ is isomorphic to $\mathcal{MTL}(\Sigma', I', J', \mathcal{P}', \mu')$. By construction, Σ' has genus $g-1$. The only thing that remains to be proved is that if μ is supported on a symplectic subsurface, then so is μ'. Letting $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ be the closed marking associated to μ and q be the interior boundary components of Σ' (as in the definition of a $\mathcal{P}$–simple subsurface in Section 6.2), Lemma 6.5 says that it is enough to prove that $\hat{\mathfrak{i}}_q(\ker(\hat{\mu})) = \widetilde{\mathbb{Z}}[q]$.

Let $\gamma = \iota|_{S^1}$. Since ι is a vertex of $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$, there exists some $x \in \ker(\hat{\mu})$ such that $\hat{\mathfrak{i}}_\gamma(x) = \gamma$. By construction, we have $q = \{\gamma_1, \gamma_2\}$, where γ_1 (resp. γ_2) is obtained by band-summing γ with a component of $\partial\Sigma$ containing a component of I (resp. J). The orientations on the γ_i are such that γ_1 is homologous in $\mathrm{H}_1(\Sigma, \partial\Sigma)$ to γ and γ_2 is homologous to $-\gamma$. It follows that $\hat{\mathfrak{i}}_q(x) = \gamma_1 - \gamma_2$, which generates $\widetilde{\mathbb{Z}}[q]$. The lemma follows. □

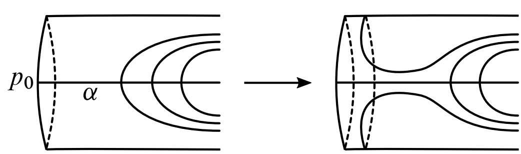
Completing a tethered loop to a double-tethered loop As a first application of Lemma 6.8 (or, rather, its proof), we prove the following:

Lemma 6.9 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf` that is supported on a symplectic subsurface. Let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent finite disjoint unions of open intervals with $I \cap J = \varnothing$. Then, for all vertices ι of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, there exists a vertex $\hat{\iota}$ of $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ such that $\hat{\iota}|_{\tau(S^1)} = \iota$.*

Proof Let $(\Sigma', \mathcal{P}')$ and I', J' and μ' be the output of applying Lemma 6.8 to the 0–simplex $\{\iota\}$ of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu) \subset \mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$. The A–homology marking μ' on $(\Sigma', \mathcal{P}')$ is thus supported on a symplectic subsurface. As in the following figure, it is enough to find an embedded arc α in Σ' connecting the endpoint p_0 of the tether of ι to a point of J such that $\mu'([\alpha]) = 0$:



Since μ' is supported on a symplectic subsurface, Lemma 6.2 implies that there exists an immersed arc α (not necessarily embedded) connecting p_0 to a point of J such that $\mu'([\alpha]) = 0$. Choose α so as to have the fewest possible self-intersections. Then α is embedded; indeed, if it has a self-intersection, then as in the following figure we can "comb" its first self-intersection over the component of $\partial\Sigma'$ containing p_0:



This has the effect of removing a self-intersection from α, but since μ' vanishes on all components of $\partial\Sigma'$ it does not change the fact that $\mu'([\alpha]) = 0$. The lemma follows. $\square$

High connectivity We close this section by proving that $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is highly connected.

Theorem 6.10 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$. Let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent finite disjoint unions of open intervals with $I \cap J = \varnothing$ and let g be the genus of Σ. Then $\mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is $\big(g-(2\,\mathrm{rk}(A)+3)\big)/(\mathrm{rk}(A)+2)$–connected.*

Proof Set $n = \big(g - (2\,\mathrm{rk}(A) + 3)\big)/(\mathrm{rk}(A) + 2)$ and

$$X = \mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu), \quad Y = \mathcal{TL}(\Sigma, I, J, \mathcal{P}, \mu) \quad \text{and} \quad Y' = \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu).$$

Theorem 6.7 says that Y is n–connected, so it is enough to prove that the pair (X, Y) is n–connected. To do this, we will apply Corollary 2.4. This requires showing the following. Let σ be a k–dimensional simplex of Y' and let L be the link of σ in X. Then we must show that $L \cap Y$ is $(n-k-1)$–connected.

Lemma 6.8 says that $L \cap Y \cong \mathcal{TL}(\Sigma', I', J', \mathcal{P}', \mu')$, where Σ', I', J', $\mathcal{P}'$ and μ' are as follows:

- $(\Sigma', \mathcal{P}') \in \mathtt{PSurf}$ with Σ' a genus-$(g-k-1)$ surface.
- μ' is an A–homology marking on $(\Sigma', \mathcal{P}')$.
- $I', J' \subset \partial\Sigma'$ are $\mathcal{P}'$–adjacent finite disjoint unions of open intervals satisfying $I' \cap J' = \varnothing$.

Theorem 6.7 thus says that $L \cap Y$ is n'–connected for

$$n' = \frac{g' - (2\,\mathrm{rk}(A) + 3)}{\mathrm{rk}(A) + 2} = \frac{g - (2\,\mathrm{rk}(A) + 3)}{\mathrm{rk}(A) + 2} - \frac{k+1}{\mathrm{rk}(A) + 2} \geq n - k - 1. \qquad \square$$

6.6 High connectivity of the complex of double-tethered vanishing loops

In this section, we finally prove that the complex of double-tethered vanishing loops is highly connected:

Theorem 6.11 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent finite disjoint unions of open intervals with $I \cap J = \varnothing$ and let g be the genus of Σ. Then $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is $\big(g-(2\,\mathrm{rk}(A)+3)\big)/(\mathrm{rk}(A)+2)$–connected.*

The proof of Theorem 6.11 requires the following lemma. Say that a simplicial map $f\colon M \to X$ between simplicial complexes is *locally injective* if $f|_{\sigma}$ is injective for all simplices σ of M.

Lemma 6.12 *Let M be a compact n–dimensional manifold (possibly with boundary) equipped with a combinatorial triangulation, let X be a simplicial complex, and let $f\colon M \to X$ be a simplicial map. Assume*

- *$f|_{\partial M}$ is locally injective,*
- *for all simplices σ of X, the link of σ in X is $(n-\dim(\sigma)-2)$–connected.*

Then after possibly subdividing simplices of M lying in its interior, f is homotopic through maps fixing ∂M to a simplicial map $f'\colon M \to X$ that is locally injective.

Proof We remark that the proof of this is very similar to Hatcher and Vogtmann's proof of Proposition 2.3 above, though it seems hard to deduce it from that proposition. This result is also related to [10, Theorem 2.4].

The proof will be by induction on n. The base case $n = 0$ is trivial, so assume that $n > 0$ and that the result is true for all smaller dimensions. Call a simplex σ of M a *noninjective simplex* if, for all vertices v of σ, there exists a vertex v' of σ with $v \neq v'$ but $f(v) = f(v')$. If M has no noninjective simplices, then we are done. Assume, therefore, that M has noninjective simplices, and let σ be a noninjective simplex of M whose dimension is as large as possible. Since no simplices of ∂M are noninjective, the simplex σ does not lie in ∂M. Letting $L \subset M$ be the link of σ, this implies that $L \cong S^{n-\dim(\sigma)-1}$. Letting L' be the link of $f(\sigma)$ in X, the maximality of the dimension of σ implies two things:

- $f(L) \subset L'$.
- The restriction of f to L is locally injective.

Our assumptions imply that L' has connectivity at least

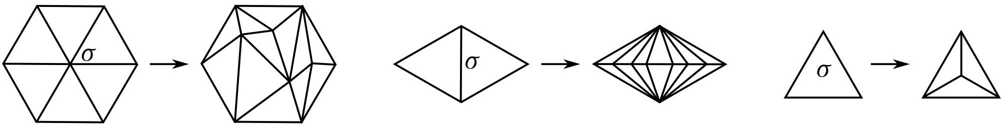
$$n-\dim(f(\sigma))-2 \geq n-(\dim(\sigma)-1)-2 = n-\dim(\sigma)-1.$$

Here we are using the fact that $f|_{\sigma}$ is not injective. We can thus extend $f|_L$ to a map

$$F\colon D^{n-\dim(\sigma)} \to L'$$

that is simplicial with respect to some combinatorial triangulation of $D^{n-\dim(\sigma)}$ that restricts to $L \cong S^{n-\dim(\sigma)-1}$ on $\partial D^{n-\dim(\sigma)}$. Since $\dim(\sigma) \geq 1$ and $F|_{\partial D^{n-\dim(\sigma)}} = f|_L$ is locally injective, we can apply our inductive hypothesis to F and ensure that F is locally injective. The star S of σ is isomorphic to the join $\sigma * L$. Subdividing M

and homotoping f, we can replace $S \subset D^{n-\dim(\sigma)}$ with $\partial\sigma * D^{n-\dim(\sigma)}$ and $f|_S$ with $f|_{\partial\sigma} * F$. Here are pictures of this operation for $n = 2$ and $\dim(\sigma) \in \{0, 1, 2\}$; on the left-hand side is S, and on the right-hand side is $\partial\sigma * D^{n-\dim(\sigma)}$:



In doing this, we have eliminated the noninjective simplex σ without introducing any new noninjective simplices. Repeating this over and over again, we can eliminate all noninjective simplices, and we are done. □

Proof of Theorem 6.11 We will prove by induction on n that $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is n–connected for $-1 \le n \le \big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2)$. The base case simply asserts that $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is nonempty when $\big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2) \ge -1$. In this case, Theorem 6.7 asserts that $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu) \neq \varnothing$, and thus Lemma 6.9 implies that $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu) \neq \varnothing$, as desired.

Assume now that $0 \le n \le \big(g-(2\operatorname{rk}(A)+3)\big)/(\operatorname{rk}(A)+2)$ and that all complexes $\mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$ as in the theorem are n'–connected for

$$n' = \min\left\{n-1, \frac{g'-(2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2}\right\},$$

where g' is the genus of Σ'. We must prove that $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is n–connected.

Set $X = \mathcal{MTL}(\Sigma, I, J, \mathcal{P}, \mu)$. The complex $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ that we want to show is n–connected is a subcomplex of X, and Theorem 6.10 says that the connectivity of X is at least

$$\frac{g-(2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2} \ge n.$$

Define Y to be the subcomplex of X consisting of simplices containing at most one vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, so

$$\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu) \subsetneq Y \subsetneq X.$$

The first step is as follows.

Claim 1 *The complex Y is n–connected.*

Proof We know that X is n–connected, so to prove that its subcomplex Y is n–connected it is enough to prove that the pair (X, Y) is $(n+1)$–connected. We will do this using Proposition 2.3. For this, we must identify a set $\mathcal{B}$ of "bad simplices" of

X and verify the three hypotheses of the proposition. Define $\mathcal{B}$ to be the set of all simplices of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu) \subset X$ whose dimension is at least 1.

We now verify the hypotheses of Proposition 2.3. The first two are easy:

(i) A simplex of X lies in Y if and only if none of its faces lie in $\mathcal{B}$, which is obvious.

(ii) If $\sigma_1, \sigma_2 \in \mathcal{B}$ are such that $\sigma_1 \cup \sigma_2$ is a simplex of X, then $\sigma_1 \cup \sigma_2 \in \mathcal{B}$, which again is obvious.

The only thing left to check is (iii), which says that for all k–dimensional $\sigma \in \mathcal{B}$, the complex $G(X, \sigma, \mathcal{B})$ has connectivity at least $(n+1) - k - 1 = n - k$.

Let L be the link of σ in X. Examining its definition in Section 2.2, we see that

$$G(X, \sigma, \mathcal{B}) \cong L \cap \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu).$$

Lemma 6.8 says that $L \cap \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu) \cong \mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$, where Σ', I', J', $\mathcal{P}'$ and μ' are as follows:

- $(\Sigma', \mathcal{P}') \in \mathtt{PSurf}$ with Σ' a genus $g' = g - k - 1$ surface.
- μ' is an A–homology marking on $(\Sigma', \mathcal{P}')$ that is supported on a symplectic subsurface.
- $I', J' \subset \partial\Sigma'$ are $\mathcal{P}'$–adjacent finite disjoint unions of open intervals satisfying $I' \cap J' = \varnothing$.

Our goal is thus to show that $\mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$ is $(n-k)$–connected. Our inductive hypothesis shows that $\mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$ is n'–connected for

$$\begin{aligned} n' = \min\left\{n-1, \frac{g' - (2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2}\right\} &= \min\left\{n-1, \frac{g - (2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2} - \frac{k+1}{\operatorname{rk}(A)+2}\right\} \\ &\geq \min\left\{n-1, n - \tfrac{1}{2}(k+1)\right\} \geq n-k. \end{aligned}$$

Here we are using the fact that, by the definition of $\mathcal{B}$, we have $k \geq 1$, and thus $k \geq \frac{1}{2}(k+1)$. □

This allows us to fill n–spheres in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ with $(n+1)$–discs in Y. We will modify these $(n+1)$–discs so that they lie in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$. For technical reasons, we will need our spheres and discs to be locally injective. That this is possible is the content of the following two steps.

Claim 2 *Equip the n–sphere S^n with a combinatorial triangulation and let $f: S^n \to \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ be a simplicial map. Then after possibly subdividing S^n, the map f is homotopic to a locally injective simplicial map.*

Proof By Lemma 6.12, this will follow if we can show that, for all k–simplices σ of $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$, the link L of σ is $(n-k-2)$–connected. Applying Lemma 6.8, we see that $L \cong \mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$, where Σ', I', J', $\mathcal{P}'$ and μ' are as follows:

- $(\Sigma', \mathcal{P}') \in \mathtt{PSurf}$ with Σ' a genus $g' = g-k-1$ surface.
- μ' is an A–homology marking on $(\Sigma', \mathcal{P}')$ that is supported on a symplectic subsurface.
- $I', J' \subset \partial\Sigma'$ are $\mathcal{P}'$–adjacent finite disjoint unions of open intervals satisfying $I' \cap J' = \varnothing$.

Our inductive hypothesis thus says that $L \cong \mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu')$ is n'–connected for

$$n' = \min\left\{n-1, \frac{g'-(2\,\mathrm{rk}(A)+3)}{\mathrm{rk}(A)+2}\right\} = \min\left\{n-1, \frac{g-(2\,\mathrm{rk}(A)+3)}{\mathrm{rk}(A)+2} - \frac{k+1}{\mathrm{rk}(A)+2}\right\}$$
$$\geq \min\left\{n-1, n - \frac{k+1}{\mathrm{rk}(A)+2}\right\} \geq n-k-2,$$

as desired. □

Claim 3 *Equip the n–sphere S^n with a combinatorial triangulation and let $f: S^n \to Y$ be a locally injective simplicial map that extends to a simplicial map of a combinatorial triangulation of D^{n+1}. Then there exists a combinatorial triangulation of D^{n+1} that restricts to our given triangulation on $\partial D^{n+1} = S^n$ and a locally injective simplicial map $F: D^{n+1} \to Y$ such that $F|_{\partial D^{n+1}} = f$.*

Proof By Lemma 6.12, this will follow if we can show that, for all k–simplices σ of Y, the link L of σ is $(n-k-1)$–connected. As temporary notation, write $Y(\Sigma, I, J, \mathcal{P}, \mu)$ for Y. By Lemma 6.8, we have either

$$L \cong \mathcal{DTL}(\Sigma', I', J', \mathcal{P}', \mu') \quad \text{or} \quad L \cong Y(\Sigma', I', J', \mathcal{P}', \mu'),$$

depending on whether or not σ contains a vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$. Here Σ', I', J', $\mathcal{P}'$ and μ' are as follows:

- $(\Sigma', \mathcal{P}') \in \mathtt{PSurf}$ with Σ' a genus $g' = g-k-1$ surface.
- μ' is an A–homology marking on $(\Sigma', \mathcal{P}')$ supported on a symplectic subsurface.
- $I', J' \subset \partial\Sigma'$ are $\mathcal{P}'$–adjacent finite disjoint unions of open intervals satisfying $I' \cap J' = \varnothing$.

Applying either our inductive hypothesis or Claim 1, we see that L is n'–connected for

$$n' = \min\left\{n-1, \frac{g'-(2\,\mathrm{rk}(A)+3)}{\mathrm{rk}(A)+2}\right\} = \min\left\{n-1, \frac{g-(2\,\mathrm{rk}(A)+3)}{\mathrm{rk}(A)+2} - \frac{k+1}{\mathrm{rk}(A)+2}\right\}$$
$$\geq \min\left\{n-1, n-\frac{k+1}{\mathrm{rk}(A)+2}\right\} \geq n-k-1,$$

as desired. $\square$

We now finally turn to proving that $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is n–connected. Our inductive hypothesis says that it is $(n-1)$–connected, so it is enough to prove that every continuous map $f\colon S^n \to \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ can be extended to a continuous map $F\colon D^{n+1} \to \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$. Using simplicial approximation, we can assume that f is simplicial with respect to a combinatorial triangulation of S^n. Next, using Claim 2 we can ensure that f is locally injective. The complex $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is a subcomplex of Y and Claim 1 says that Y is n–connected, so we can extend f to a continuous map $F\colon D^{n+1} \to Y$, which by the relative version of simplicial approximation we can ensure is simplicial with respect to a combinatorial triangulation of D^{n+1} that restricts to our given triangulation on S^n. Finally, applying Claim 3 we can ensure that F is locally injective.

If F does not map any vertices of D^{n+1} to $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, then the image of F lies in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ and we are done. Assume, therefore, that x is a vertex of D^{n+1} such that $F(x)$ is a vertex $\iota\colon \tau(S^1) \to \Sigma$ of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$. Let $L \subset D^{n+1}$ be the link of x and let $\mathcal{L} \subset Y$ be the link of $\iota = F(x)$. Since F is locally injective, we have $F(L) \subset \mathcal{L}$. Also, since simplices of Y can contain at most one vertex of $\mathcal{TL}(\Sigma, I, \mathcal{P}, \mu)$, we have $\mathcal{L} \subset \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$.

By Lemma 6.9, we can find a vertex $\hat{\iota}\colon \tau^2(S^1) \to \Sigma$ of $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ such that $\hat{\iota}|_{\tau(S^1)} = \iota$. Let $\widehat{\mathcal{L}}$ be the link of $\hat{\iota}$ in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$, so $\widehat{\mathcal{L}} \subset \mathcal{L}$. As we said above, we have $F(L) \subset \mathcal{L}$. If $F(L) \subset \widehat{\mathcal{L}}$, then we could redefine F to take x to $\hat{\iota}$ instead of ι. Repeating this process would modify F so that its image would lie in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$, and we would be done.

Unfortunately, it might not be the case that $F(L) \subset \widehat{\mathcal{L}}$. We will therefore have to perform a more complicated modification to F. Since $L \subset D^{n+1}$ is the link of the vertex x and x does not lie in ∂D^{n+1}, we have $L \cong S^n$. Recall that

$$F(L) \subset \mathcal{L} \subset \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu).$$

Among all simplicial maps

$$G\colon L \to \mathcal{L} \subset \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$$

that are homotopic to $F|_L$ through maps $S^n \to \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$, pick the one that minimizes the total number of intersections between the image of $\hat{\iota}\colon \tau^2(S^1) \to \Sigma$ and the images of $G(y)\colon \tau^2(S^1) \to \Sigma$ as y ranges over the vertices of L.

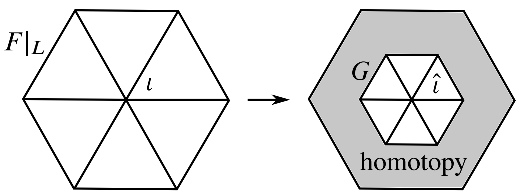
Below in Claim 4 we will prove that with this choice there are in fact no such intersections, and thus the image $G(L)$ lies in the link $\hat{\mathcal{L}}$ of $\hat{\iota}$ in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$. Letting $*$ denote the join, we can then replace the restriction of F to the subset

$$x * L \cong \mathrm{pt} * S^n \cong D^{n+1}$$

of D^{n+1} with the following two pieces:

- The first is an annular region that is a combinatorial triangulation of $S^n \times [0,1]$, both of whose boundary components are L. On this region, F maps to a homotopy from $F|_L$ to G.
- The second is the cone $x * L \cong \mathrm{pt} * S^n \cong D^{n+1}$, on which F is defined to equal G on L and to take x to $\hat{\iota}$.

See the following figure, where the shaded region is the homotopy from $F|_L$ to G:



This redefines F so that $F(x) = \hat{\iota}$ without introducing any other vertices mapping to vertices of $\mathcal{TL}(\Sigma, I, J, \mathcal{P}, \mu)$, completing the proof.

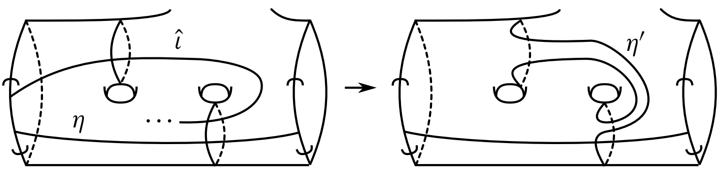
It remains to prove the aforementioned claim about $G\colon L \to \mathcal{L} \subset \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$.

Claim 4 *For all vertices y of L, we can choose a representative of $G(y)\colon \tau^2(S^1) \to \Sigma$ whose image is disjoint from the image of $\hat{\iota}\colon \tau^2(S^1) \to \Sigma$.*

Proof Assume otherwise. Since the image of G lies in the link $\mathcal{L}$ of ι, we can choose representatives of the $G(y)$ for $y \in L$ that are disjoint from the image of $\iota\colon \tau(S^1) \to \Sigma$. Pick these representatives so that their intersections with the image of $\hat{\iota}|_{[1,2]}\colon [1,2] \to \Sigma$ are transverse and all distinct. Let y be the vertex of L such that the image of $\eta := G(y)\colon \tau^2(S^1) \to \Sigma$ intersects the image of $\hat{\iota}|_{[1,2]}\colon [1,2] \to \Sigma$ in the first of these intersection points (enumerated from $\hat{\iota}(1)$ to $\hat{\iota}(2)$).

The argument is slightly different depending on whether this intersection point is contained in the image under $\eta\colon \tau^2(S^1) \to \Sigma$ of $[0,1]$, S^1 or $[1,2]$. We will give the details for when this intersection point is contained in $\eta(S^1)$; the other cases are similar.

As in the following figure, let $\eta'\colon \tau^2(S^1) \to \Sigma$ be the result of "sliding" the intersection point of η in question across $\iota(S^1)$ via the initial segment of $\iota([1,2])$:

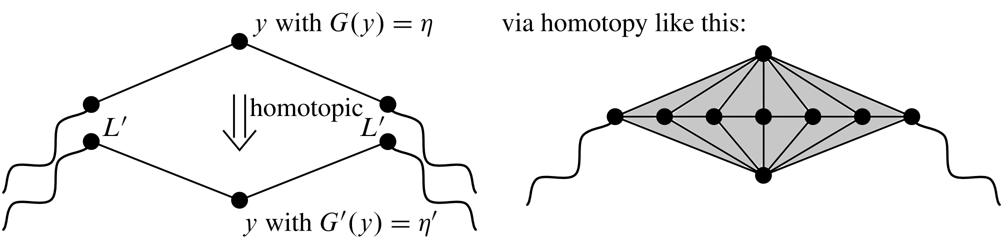


The image of η' intersects the image of $\hat{\iota}$ in one fewer place than the image of η. Define

$$G'\colon L \to \mathcal{L} \subset \mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$$

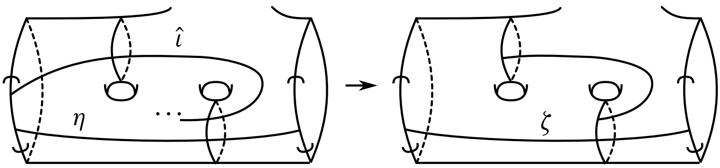
to be the map which equals G except at the vertex y, where $G'(y) = \eta'$ instead of η. It is easy to see that G' is indeed a simplicial map. Since the image of η' intersects the image of $\hat{\iota}$ in one fewer place than the image of η, to derive a contradiction to the minimality of the total number of these intersections it is enough to prove that G and G' are homotopic through maps landing in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$.

Define $L' \cong S^{n-1}$ to be the link of y in S^n, define $\mathcal{L}_\eta$ to be the link of η in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$, and define $\mathcal{L}_{\eta'}$ to be the link of η' in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$. We have $G|_{L'} = G'|_{L'}$, and the image $G(L') = G'(L')$ lies in $\mathcal{L}_\eta \cap \mathcal{L}_{\eta'}$. Below we will prove that the map $G|_{L'}\colon L' \to \mathcal{L}_\eta \cap \mathcal{L}_{\eta'}$ can be homotoped to a constant map. This will imply that G and G' are homotopic through maps lying in $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ via a homotopy like the one in this figure:



This figure depicts the case $n = 1$; pictured is a fragment of $L \cong S^1$, along with the vertex y and $L' \cong S^0$.

Since $L' \cong S^{n-1}$, to prove that the map $G|_{L'}\colon L' \to \mathcal{L}_\eta \cap \mathcal{L}_{\eta'}$ can be homotoped to a constant map, it is enough to prove that $\mathcal{L}_\eta \cap \mathcal{L}_{\eta'}$ is $(n-1)$–connected. Define ζ to be the union of $\eta(\tau^2(S^1))$, $\iota(S^1)$, and the portion of the arc of $\iota([1,2])$ connecting $\iota(0) \in \iota(S^1)$ to a point of $\eta(S^1)$; see here:



The images of both η and η' are contained in a regular neighborhood of ζ. Let Σ' be the surface obtained by cutting open Σ along ζ. The surface Σ' thus has genus $g' = g-2$. Moreover, an argument identical to that in the proof of Lemma 6.8 shows that there exist a partition $\mathcal{P}'$ of the components of $\partial\Sigma'$, an A–homology marking μ' on $(\Sigma', \mathcal{P}')$, and $\mathcal{P}'$–adjacent finite disjoint unions of open intervals $I', J' \subset \partial\Sigma'$ with $I' \cap J' = \varnothing$ such that

- $\mathcal{L}_\eta \cap \mathcal{L}_{\eta'} \cong \mathcal{MTL}(\Sigma', I', J', \mathcal{P}', \mu')$,
- μ' is supported on a symplectic subsurface.

Our inductive hypothesis thus says that $\mathcal{L}_\eta \cap \mathcal{L}_{\eta'} \cong \mathcal{MTL}(\Sigma', I', J', \mathcal{P}', \mu')$ is n'–connected for

$$n' = \min\left\{n-1, \frac{g'-(2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2}\right\} = \min\left\{n-1, \frac{g-(2\operatorname{rk}(A)+3)}{\operatorname{rk}(A)+2} - \frac{2}{\operatorname{rk}(A)+2}\right\}$$
$$\geq \min\left\{n-1, n - \frac{2}{\operatorname{rk}(A)+2}\right\} = n-1,$$

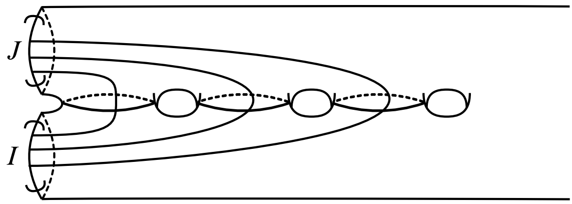
as desired. □

This completes the proof of Theorem 6.11. □

6.7 The complex of order-preserving double-tethered vanishing loops

We finally come to the complex of order-preserving double-tethered vanishing loops.

Complex of order-preserving double-tethered loops Let $\Sigma \in \texttt{Surf}$ be a surface and let $I, J \subset \partial\Sigma$ be disjoint open intervals. Orient I so that Σ lies on its right and J so that Σ lies on its left. These two orientations induce two natural orderings on simplices of $\mathcal{DTL}(\Sigma, I, J)$. The *complex of order-preserving (I, J)–double-tethered* *loops*, denoted by $\mathcal{ODTL}(\Sigma, I, J)$, is the subcomplex of $\mathcal{DTL}(\Sigma, I, J)$ consisting of simplices such that these two orderings agree. Here is an example of such a simplex:



The complex $\mathcal{ODTL}(\Sigma, I, J)$ was introduced by Hatcher and Vogtmann [12], who proved that if Σ has genus g then (like $\mathcal{TL}(\Sigma, I)$ and $\mathcal{DTL}(\Sigma, I, J)$) it is $\frac{1}{2}(g-3)$–connected; see [12, Proposition 5.3].

Complex of order-preserving double-tethered vanishing loops Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \texttt{PSurf}$ and let $I, J \subset \partial\Sigma$ be disjoint $\mathcal{P}$–adjacent open intervals in $\partial\Sigma$. Define the complex $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ to be the intersection of $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ with $\mathcal{ODTL}(\Sigma, I, J)$. The orientations on I and J endow $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ with a natural ordering on its simplices, and thus with the structure of a semisimplicial set.

High connectivity The following theorem asserts that $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ has the same connectivity that Theorem 6.11 says $\mathcal{DTL}(\Sigma, I, J, \mathcal{P}, \mu)$ enjoys.

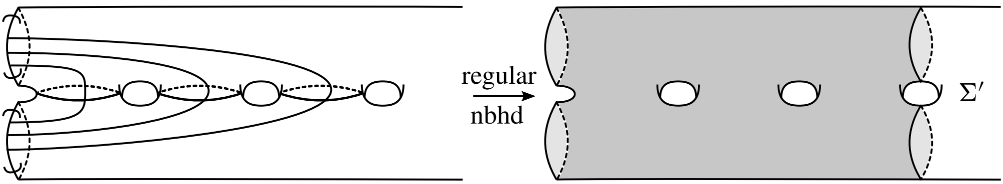
Theorem 6.13 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \texttt{PSurf}$ that is supported on a symplectic subsurface. Let $I, J \subset \partial\Sigma$ be $\mathcal{P}$–adjacent disjoint open intervals and let g be the genus of Σ. Then $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is $\bigl(g-(2\,\mathrm{rk}(A)+3)\bigr)/(\mathrm{rk}(A)+2)$–connected.*

Proof In [12, Proposition 5.3], Hatcher and Vogtmann show how to derive the fact that $\mathcal{ODTL}(\Sigma, I, J)$ is $\frac{1}{2}(g-3)$–connected from the fact that $\mathcal{DTL}(\Sigma, I, J)$ is $\frac{1}{2}(g-3)$–connected. Their argument works word-for-word to prove this theorem. $\square$

Stabilizers In the remainder of this section, we will be interested in the case where I and J are open intervals in distinct components ∂_I and ∂_J of $\partial\Sigma$ (much of what we say will also hold if $\partial_I = \partial_J$, but the pictures would be a bit different). The $\mathrm{Mod}(\Sigma)$–stabilizer of a simplex $\sigma = \{\iota_0, \ldots, \iota_k\}$ of $\mathcal{ODTL}(\Sigma, I, J)$ is the mapping class group of the complement Σ' of an open regular neighborhood of

$$\partial_I \cup \partial_J \cup \iota_0(\tau^2(S^1)) \cup \cdots \cup \iota_k(\tau^2(S^1)).$$

We will call this the *stabilizer subsurface* of σ. See here:



If ∂_I and ∂_J are $\mathcal{P}$–adjacent, then the surface Σ' is a $\mathcal{P}$–simple subsurface of Σ, and thus has an induced partition $\mathcal{P}'$. The following lemma records some of its properties if σ is a simplex of $\mathcal{ODTL}(\Sigma, I, \mathcal{P}, \mu)$ for an A–homology marking μ on $(\Sigma, \mathcal{P})$:

Lemma 6.14 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P})$ and let I and J be open intervals in distinct $\mathcal{P}$–adjacent components of $\partial\Sigma$. Let σ be a simplex of $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$, let Σ' be its stabilizer subsurface, and let $\mathcal{P}'$ be the induced partition of $\partial\Sigma'$. Then there exists an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ such that μ is the stabilization of μ'. Moreover, if μ is supported on a symplectic subsurface then so is μ'.*

Proof The proof is identical to that of Lemma 6.8. □

Transitivity The final fact we need about these complexes is as follows:

Lemma 6.15 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in$ `PSurf` that is supported on a symplectic subsurface and let I and J be open intervals in distinct $\mathcal{P}$–adjacent components of $\partial\Sigma$. The group $\mathcal{I}(\Sigma, \mathcal{P}, \mu)$ acts transitively on the k–simplices of $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ if the genus of Σ is at least $2\,\mathrm{rk}(A) + 3 + k$.*

Proof Just like in the proof of Lemma 3.9, this will be by induction on k. In fact, once we prove the base case $k = 0$, the inductive step is handled exactly like Lemma 3.9, so we will only give the details for $k = 0$.

Assume that the genus of Σ is at least $2\,\mathrm{rk}(A) + 3$. Theorem 6.13 then implies that $\mathcal{ODTL}(\Sigma, I, J, \mathcal{P}, \mu)$ is connected, so to prove that $\mathcal{I}(\Sigma, \mathcal{P}, \mu)$ acts transitively on its vertices it is enough to prove that if $\iota_0, \iota_1 \colon \tau^2(S^1) \to \Sigma$ are vertices that are joined by an edge, then there exists some $f \in \mathcal{I}(\Sigma, \mathcal{P}, \mu)$ such that $f(\iota_0) = \iota_1$. Let Σ' be the stabilizer subsurface of $\{\iota_0, \iota_1\}$ and let $\mathcal{P}'$ be the induced partition of $\partial\Sigma'$. By Lemma 6.14, there exists an A–homology marking μ' on $(\Sigma', \mathcal{P}')$ that is supported on a symplectic subsurface such that μ is the stabilization of μ' to $(\Sigma, \mathcal{P})$. Let $S \cong \Sigma_h^1$ be a subsurface of Σ' on which μ' is supported.

The change of coordinates principle from [8, Section 1.3.2] implies that there is a mapping class f' on $\Sigma' \setminus \mathrm{Int}(S)$ with $f'(\iota_0) = \iota_1$. Let $f \in \mathrm{Mod}(\Sigma)$ be the result of extending f' over S by the identity. Since μ is supported on S, we have $f \in \mathcal{I}(\Sigma, \mathcal{P}, \mu)$ and $f(\iota_0) = \iota_1$, as desired. $\square$

6.8 The double boundary stabilization proof

We now prove Proposition 5.9.

Proof of Proposition 5.9 We start by recalling the statement and introducing some notation. Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a double boundary stabilization and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Setting

$$c = \mathrm{rk}(A) + 2 \quad \text{and} \quad d = 2\,\mathrm{rk}(A) + 2,$$

we want to prove that the induced map $\mathrm{H}_k(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_k(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is an isomorphism if the genus of Σ is at least $ck + d$ and a surjection if the genus of Σ is $ck + d - 1$. We will prove this using Theorem 3.1. This requires fitting $\mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathcal{I}(\Sigma', \mathcal{P}', \mu')$ into an increasing sequence of group $\{G_n\}$ and constructing appropriate simplicial complexes.

As notation, let $(S_g, \mathcal{P}_g) = (\Sigma, \mathcal{P})$, $\mu_g = \mu$, $(S_{g+1}, \mathcal{P}_{g+1}) = (\Sigma', \mathcal{P}')$ and $\mu_{g+1} = \mu'$. In a double boundary stabilization like $(S_g, \mathcal{P}_g) \to (S_{g+1}, \mathcal{P}_{g+1})$, two boundary components of Σ_0^4 are glued to two boundary components of S_g to form S_{g+1}. We will call the two boundary components of S_g to which Σ_0^4 is glued the *attaching components* and the two components of $\partial\Sigma_0^4 \cap \partial S_{g+1}$ the *new components*.

By assumption, μ_g is supported on a genus-h symplectic subsurface for some h, ie there exists a $\mathtt{PSurf}$–morphism $(T, \mathcal{P}_T) \to (S_g, \mathcal{P}_g)$ with $T \cong \Sigma_h^1$ and an A–homology marking μ_T on $(T, \mathcal{P}_T)$ such that μ_g is the stabilization of μ_T to $(S_g, \mathcal{P}_g)$. Applying Corollary 3.6 to μ_T, we can assume without loss of generality that $h \le \mathrm{rk}(A)$. We can then factor $(T, \mathcal{P}_T) \to (S_g, \mathcal{P}_g)$ into an increasing sequence of subsurfaces

$$(T, \mathcal{P}_T) \to (S_h, \mathcal{P}_h) \to (S_{h+1}, \mathcal{P}_{h+1}) \to \cdots \to (S_g, \mathcal{P}_g)$$

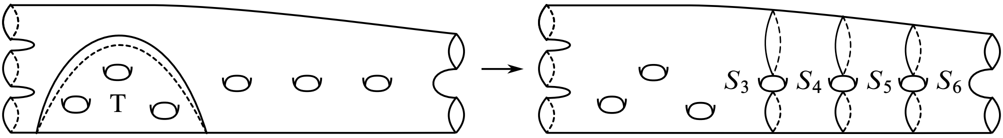
such that

(i) each S_r has genus r,

(ii) each $(S_r, \mathcal{P}_r) \to (S_{r+1}, \mathcal{P}_{r+1})$ is a double boundary stabilization, and

(iii) for $r > h$, the attaching components of $(S_r, \mathcal{P}_r) \to (S_{r+1}, \mathcal{P}_{r+1})$ equal the new components of $(S_{r-1}, \mathcal{P}_{r-1}) \to (S_r, \mathcal{P}_r)$.

See here:



This can then be continued indefinitely to form an increasing sequence of subsurfaces

$$(T, \mathcal{P}_T) \to (S_h, \mathcal{P}_h) \to \cdots \to (S_g, \mathcal{P}_g) \to (S_{g+1}, \mathcal{P}_{g+1}) \to (S_{g+2}, \mathcal{P}_{g+2}) \to \cdots$$

satisfying (i)–(iii). Here $(S_{g+1}, \mathcal{P}_{g+1})$ is as defined above. For $r \geq h$, let μ_r be the stabilization of μ_T to (S_r, μ_r). This agrees with our previous definitions of μ_g and μ_{g+1}.

We thus have an increasing sequence of groups

$$\mathcal{I}(S_h, \mathcal{P}_h, \mu_h) \subset \mathcal{I}(S_{h+1}, \mathcal{P}_{h+1}, \mu_{h+1}) \subset \mathcal{I}(S^1_{h+2}, \mathcal{P}_{h+2}, \mu_{h+2}) \subset \cdots.$$

For $r \geq h$, let $I_r, J_r \subset \partial S_r$ be open intervals in the two attaching components for $(S_r, \mathcal{P}_r) \to (S_{r+1}, \mathcal{P}_{r+1})$. According to Theorem 6.13, $\mathcal{ODTL}(S_r, I_r, J_r, \mathcal{P}_r, \mu_r)$ is $(r-(d+1))/c$–connected (where c and d are as defined in the first paragraph).

For $n \geq 0$, let

$$G_n = \mathcal{I}(S_{d+n}, \mathcal{P}_{d+n}, \mu_{d+n}) \quad \text{and} \quad X_n = \mathcal{ODTL}(S_{d+n}, I_{d+n}, J_{d+n}, \mathcal{P}_{d+n}, \mu_{d+n}).$$

For this to make sense, we must have $d+n \geq h$, which follows from

$$d+n = 2\,\mathrm{rk}(A) + 2 + n \geq \mathrm{rk}(A) \geq h.$$

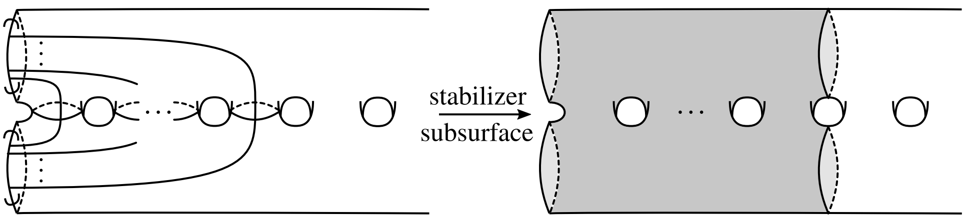
We thus have an increasing sequence of groups

$$G_0 \subset G_1 \subset G_2 \subset \cdots$$

with G_n acting on X_n. The indexing convention here is chosen so that X_1 is 0–connected, and more generally so that X_n is $(n-1)/c$–connected, as in Theorem 3.1. Our goal is to prove that the map $\mathrm{H}_k(G_{n-1}) \to \mathrm{H}_k(G_n)$ is an isomorphism for $n \geq ck+1$ and a surjection for $n = ck$, which will follow from Theorem 3.1 once we check its conditions:

(1) The first is that X_n is $(n-1)/c$–connected, which follows from Theorem 6.13.

(2) The second is that, for $0 \leq i < n$, the group G_{n-i-1} is the G_n–stabilizer of some i–simplex of X_n, which follows from Lemma 6.14 via the following picture:



(3) The third is that, for all $0 \leq i < n$, the group G_n acts transitively on the i–simplices of X_n, which follows from Lemma 6.15.

(4) The fourth is that, for all $n \geq c+1$ and all 1–simplices e of X_n whose boundary consists of vertices v and v', there exists some $\lambda \in G_n$ such that $\lambda(v) = v'$ and such that λ commutes with all elements of $(G_n)_e$. Let S' be the stabilizer subsurface of e, so by Lemma 6.14 the stabilizer $(G_n)_e$ consists of mapping classes supported on S'. The surface $S_{d+n} \setminus \mathrm{Int}(S')$ is diffeomorphic to Σ_1^4 (as in the picture above), and in particular is connected. The change of coordinates principle from [8, Section 1.3.2] implies that we can find a mapping class λ supported on $S_{d+n} \setminus \mathrm{Int}(S')$ taking the double-tethered loop v to v'. Lemma 6.14 implies that μ_{d+n} can be destabilized to an A–homology marking on S' (with respect to an appropriate partition) that is supported on a symplectic subsurface. This implies that λ lies in $G_n = \mathcal{I}(S_{d+n}, \mathcal{P}_{d+n}, \mu_{d+n})$ and commutes with $(G_n)_e$. □

7 Nonstability

This section concerns situations where homological stability does *not* occur. The highlights are the proofs of Theorems B and 5.3.

Disc-pushing subgroup Let $\Sigma \in \mathtt{Surf}$ be a surface and let ∂ be a component of $\partial\Sigma$. Let $\widehat{\Sigma}$ be the result of gluing a disc to ∂. The embedding $\Sigma \hookrightarrow \widehat{\Sigma}$ induces a homomorphism $\mathrm{Mod}(\Sigma) \to \mathrm{Mod}(\widehat{\Sigma})$, which is easily seen to be surjective. Its kernel, denoted by $\mathrm{DP}(\partial)$, is the *disc-pushing subgroup* and is isomorphic to the fundamental group of the unit tangent bundle $U\widehat{\Sigma}$ of $\widehat{\Sigma}$; see [8, Section 4.2.5]. Elements of $\mathrm{DP}(\partial)$ "push" ∂ around paths in $\widehat{\Sigma}$ while allowing it to rotate.

Disc-pushing and partial Torelli If ∂ is the single component of $\partial\Sigma_g^1$, then $\mathrm{DP}(\partial) \subset \mathrm{Mod}(\Sigma_g^1)$ is contained in the Torelli group $\mathcal{I}(\Sigma_g^1)$, and thus is also contained in $\mathcal{I}(\Sigma_g^1, \mu)$ for any A–homology marking μ on Σ_g^1. he following lemma generalizes this to the partial Torelli groups on surfaces with multiple boundary components:

Lemma 7.1 *Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \texttt{PSurf}$ and let ∂ be a component of $\partial\Sigma$ such that $\{\partial\} \in \mathcal{P}$. Then* $\mathrm{DP}(\partial) \subset \mathcal{I}(\Sigma, \mathcal{P}, \mu)$.

Proof Let $f \in \mathrm{DP}(\partial)$ and let $x \in \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$. It is enough to prove that $f(x) = x$. Let $\widehat{\Sigma}$ be the result of gluing a disc to ∂ and let $\widehat{\mathcal{P}} = \mathcal{P} \setminus \{\{\partial\}\}$. We thus have a $\texttt{PSurf}$–morphism $\iota\colon (\Sigma, \mathcal{P}) \to (\widehat{\Sigma}, \widehat{\mathcal{P}})$. Since the homology classes of arcs connecting ∂ to other components of $\partial\Sigma$ do not contribute to $\mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$, the map $\iota^*\colon \mathrm{H}_1^{\widehat{\mathcal{P}}}(\widehat{\Sigma}, \partial\widehat{\Sigma}) \to \mathrm{H}_1^{\mathcal{P}}(\Sigma, \partial\Sigma)$ is a surjection (in fact, it is an isomorphism, but we will not need this). We can thus write $x = \iota^*(\hat{x})$ for some $\hat{x} \in \mathrm{H}_1^{\widehat{\mathcal{P}}}(\widehat{\Sigma}, \partial\widehat{\Sigma})$. Since

$$f \in \mathrm{DP}(\partial) = \ker(\mathrm{Mod}(\Sigma) \xrightarrow{\iota_*} \mathrm{Mod}(\widehat{\Sigma})),$$

we clearly have $\iota_*(f)(\hat{x}) = \hat{x}$, so Lemma 4.5 implies that

$$x = \iota^*(\hat{x}) = \iota^*(\iota_*(f)(\hat{x})) = f(\iota^*(\hat{x})) = f(x),$$

as desired. □

Johnson homomorphism Fix some $g \geq 2$ and let $H = \mathrm{H}_1(\Sigma_g^1)$. The Johnson homomorphism [14] is an important homomorphism $\tau\colon \mathcal{I}(\Sigma_g^1) \to \bigwedge^3 H$. Letting ∂ be the single component of $\partial\Sigma_g^1$, it interacts with the disc-pushing subgroup $\mathrm{DP}(\partial) \cong \pi_1(U\Sigma_g)$ in the following way. Let $\omega \in \bigwedge^2 H$ be the *symplectic element*, ie the element corresponding to the algebraic intersection pairing under the isomorphism

$$\big(\textstyle\bigwedge^2 H\big)^* \cong \bigwedge^2 H^* \cong \bigwedge^2 H,$$

where we identify H with its dual H^* via Poincaré duality. We then have an injection $H \hookrightarrow \bigwedge^3 H$ taking $h \in H$ to $h \wedge \omega$. The restriction of τ to $\mathrm{DP}(\partial)$ is the composition

$$\mathrm{DP}(\partial) \cong \pi_1(U\Sigma_g) \to \pi_1(\Sigma_g) \to H \xrightarrow{-\wedge\omega} \textstyle\bigwedge^3 H.$$

Symplectic nondegeneracy Let μ be an A–homology marking on $(\Sigma, \mathcal{P}) \in \texttt{PSurf}$. The *$\mu$–symplectic element* $\omega_\mu \in \bigwedge^2 A$ is as follows. Let H be the quotient of $\mathrm{H}_1(\Sigma)$ by the subgroup generated by the loops around the boundary components. Since H is the first homology group of the closed surface obtained by gluing discs to all components of $\partial\Sigma$, there is a symplectic element $\omega \in \bigwedge^2 H$. The closed marking $\hat{\mu}\colon \mathrm{H}_1(\Sigma) \to A$ factors through a homomorphism $H \to A$, and ω_μ is the image of $\omega \in \bigwedge^2 H$ under the induced map $\bigwedge^2 H \to \bigwedge^2 A$. We then have a map $A \to \bigwedge^3 A$ taking $a \in A$ to $a \wedge \omega_\mu$. We will say that μ is *symplectically nondegenerate* if this map is nonzero.

Example 7.2 Let V be a symplectic subspace of $\mathrm{H}_1(\Sigma_g^1)$, so $\mathrm{H}_1(\Sigma_g^1) = V \oplus V^\perp$, and let $\mu\colon \mathrm{H}_1(\Sigma_g^1) \to V$ be the orthogonal projection. We claim that μ is symplectically nondegenerate if and only if V has genus at least 2. Indeed, $\omega_\mu \in \bigwedge^2 V$ equals the symplectic element arising from the symplectic form on V, and the map $V \to \bigwedge^3 V$ taking $v \in V$ to $v \wedge \omega_\mu$ is nonzero precisely when V has genus at least 2. We remark that if V has genus 0 or 1 then $\bigwedge^3 V = 0$, so the map $V \to \bigwedge^3 V$ is automatically the zero map.

Partial Johnson homomorphism The homomorphism given by the following lemma is a version of the Johnson homomorphism for the partial Torelli groups:

Lemma 7.3 *Let μ be an symplectically nondegenerate A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ and let ∂ be a component of $\partial\Sigma$ such that $\{\partial\} \in \mathcal{P}$ (and thus by Lemma 7.1 such that $\mathrm{DP}(\partial) \subset \mathcal{I}(\Sigma, \mathcal{P}, \mu)$). Then there exists a homomorphism $\tau\colon \mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathrm{H}_3(A)$ whose restriction to $\mathrm{DP}(\partial)$ is nontrivial.*

Remark 7.4 The target group $\mathrm{H}_3(A)$ contains $\bigwedge^3 A$, though sometimes it is a bit larger.

Proof of Lemma 7.3 Let Σ' be the result of gluing discs to all components of $\partial\Sigma$ except for ∂, let $\mathcal{P}' = \{\{\partial\}\}$, and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. From their definitions, the μ'–symplectic element $\omega_{\mu'} \in \bigwedge^2 A$ is the same as the μ–symplectic element $\omega_\mu \in \bigwedge^2 A$, so μ' is symplectically nondegenerate. In [3, Theorem 5.8], Broaddus, Farb and Putman construct a homomorphism

$$\tau'\colon \mathcal{I}(\Sigma', \mathcal{P}', \mu') \to \mathrm{H}_3(A).$$

We remark that their notation is a little different from ours — the group W in the statement of [3, Theorem 5.8] should be taken to be $W = \ker(\mu')$. Let $\mathrm{DP}'(\partial)$ be the disc-pushing subgroup of $\mathcal{I}(\Sigma', \mathcal{P}', \mu')$, let $\hat{\Sigma}'$ be the result of gluing a disc to the component ∂ of $\partial\Sigma'$, and let $\hat{\mu}'\colon \mathrm{H}_1(\Sigma') \to A$ be the closed marking associated to μ'. One of the characteristic properties of τ' is that its restriction to $\mathrm{DP}'(\partial)$ is

$$\mathrm{DP}'(\partial) = \pi_1(U\hat{\Sigma}') \to \pi_1(\hat{\Sigma}') \to \mathrm{H}_1(\hat{\Sigma}') = \mathrm{H}_1(\Sigma') \xrightarrow{\hat{\mu}'} A \xrightarrow{-\wedge\omega_{\mu'}} \textstyle\bigwedge^3 A \hookrightarrow \mathrm{H}_3(A).$$

In particular, since μ' is symplectically nondegenerate, the restriction of τ' to $\mathrm{DP}'(\partial)$ is nontrivial. Let $\tau\colon \mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathrm{H}_3(A)$ be the composition of τ' with the map $\mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathcal{I}(\Sigma', \mathcal{P}', \mu')$. The restriction of this latter map to $\mathrm{DP}(\partial)$ is a surjection $\mathrm{DP}(\partial) \to \mathrm{DP}'(\partial)$, so the restriction of τ to $\mathrm{DP}(\partial)$ is nontrivial, as desired. □

Closing up surfaces and nonstability In light of Example 7.2 above, the following theorem generalizes Theorem B:

Theorem 7.5 *Let μ be a symplectically nondegenerate A–homology marking on $(\Sigma, \mathcal{P})$, let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a $\mathtt{PSurf}$–morphism, and let μ' be the stabilization* *of μ to $(\Sigma', \mathcal{P}')$. Assume there is a component ∂ of $\partial\Sigma$ with $\{\partial\} \in \mathcal{P}$ whose image in Σ' bounds a disc. Then the map $\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is not injective.*

Proof Lemma 7.1 implies that $\mathrm{DP}(\partial) \subset \mathcal{I}(\Sigma, \mathcal{P}, \mu)$, and Lemma 7.3 implies that there exists a homomorphism from $\mathcal{I}(\Sigma, \mathcal{P}, \mu)$ to an abelian group whose restriction to $\mathrm{DP}(\partial)$ is nontrivial. Since

$$\mathrm{DP}(\partial) \subset \ker(\mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathcal{I}(\Sigma', \mathcal{P}', \mu')),$$

this implies that the induced map on abelianizations is not injective, as desired. □

General nonstability We now prove Theorem 5.3.

Proof of Theorem 5.3 We start by recalling what we must prove. Let μ be a symplectically nondegenerate A–homology marking on $(\Sigma, \mathcal{P}) \in \mathtt{PSurf}$ that is supported on a symplectic subsurface. Let $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ be a non-partition-bijective $\mathtt{PSurf}$–morphism and let μ' be the stabilization of μ to $(\Sigma', \mathcal{P}')$. Assume that the genus of Σ is at least $3\,\mathrm{rk}(A) + 4$. We must prove that the induced map $\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is not an isomorphism. We will ultimately prove this by reducing it to Theorem 7.5 above.

Identify Σ with its image in Σ'. We start with the following reduction. Recall that, for a surface S, the discrete partition of the components of ∂S is

$$\{\{\partial\} \mid \partial \text{ is a component of } \partial S\}.$$

Claim *We can assume without loss of generality that $\mathcal{P}$ and $\mathcal{P}'$ are the discrete partitions of the components of $\partial\Sigma$ and $\partial\Sigma'$ and that the genera of Σ and Σ' are equal.*

Proof We do this in three steps:

- First, let $(\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$ be an open capping (see Section 5.2; this implies in particular that $\mathcal{P}''$ is the discrete partition of $\partial\Sigma''$) and let μ'' be the stabilization of μ' to $(\Sigma'', \mathcal{P}'')$. Since open cappings are partition-bijective, Theorem F implies that the map $\mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu')) \to \mathrm{H}_1(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu''))$ is an isomorphism. The composition

$$(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}') \to (\Sigma'', \mathcal{P}'')$$

is still not partition-bijective, so replacing $(\Sigma', \mathcal{P}')$ and μ' with $(\Sigma'', \mathcal{P}'')$ and μ'', we can assume without loss of generality that $\mathcal{P}'$ is the discrete partition of $\partial\Sigma'$.

- Next, just like in Case 2 of the proof of Theorem F in Section 5.2, we can use the fact that μ is supported on a symplectic subsurface to find a partition-bijective $\mathtt{PSurf}$–morphism $(\Sigma''', \mathcal{P}''') \to (\Sigma, \mathcal{P})$ and an A–homology marking μ''' on $(\Sigma''', \mathcal{P}''')$

such that μ is the stabilization of μ''' to $(\Sigma, \mathcal{P})$, such that $\mathcal{P}'''$ is the discrete partition of $\partial\Sigma'''$, and such that the genera of Σ''' and Σ are the same. Theorem F implies that the map $\mathrm{H}_1(\mathcal{I}(\Sigma''', \mathcal{P}''', \mu''')) \to \mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu))$ is an isomorphism. The composition

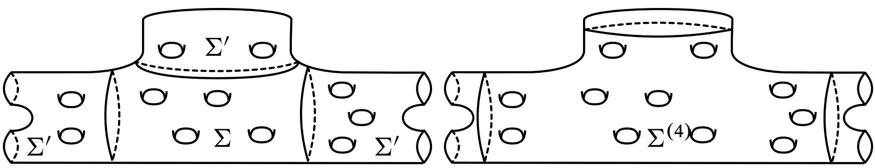
$$(\Sigma''', \mathcal{P}''') \to (\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$$

is still not partition-bijective, so replacing $(\Sigma, \mathcal{P})$ and μ with $(\Sigma''', \mathcal{P}''')$ and μ''', we can assume without loss of generality that $\mathcal{P}$ is the discrete partition of $\partial\Sigma$.

- We have now ensured that $\mathcal{P}$ and $\mathcal{P}'$ are the discrete partitions, and it remains to show that we can ensure that the genera of Σ and Σ' are the same. As in the following picture, we can factor $(\Sigma, \mathcal{P}) \to (\Sigma', \mathcal{P}')$ into

$$(\Sigma, \mathcal{P}) \to (\Sigma^{(4)}, \mathcal{P}^{(4)}) \to (\Sigma', \mathcal{P}'),$$

where $(\Sigma, \mathcal{P}) \to (\Sigma^{(4)}, \mathcal{P}^{(4)})$ is partition-bijective, where $\mathcal{P}^{(4)}$ is the discrete partition of $\partial\Sigma^{(4)}$, and where the genera of $\Sigma^{(4)}$ and Σ' are the same:



Theorem F implies that the map $\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P})) \to \mathrm{H}_1\big(\mathcal{I}(\Sigma^{(4)}, \mathcal{P}^{(4)})\big)$ is an isomorphism. Since the map $(\Sigma^{(4)}, \mathcal{P}^{(4)}) \to (\Sigma', \mathcal{P}')$ is still not partition-bijective, we can replace $(\Sigma, \mathcal{P})$ with $(\Sigma^{(4)}, \mathcal{P}^{(4)})$ and ensure that the genera of Σ and Σ' are the same. □

Since the genera of Σ and Σ' are the same, all components of $\overline{\Sigma' \setminus \Sigma}$ are genus-0 surfaces intersecting Σ in a single boundary component. If any of these components are discs, then Theorem 7.5 implies that the map $\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is not injective, and we are done. We can thus assume that no components of $\overline{\Sigma' \setminus \Sigma}$ are discs. Furthermore, if any of these components are annuli, then we can deformation retract Σ' over them without changing anything; doing this, we can assume that none of them are annuli.

It follows that all the components of $\overline{\Sigma' \setminus \Sigma}$ are genus-0 surfaces with at least three boundary components intersecting Σ in a single boundary component. Let $\{\partial_1, \dots, \partial_k\}$ be a set of components of $\partial\Sigma'$ containing precisely one component in each component of $\overline{\Sigma' \setminus \Sigma}$. Let Σ'' be the result of gluing discs to all components of Σ' except for the ∂_i, let $\mathcal{P}''$ be the discrete partition of $\partial\Sigma''$ (so in particular $\{\partial_i\} \in \mathcal{P}''$ for all i), and let μ'' be the stabilization of μ' to $(\Sigma'', \mathcal{P}'')$. All components of $\overline{\Sigma'' \setminus \Sigma}$ are annuli, so Σ'' deformation retracts to Σ'.

From this, we see that the composition

$$\mathcal{I}(\Sigma, \mathcal{P}, \mu) \to \mathcal{I}(\Sigma', \mathcal{P}', \mu') \to \mathcal{I}(\Sigma'', \mathcal{P}'', \mu'')$$

is an isomorphism, and thus the composition

$$\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu')) \to \mathrm{H}_1(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu'')) \tag{7-1}$$

is also an isomorphism. Since $\mathcal{P}'$ is the discrete partition and at least one disc was glued to a component of $\partial\Sigma'$ when we formed Σ'', Theorem 7.5 implies that the map $\mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu')) \to \mathrm{H}_1(\mathcal{I}(\Sigma'', \mathcal{P}'', \mu''))$ is not injective. Since the composition (7-1) is an isomorphism, we conclude that the map $\mathrm{H}_1(\mathcal{I}(\Sigma, \mathcal{P}, \mu)) \to \mathrm{H}_1(\mathcal{I}(\Sigma', \mathcal{P}', \mu'))$ is not surjective, and we are done. □

References

[1] **M Bestvina**, **K-U Bux**, **D Margalit**, *The dimension of the Torelli group*, J. Amer. Math. Soc. 23 (2010) 61–105 MR Zbl

[2] **S K Boldsen**, **M Hauge Dollerup**, *Towards representation stability for the second homology of the Torelli group*, Geom. Topol. 16 (2012) 1725–1765 MR Zbl

[3] **N Broaddus**, **B Farb**, **A Putman**, *Irreducible* Sp*–representations and subgroup distortion in the mapping class group*, Comment. Math. Helv. 86 (2011) 537–556 MR Zbl

[4] **T Church**, **B Farb**, *Representation theory and homological stability*, Adv. Math. 245 (2013) 250–314 MR Zbl

[5] **J F Davis**, **P Kirk**, *Lecture notes in algebraic topology*, Graduate Studies in Math. 35, Amer. Math. Soc., Providence, RI (2001) MR Zbl

[6] **N M Dunfield**, **W P Thurston**, *Finite covers of random* 3*–manifolds*, Invent. Math. 166 (2006) 457–521 MR Zbl

[7] **J S Ellenberg**, **A Venkatesh**, **C Westerland**, *Homological stability for Hurwitz spaces and the Cohen–Lenstra conjecture over function fields*, Ann. of Math. 183 (2016) 729–786 MR Zbl

[8] **B Farb**, **D Margalit**, *A primer on mapping class groups*, Princeton Math. Series 49, Princeton Univ. Press (2012) MR Zbl

[9] **G Friedman**, *Survey article: an elementary illustrated introduction to simplicial sets*, Rocky Mountain J. Math. 42 (2012) 353–423 MR Zbl

[10] **S Galatius**, **O Randal-Williams**, *Homological stability for moduli spaces of high dimensional manifolds, I*, J. Amer. Math. Soc. 31 (2018) 215–264 MR Zbl

[11] **J L Harer**, *Stability of the homology of the mapping class groups of orientable surfaces*, Ann. of Math. 121 (1985) 215–249 MR Zbl

[12] **A Hatcher**, **K Vogtmann**, *Tethers and homology stability for surfaces*, Algebr. Geom. Topol. 17 (2017) 1871–1916 MR Zbl

[13] **N V Ivanov**, *Fifteen problems about the mapping class groups*, from "Problems on mapping class groups and related topics" (B Farb, editor), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 71–80 MR Zbl

[14] **D Johnson**, *An abelian quotient of the mapping class group $\mathcal{I}_g$*, Math. Ann. 249 (1980) 225–242 MR Zbl

[15] **D Johnson**, *Conjugacy relations in subgroups of the mapping class group and a group-theoretic description of the Rochlin invariant*, Math. Ann. 249 (1980) 243–263 MR Zbl

[16] **D Johnson**, *The structure of the Torelli group, III: The abelianization of $\mathcal{I}$*, Topology 24 (1985) 127–144 MR Zbl

[17] **W van der Kallen**, *Homology stability for linear groups*, Invent. Math. 60 (1980) 269–295 MR Zbl

[18] **M Kassabov**, **A Putman**, *Equivariant group presentations and the second homology group of the Torelli group*, Math. Ann. 376 (2020) 227–241 MR Zbl

[19] **M Krannich**, *Homological stability of topological moduli spaces*, Geom. Topol. 23 (2019) 2397–2474 MR Zbl

[20] **E Looijenga**, *Connectivity of complexes of separating curves*, Groups Geom. Dyn. 7 (2013) 443–450 MR Zbl

[21] **J Miller**, **P Patzt**, **J C H Wilson**, *Central stability for the homology of congruence subgroups and the second homology of Torelli groups*, Adv. Math. 354 (2019) art. id. 106740 MR Zbl

[22] **A Putman**, *Cutting and pasting in the Torelli group*, Geom. Topol. 11 (2007) 829–865 MR Zbl

[23] **A Putman**, *The stable cohomology of the moduli space of curves with level structures*, preprint (2022) arXiv 2209.06183

[24] **A Putman**, **S V Sam**, *Representation stability and finite linear groups*, Duke Math. J. 166 (2017) 2521–2598 MR Zbl

[25] **O Randal-Williams**, **N Wahl**, *Homological stability for automorphism groups*, Adv. Math. 318 (2017) 534–626 MR Zbl

Department of Mathematics, University of Notre Dame
Notre Dame, IN, United States

andyp@nd.edu

Received: 29 May 2020 Revised: 18 April 2022

Algebraic & Geometric Topology **23**:8 (2023) 3497–3530
DOI: 10.2140/agt.2023.23.3497
Published: 5 November 2023

On symplectic fillings of small Seifert 3–manifolds

HAKHO CHOI
JONGIL PARK

We investigate the minimal symplectic fillings of small Seifert 3–manifolds with a canonical contact structure. As a result, we list all minimal symplectic fillings using curve configurations for small Seifert 3–manifolds satisfying certain conditions. Furthermore, we also demonstrate that every such a minimal symplectic filling is obtained by a sequence of rational blowdowns from the minimal resolution of the corresponding weighted homogeneous complex surface singularity.

53D05, 57R17; 32S25

1 Introduction

One of the fundamental problems in symplectic 4–manifold topology is to classify symplectic fillings of certain 3–manifolds equipped with a natural contact structure. Among them, researchers have long studied symplectic fillings of the link of a normal complex surface singularity. Note that the link of a normal surface singularity carries a canonical contact structure also known as the Milnor fillable contact structure. For example, P Lisca [8], M Bhupal and K Ono [1], and H Park, J Park, D Shin and G Urzúa [12] completely classified all minimal symplectic fillings of lens spaces and certain small Seifert 3–manifolds coming from the link of quotient surface singularities. L Starkston [16] also investigated minimal symplectic fillings of the link of some weighted homogeneous surface singularities.

On the one hand, topologists working on 4–manifold topology are also interested in finding a surgical interpretation for symplectic fillings of a given 3–manifold. More specifically, one may ask whether there is any surgical description of those fillings. In fact, it has been known that *rational blowdown* surgery, introduced by R Fintushel and R Stern [5] and generalized by the second author [14] and A Stipsicz, Z Szabó and J Wahl [18], is a powerful tool to answer this question. For example, for the link of

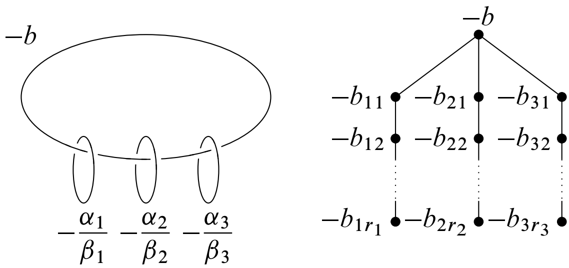


Figure 1: Surgery diagram of Y and its associated plumbing graph Γ.

quotient surface singularities equipped with a canonical contact structure, it was proven that every minimal symplectic filling is obtained by a sequence of rational blowdowns from the minimal resolution of the singularity; see Bhupal and Ozbagci [2], Choi and Park [4]. On the other hand, L Starkston [17] showed that there are symplectic fillings of some Seifert 3–manifolds that cannot be obtained by a sequence of rational blowdowns from the minimal resolution of the singularity. Note that Seifert 3–manifolds can be viewed as the link of weighted homogeneous surface singularities. Hence, it is an intriguing question as to which Seifert 3–manifolds have a rational blowdown interpretation for their minimal symplectic fillings.

In this paper, we investigate the minimal symplectic fillings of small Seifert 3–manifolds over the 2–sphere satisfying certain conditions. By a *small* Seifert (fibered) 3–manifold, we assume that it admits at most 3 singular fibers when it is considered as an S^1–fibration over the 2–sphere. In general, a Seifert 3–manifold as an S^1–fibration over a Riemann surface can have any number of singular fibers. We denote a small Seifert 3–manifold Y by $Y(-b;(\alpha_1,\beta_1),(\alpha_2,\beta_2),(\alpha_3,\beta_3))$ whose surgery diagram is given in Figure 1 and which is also given as a boundary of a plumbing 4–manifold of disk bundles of 2–spheres according to the graph Γ in Figure 1. The integers $b_{ij}\geq 2$ in Figure 1 are uniquely determined by the continued fraction

$$\frac{\alpha_i}{\beta_i}=[b_{i1},b_{i2},\dots,b_{ir_i}]=b_{i1}-\cfrac{1}{b_{i2}-\cfrac{1}{\dots-\cfrac{1}{b_{ir_i}}}}.$$

If the intersection matrix of a plumbing graph Γ is negative definite, which is always true for $b\geq 3$, then there is a canonical contact structure on Y induced from a symplectic structure of the plumbing 4–manifold, where each vertex corresponds to a symplectic 2–sphere and each edge represents an orthogonal intersection between the symplectic 2–spheres; see Gay and Stipsicz [7]. Furthermore, the canonical contact structure on Y is contactomorphic to the contact structure defined by the complex tangency of a complex structure on the link of the corresponding singularity, which is called the *Milnor fillable* contact structure; see Park and Stipsicz [13].

This paper aims to classify all possible list of minimal symplectic fillings of small Seifert 3–manifolds satisfying certain conditions, and to prove that every such a minimal symplectic filling is obtained by a sequence of rational blowdowns from the minimal resolution of the corresponding weighted homogeneous surface singularity, as it is true for a quotient surface singularity. Our strategy is as follows. For a given minimal symplectic filling W of $Y(-b;(\alpha_1,\beta_1),(\alpha_2,\beta_2),(\alpha_3,\beta_3))$ with $b \geq 4$, we glue W with a concave cap K to get a closed symplectic 4–manifold X. Then, since the concave cap K always contains an embedded $(+1)$ 2–sphere corresponding the central vertex, X is a rational symplectic 4–manifold by McDuff [9]. Furthermore, the adjunction formula and intersection data impose a constraint on the homological data of K in $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$. Under blowdowns along all exceptional 2–spheres away from the $(+1)$ 2–sphere in $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$, the concave cap K becomes a neighborhood of symplectic 2–spheres which are isotopic to b number of complex lines through symplectic 2–spheres in $\mathbb{CP}^2$; see Starkston [16; 17] for details. Since the symplectic deformation type of $W \cong X \setminus K$ is determined by the isotopy class of a symplectic embedding of K within a fixed homological embedding, we investigate a symplectic embedding of K using a *curve configuration* corresponding to W, which consists of strands representing irreducible components of K and exceptional 2–spheres intersecting them (Definition 3.1 and Figure 5). Since the curve configuration corresponding to W determines a symplectic embedding of K, we can recover all minimal symplectic fillings by investigating all possible curve configurations of Y. Sometimes, we find a certain chain of symplectic 2–spheres lying in W, which can be rationally blowing down, from the homological data of K. Note that by rationally blowing down the chain of symplectic 2–spheres lying in W, we obtain another minimal symplectic W' from W. In this case, we keep track of changes in the homological data of K so that we get a curve configuration of W' from that of W. Finally, by analyzing the effect of rational blowdown surgery on the curve configuration of minimal symplectic fillings, we obtain our main result.

Theorem 1.1 *For a small Seifert* 3*–manifold* $Y(-b;(\alpha_1,\beta_1),(\alpha_2,\beta_2),(\alpha_3,\beta_3))$ *with its canonical contact structure and* $b \geq 4$*, all minimal symplectic fillings of* Y *are listed explicitly by curve configurations. Furthermore, they are also obtained by a sequence of rational blowdowns from the minimal resolution of the corresponding weighted homogeneous surface singularity.*

Remark L Starkston [16] originally described a general scheme for how to obtain minimal symplectic fillings from given homological data of an embedding of the cap and got some results in special cases. L Starkston [17] also showed that the isotopy type of a symplectic line arrangement is uniquely (up to deformation equivalence) determined by its intersection data in the cases that multi-intersection points of a symplectic line arrangement satisfy some mild conditions, which contain the cases appearing in Proposition 3.4. Thus, by combining Propositions 3.3 and 3.4 with Starkston's result [17, Proposition 4.2], we conclude that there exists at most one minimal symplectic filling for each possible curve configuration. Then we prove in Section 4 that every such a curve configuration gives the corresponding minimal symplectic filling, which implies the first statement in Theorem 1.1 above.

Acknowledgements The authors would like to thank the referees for their valuable comments. Jongil Park was supported by Samsung Science and Technology Foundation under Project Number SSTF-BA1602-02 and by the National Research Foundation of Korea (NRF) Grant 2020R1A5A1016126 funded by the Korea government. He also holds a joint appointment in the Research Institute of Mathematics, SNU.

2 Preliminaries

2.1 Weighted homogeneous surface singularities and Seifert 3–manifolds

We briefly recall some basics of weighted homogeneous surface singularities and Seifert 3–manifolds; see [10] for details. Suppose that $(w_0, \dots, w_n)$ are nonzero rational numbers. A polynomial $f(z_0, \dots, z_n)$ is called *weighted homogeneous* of type $(w_0, \dots, w_n)$ if it can be expressed as a linear combination of monomials $z_0^{i_0} \cdots z_n^{i_n}$ for which

$$\frac{i_0}{w_0} + \frac{i_1}{w_1} + \cdots + \frac{i_n}{w_n} = 1.$$

Equivalently, there exist nonzero integers $(q_0, \dots, q_n)$ and a positive integer d satisfying $f(t^{q_0} z_0, \dots t^{q_n} z_n) = t^d f(z_0, \dots, z_n)$. Then, a weighted homogeneous surface singularity $(X, 0)$ is a normal surface singularity that is defined as the zero loci of weighted homogeneous polynomials of the same type. Note that there is a natural $\mathbb{C}^*$–action given by

$$t \cdot (z_0, \dots, z_n) = (t^{q_0} z_0, \dots, t^{q_n} z_n),$$

with a single fixed point $0 \in X$. This $\mathbb{C}^*$–action induces a fixed-point-free $S^1 \subset \mathbb{C}^*$ action on the link $L := X \cap \partial B$ of the singularity, where B is a small ball centered at the origin. Hence, the link L is a Seifert fibered 3–manifold over a genus g Riemann surface, denoted by $Y(-b; g; (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_k, \beta_k))$ for some integers b, α_i and β_i with $0 < \beta_i < \alpha_i$ and $(\alpha_i, \beta_i) = 1$. Note that k is the number of singular fibers, and there is an associated star-shaped plumbing graph Γ: the central vertex has genus g and weight $-b$, and each vertex in k arms has genus 0 and weight $-b_{ij}$ uniquely determined by the continued fraction

$$\frac{\alpha_i}{\beta_i} = [b_{i1}, b_{i2}, \dots, b_{ir_i}] = b_{i1} - \cfrac{1}{b_{i2} - \cfrac{1}{\cdots - \cfrac{1}{b_{ir_i}}}}$$

with $b_{ij} \geq 2$. For example, Figure 1 shows the case of $g = 0$ and $k = 3$, which is called a *small* Seifert (fibered) 3–manifold. By P Orlik and P Wagreich [11], it is well known that the plumbing graph Γ is a dual graph of the minimal resolution of $(X, 0)$. Conversely, if the intersection matrix of Γ is negative definite, there is a weighted homogeneous surface singularity whose dual graph of the minimal resolution is Γ; see [15]. Note that a Seifert 3–manifold Y, as a boundary of a plumbed 4–manifold according to Γ, inherits a canonical contact structure providing that each vertex represents a symplectic 2–sphere, all intersections between them are orthogonal, and the intersection matrix of Γ is negative definite; see [7]. Furthermore, if the Seifert 3–manifold Y can be viewed as the link L of a weighted homogeneous surface singularity, then the canonical contact structure above is contactomorphic to the *Milnor fillable* contact structure, which is given by $TL \cap JTL$; see [13].

2.2 Rational blowdowns and symplectic fillings

Rational blowdown surgery, first introduced by R Fintushel and R Stern [5], is one of the most powerful cut-and-paste techniques. It replaces a certain linear plumbing C_p of disk bundles over a 2–sphere whose boundary is a lens space $L(p^2, p-1)$ with a rational homology 4–ball B_p which has the same boundary. Later, Fintushel and Stern's rational blowdown surgery was generalized by J Park [14] using a configuration $C_{p,q}$ obtained by linear plumbing disk bundles over a 2–sphere according to the dual

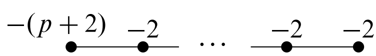


Figure 2: Linear plumbing C_p.

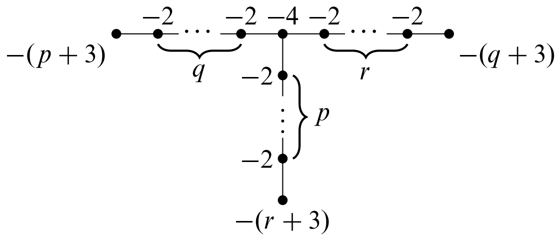


Figure 3: Plumbing graph $\Gamma_{p,q,r}$.

resolution graph of $L(p^2, pq-1)$, which also bounds a rational homology 4–ball $B_{p,q}$. In the case of a symplectic 4–manifold (X, ω), rational blowdown surgery can be performed in the symplectic category: if all 2–spheres in the plumbing graph are symplectically embedded and their intersections are ω–orthogonal, then the surgered 4–manifold $X_{p,q} = (X - C_{p,q}) \cup B_{p,q}$ also admits a symplectic structure induced from the symplectic structure of X; see [19; 20]. In fact, the rational homology 4–ball $B_{p,q}$ admits a symplectic structure compatible with the canonical contact structure on the boundary $L(p^2, pq-1)$. More generally, in addition to the linear plumbing of 2–spheres, there is a plumbing of 2–spheres according to star-shaped plumbing graphs with 3 or 4 legs admitting a symplectic rational homology 4–ball; see [18; 3]. That is, the corresponding Seifert 3–manifold $Y(-b, (\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3))$, or $Y(-b, (\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3), (\alpha_4, \beta_4))$ with a canonical contact structure, has a minimal symplectic filling whose rational homology is isomorphic to that of the 4–ball [6]. For example, a plumbing graph $\Gamma_{p,q,r}$ in Figure 3 can be rationally blown down. We will use this later in the proof of the main theorem.

As rational blowdown surgery does not affect the symplectic structure near the boundary, if there is a plumbing of disk bundles over symplectically embedded 2–spheres that can be rationally blown down, then one can obtain another symplectic filling by replacing the plumbing with a rational homology 4–ball. In the case of the link of quotient surface singularities, it was proven [2; 4] that every minimal symplectic filling is obtained by a sequence of rational blowdowns from the minimal resolution of the singularity, which is diffeomorphic to a plumbing of disk bundles over symplectically embedded 2–spheres. First, they constructed a genus-0 or genus-1 Lefschetz fibration X on each minimal symplectic filling of the link of a quotient surface singularity. Suppose that w_1 and w_2 are two words consisting of right-handed Dehn twists along curves in a generic fiber, which represent the same element in the mapping class group of the generic fiber. If the monodromy factorization of X is given by $w_1 \cdot w'$, one can construct another Lefschetz fibration X' whose monodromy factorization is given by $w_2 \cdot w'$. The operation of replacing w_1 with w_2 is called a *monodromy substitution*. Next, they showed that the monodromy factorization of each minimal symplectic filling of the link of a quotient surface singularity is obtained by a sequence of monodromy substitutions from that of the minimal resolution. Furthermore, these monodromy substitutions can be interpreted as rational blowdown surgeries topologically. Note that all rational blowdown surgeries mentioned here are linear: a certain linear chain $C_{p,q}$ of 2–spheres is replaced with a rational homology 4–ball.

2.3 Minimal symplectic fillings of a small Seifert 3–manifold

In this subsection, we briefly review Starkston's results [16; 17] for minimal symplectic fillings of a small Seifert fibered 3–manifold $Y(-b; (\alpha_1, \beta_1), (\alpha_2, \beta_2), (\alpha_3, \beta_3))$ with $b \geq 4$. The condition $b \geq 4$ on the weight (equivalently, degree) of a central vertex of the plumbing graph Γ ensures that one can always choose a concave cap K, which is also star-shaped, with a $(+1)$ central 2–sphere and $b-4$ arms, each of which consists of a single (-1) 2–sphere as in Figure 4. Here $[a_{i1}, a_{i2}, \dots, a_{in_i}]$ denotes a dual continued fraction of $[b_{i1}, b_{i2}, \dots, b_{ir_i}]$; that is, $\alpha_i/(\alpha_i - \beta_i) = [a_{i1}, a_{i2}, \dots, a_{in_i}]$ while $\alpha_i/\beta_i = [b_{i1}, b_{i2}, \dots, b_{ir_i}]$.

For a given minimal symplectic filling W of Y, we glue W and K along Y to get a closed symplectic 4–manifold X. Then, the existence of a $(+1)$ 2–sphere implies that X is a rational symplectic 4–manifold and, after a finite number of blowdowns, X becomes $\mathbb{CP}^2$ and the $(+1)$ 2–sphere in K becomes a complex line $\mathbb{CP}^1 \subset \mathbb{CP}^2$; see McDuff [9] for details. Under these circumstances, it is natural to ask the following question: What is the image of K in $\mathbb{CP}^2$ under blowdowns? In the case that K is linear, which means that the corresponding Y is a lens space, Lisca showed that the image of K is two symplectic 2–spheres in $\mathbb{CP}^2$, each of which is homologous to $\mathbb{CP}^1 \subset \mathbb{CP}^2$.

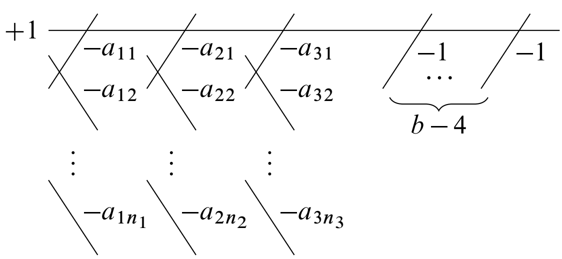


Figure 4: Concave cap K.

By analyzing the proof of Lisca's result [8, Theorem 4.2], Starkston [16] showed that the image of K is b symplectic 2–spheres in $\mathbb{CP}^2$, each of which is homologous to $\mathbb{CP}^1 \subset \mathbb{CP}^2$. For the complete classification of minimal symplectic fillings of Y, one needs to classify the isotopy classes of these b symplectic 2–spheres, which are called *symplectic line arrangements*. Since all these spheres are J–holomorphic for some J tamed by the standard Kähler form of $\mathbb{CP}^2$ and are homologous to $\mathbb{CP}^1 \subset \mathbb{CP}^2$, they intersect each other at a single point for each pair of 2–spheres. Note that these intersection points need not be all distinct. These intersection data of a symplectic line arrangement are determined by the homological data of K, which also have constraints from the adjunction formula. In [17], Starkston showed that symplectic line arrangements with certain types of intersections are isotopic to *complex line arrangements*, that is, the corresponding b symplectic 2–spheres are isotopic (through symplectic spheres) to b complex lines in $\mathbb{CP}^2$. For example, Starkston classified minimal symplectic fillings by an explicit computation of all possible homological embeddings of K for some families of Seifert fibered spaces; see [16, Sections 3 and 4.4; 17, Section 5].

3 Strategy for the main theorem

As we saw in the previous section, for each minimal symplectic filling W of Y, we obtain a rational symplectic 4–manifold X which is symplectomorphic to $\mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$ for some integer N by gluing K to W along Y. Conversely, given an embedding of a concave cap K into $\mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$, we obtain a symplectic filling W of Y by taking a complement of K in $\mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$. So the classification of minimal symplectic fillings of Y is equivalent to the classification of the embeddings of K into $\mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$ for some N. Hence, in order to investigate minimal symplectic fillings W of Y, we first introduce two notions, *homological data* and *curve configuration* of the corresponding embedding of K, which are defined as follows.

Definition 3.1 Suppose that W is a minimal symplectic filling of a small Seifert 3–manifold Y equipped with a concave cap K. Then we have an embedding of K into a rational symplectic 4–manifold $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$ such that the $(+1)$ 2–sphere in K is identified with $\mathbb{CP}^1 \subset \mathbb{CP}^2$. Let l be a homology class represented by a complex line $\mathbb{CP}^1$ in $\mathbb{CP}^2$ and e_i be homology classes of exceptional spheres coming from blowups. Then $\{l, e_1, \dots, e_N\}$ becomes a basis for $H_2(X;\mathbb{Z})$, so that the homology class of each irreducible component of K can be expressed in terms of this basis, which we call the *homological data of K* for W.

Note that K is symplectically embedded in $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$ and each irreducible component of K can be assumed to be J–holomorphic for some J tamed by standard Kähler form on X. Then there is a sequence of blowdowns from X to $\mathbb{CP}^2$ and we can find a J–holomorphic exceptional sphere Σ_i whose homology class is e_i disjoint from the central $(+1)$ 2–sphere of K at each stage of blowing down. Because of the J–holomorphic condition and homological restrictions from the adjunction formula together with intersection data of K, the exceptional sphere Σ_i intersects positively at most once with the image of an irreducible component of K or is one of the image of irreducible components of K; see [8, Proposition 4.4]. In particular, for each image C_j of irreducible components of K, the intersection number between e_i and $[C_j]$ lies in $\{-1, 0, 1\}$.

As mentioned in the previous section, we finally get a symplectic line arrangement in $\mathbb{CP}^2$ which consists of J–holomorphic 2–spheres, each of which is the image of the first component of each arm under blowdowns. The intersection data of the symplectic line arrangement are determined by the homological data of K, so that it can be represented as a configuration of strands: each strand represents a J–holomorphic 2–sphere of a symplectic line arrangement in $\mathbb{CP}^2$, while the intersection of two strands represents a geometric intersection of two 2–spheres. Then, starting from the configuration of the symplectic line arrangement, we can draw a configuration C of strands with degrees by blowups according to the homological data of K until we get K in the configuration. Here the degree of each strand in C means a self-intersection number of the strand. To be more precise, when we blow up a point p on a strand in a configuration, we introduce a new strand with degree -1 to the point p so that we resolve intersection of strands at p and we decrease the degree of the strands containing p by one. Hence the configuration C, which represents the total transform of a symplectic line arrangement, contains strands representing irreducible components of K and exceptional (-1) 2–spheres intersecting with the irreducible components. We say that two configurations C_1 and C_2 for W are *equivalent* if there is a bijective map between (-1) strands preserving intersections with the irreducible components of K.

Definition 3.2 If there are no strands with degree less than or equal to -2 in C except for irreducible components of K, we call the configuration C the *curve configuration* of a minimal symplectic filling W.

Remark A curve configuration C of W consists of strands representing irreducible components of K and exceptional 2–spheres intersecting the irreducible components

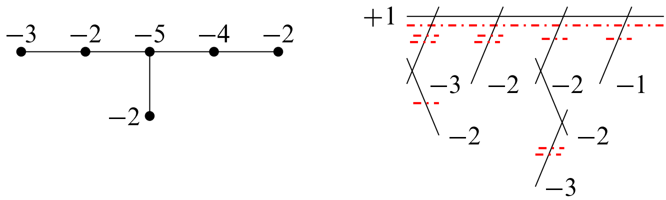


Figure 5: Plumbing graph Γ and curve configuration for corresponding concave cap K.

of K. We denote the exceptional 2–spheres by dash-dotted strands. See Figure 5 for example.

Remark We often use the terminology *configuration of strands* when we deal with an intermediate configuration between a symplectic line arrangement and a curve configuration, or a configuration containing K but with strands with degree less than or equal to -2 other than irreducible components of K.

Proposition 3.3 *For given homological data of K for W, there is a unique curve configuration C up to equivalence.*

Proof Since each strand in a curve configuration C represents a J–holomorphic 2–sphere for some J tamed by standard Kähler form on $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$, all intersections between the strands represent positive geometric intersections between the corresponding J–holomorphic 2–spheres. Note that there is at most one intersection point between any two strands due to homological restrictions. Furthermore, if e_i is a homology class of an exceptional 2–sphere satisfying $e_i \cdot [C_j] \in \{0, 1\}$ for any irreducible component C_j of K, then there is a (-1) strand L_i in C whose homology class is e_i; otherwise, there is a blowup on the strand L_i so that the proper transform of L_i becomes an irreducible component C_j of K whose intersection with e_i is -1, contradicting the assumption. Hence there is a (-1) strand L_i representing a J–holomorphic exceptional sphere Σ_i whose homology class is e_i in C if and only if $e_i \cdot [C_j] \in \{0, 1\}$ for any irreducible component C_j of K.

Let C and C' be two curve configurations for a fixed homological data of K for W. Then, the numbers of (-1) strands in C and C' are equal to the number of e_i satisfying the condition $e_i \cdot [C_j] \in \{0, 1\}$ for any irreducible component C_j of K. Hence we can construct a desired bijection between the (-1) strands by finding correspondence between such e_i and (-1) strands in two curve configurations. □

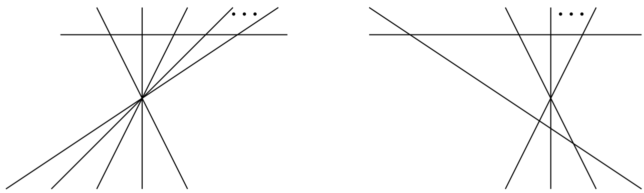


Figure 6: Symplectic line arrangements.

Now, we investigate minimal symplectic fillings of a given small Seifert 3–manifold Y by analyzing all the possible curve configurations. For this, we first determine all possible symplectic line arrangements.

Proposition 3.4 *For minimal symplectic fillings of a small Seifert fibered 3–manifold $Y(-b;(\alpha_1,\beta_1),(\alpha_2,\beta_2),(\alpha_3,\beta_3))$ with $b \geq 4$, there are only two possible intersection relations of symplectic line arrangements, which can be drawn as in Figure 6.*

Proof Since Y is a small Seifert 3–manifolds with $b \geq 4$, we can always choose a concave cap K with a $(+1)$ central 2–sphere and $b-4$ arms, each of which consists of a single (-1) 2–sphere as in Figure 4. Furthermore, since the blowdowns are disjoint from the central 2–sphere in K, each of $b-1$ arms in K descends to a single $(+1)$ J–holomorphic 2–sphere intersecting at a distinct point with an image of the central 2–sphere of K under the blowdowns. Let $C_1, C_2, \dots, C_{b-4}$ be the images of $b-4$ many (-1) 2–spheres in K under the blowdowns. Then they should have a common intersection point in $\mathbb{CP}^2$: otherwise, we have distinct two points p and q on some C_i such that C_i intersects C_j and C_k at p and q, respectively. Let r be an intersection point of C_j and C_k. Then any J–holomorphic 2–sphere coming from an arm of K other than $C_1, \dots C_{b-4}$ must pass two of p, q and r, which is a contradiction.

If $b \geq 6$, a similar argument shows that there is at most one J–holomorphic 2–sphere coming from an arm of K intersecting at a different point from the common intersection point p with C_i, which proves the proposition.

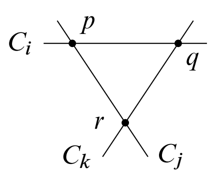


Figure 7: Configuration for C_i, C_j and C_k.

In the case of $b \le 5$, we can easily check that Figure 6 gives all possible symplectic line arrangements: if $b = 5$, then there is only one C_1 coming from (-1) 2–sphere from K. Recall that there are at most two intersection points on C_1. If there is only one intersection point on C_1, then we get the left-hand figure in Figure 6. If there are two intersection points p and q on C_1, then two of three J–holomorphic 2–spheres coming from the arms of K other than C_1 pass p, and the other passes q (or vice versa), so that we get the right-hand figure in Figure 6. For $b = 4$ case, we have only three strands in a figure for a symplectic line arrangement except the strand from $(+1)$ 2–sphere so that we have only two possibilities. $\square$

Next, for the complete classification of minimal symplectic fillings of Y, we need to consider the isotopy classes of embeddings of K with a fixed homological data in $X \cong \mathbb{CP}^2 \sharp N\overline{\mathbb{CP}^2}$. By blowing down J–holomorphic 2–spheres, it descends to isotopic types of corresponding symplectic line arrangement in $\mathbb{CP}^2$. By [17, Propositions 4.1 and 4.2], two symplectic line arrangements in Figure 6 are actually isotopic to complex line arrangements through symplectic configurations, which means that there is a unique minimal symplectic filling up to symplectic deformation equivalence for each possible choice of homological data of K. Since a choice of homological data of K gives a unique curve configuration C up to equivalence by Proposition 3.3, we analyze minimal symplectic fillings of a small Seifert 3–manifold Y by considering all possible curve configurations obtained from the complex line arrangements in Figure 6.

As previously mentioned, in the case of quotient surface singularities that include all lens spaces and some small Seifert 3–manifolds, every minimal symplectic filling is obtained by linear rational blowdown surgeries from the minimal resolution of the corresponding singularity. However, this is not true anymore for small Seifert 3–manifolds in general. For example, a rational homology 4–ball of $\Gamma_{p,q,r}$ in Figure 3 might not be obtained by linear rational blowdown surgeries. Nevertheless, many cases such as $b \ge 5$ are in fact obtained by linear rational blowdowns from their minimal resolutions. For the case of $b = 4$, one might need 3–legged rational blowdown surgeries to get a minimal symplectic filling. Hence, it is natural to prove the two cases $b \ge 5$ and $b = 4$ separately.

3.1 The case $b \ge 5$

We consider all possible curve configurations coming from two complex line arrangements in Figure 6 which can be divided into three types. First, we need to blow up all intersection points in the line arrangements so that we get two configurations as

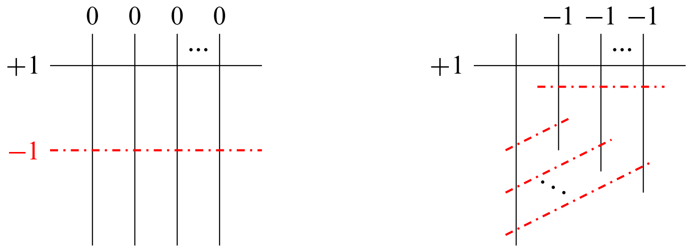


Figure 8: Blowups of the line arrangements.

in Figure 8. There are two possibilities for a strand representing exceptional sphere in intermediate configurations coming from blowups: to blow up some intersection points, or not. Once we blow up an intersection point on a strand representing an exceptional sphere Σ, which means the proper transform of Σ becomes an irreducible component of K, we should blow up all the intersection points except one intersection point because each strand intersecting the strand for Σ become irreducible components of distinct arms in K. We can also blow up the last intersection point we did not blow up to get another curve configuration, but it is not necessary in general.

If we do blow up an intersection point on the dash-dotted strand of the left-hand side of Figure 8, we get the configuration on the left-hand side of Figure 9. When we start with two configurations in Figure 9, we can assume without loss of generality that the first three arms become *essential arms* in K, which consist of strands with degree less than or equal to -2. Since the degree of the other arms is already -1, we can only blow up e_1 and e_2 among dotted exceptional strands. In conclusion, we can divide all the possible curve configurations into following three types:

- **Type A** Curve configurations obtained from Figure 8, left, without blowing up the exceptional strand.

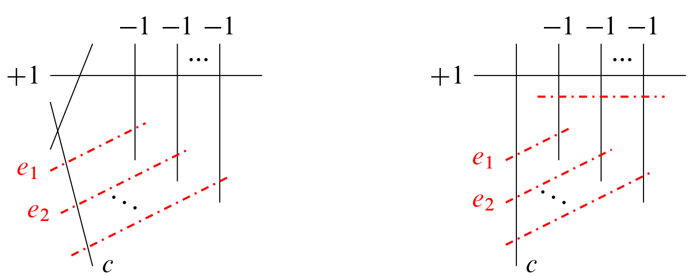


Figure 9: Two configurations.

- **Type B** Curve configurations obtained from Figure 9, left or right, by blowing up at most one e_i, with $1 \leq i \leq 2$.
- **Type C** Curve configurations obtained from Figure 9, left or right, by blowing up both e_1 and e_2.

3.2 The case $b = 4$

We divide all curve configurations for $b = 4$ into the following two cases:

- Curve configurations of types A, B or, C as in the $b \geq 5$ case.
- **Type D** Curve configurations obtained from Figure 9, right, by blowing up all exceptional (-1) strands.

Then, since we can deal with the first case using the same argument in the $b \geq 5$ case, it suffices to prove the type D case whose corresponding curve configurations come from some configurations $C_{p,q,r}$ in Figure 22, which are obtained from the right-hand figure in Figure 9; see Section 4.4 for details. The main difference between the $b = 4$ case and the $b \geq 5$ case is that one can use all three exceptional 2–spheres to get a concave cap K for $b = 4$, while one can use only e_1 and e_2 for $b \geq 5$ from the right-hand figure in Figure 9.

4 Proof of main theorem

In this section, for a given possible curve configuration C, we show that there is a sequence of rational blowdowns from the minimal resolution $\widetilde{M}$ to the minimal symplectic filling W of Y corresponding to C. Since any minimal symplectic filling of a lens space is obtained by a sequence of rational blowdowns from a linear plumbing which is the minimal resolution corresponding to the lens space [2], it suffices to construct a sequence of curve configurations $C = C_0, C_1, \dots, C_n$ such that each minimal symplectic filling W_i corresponding to C_i is obtained from W_{i+1} by replacing a certain linear plumbing L_i with its minimal symplectic filling. Here C_n denotes a curve configuration for the minimal resolution $\widetilde{M}$. As previously mentioned, since our possible symplectic line arrangements are isotopic to complex line arrangements, it suffices to work in complex category with a symplectic form ω coming from the standard Kähler form on $\mathbb{CP}^2$. In order to show that there is a symplectic embedding of L_i in W_{i+1}, we construct a configuration C'_{i+1} of strands, which is not a curve configuration for W_{i+1}, from a complex line arrangement by blowups with the same homological data of K for W_{i+1} so that we have L_i disjoint from K in C'_{i+1}. Since we work in complex category, each strand in C'_{i+1} can be considered as a complex rational curve in a rational surface X while the intersections between strands represent positive geometric intersections between the corresponding rational curves. This observation implies that L_i is symplectically embedded in W_{i+1}.

Now we introduce the notion of *standard blowups*, which frequently appears in the construction of W_i from W_{i+1}. Let K and K' be two star-shaped plumbing graphs having the same number of arms together with a $(+1)$ central vertex, and let $-a_{ij}$ for $1 \le j \le n_i$ and $-a'_{ij}$ for $1 \le j \le n'_i$ be the weights (equivalently, degrees) of the j^{th} vertex in the i^{th} arm of K and K', respectively. We say $K' \le K$ if $n'_i \le n_i$ and $a'_{ij} \le a_{ij}$ for any i and j except for $a'_{in'_i} < a_{in'_i}$ in the case of $n'_i < n_i$. The condition $K' \le K$ guarantees that we can obtain a configuration of strands representing K by blowups from a configuration representing K' in the following way: we blow up nonintersection points of the last component of each i^{th} arm in K' consecutively until we get n_i components, and then we blow up each component at nonintersection points to get the right weights.

Definition 4.1 Let C' be a configuration of strands obtained from a complex line arrangement by blowups containing a star-shaped plumbing graph K' with a homological data. If $K' \le K$ and the degree of all strands in $C' \setminus K'$ is -1, then we can obtain a curve configuration $\widetilde{C}'$ from C' by blowing up at nonintersection points only. In this case, we say that the curve configuration $\widetilde{C}'$ is obtained by *standard blowups* from C'.

Remark With given homological data of K' in C', the standard blowups induce a unique homological data of K for $\widetilde{C}'$: Let e be a homology class of an exceptional sphere coming from blowing-ups from C' to $\widetilde{C}'$. Since we blow up nonintersection points, e appears in at most two $[C^{i_1}_{j_1}]$ and $[C^{i_2}_{j_2}]$, where C^i_j denotes the j^{th} component in the i^{th} arm of K. Moreover, if e appears in two $[C^{i_1}_{j_1}]$ and $[C^{i_2}_{j_2}]$, then $i_1 = i_2 = i$ and $j_2 = j_1 + 1$ with $e \cdot [C^i_{j_1}] = 1$ and $e \cdot [C^i_{j_1+1}] = -1$.

For a given star-shaped plumbing graph $K' \le K$, in general if $n'_i < n_i$ for some i, where n'_i and n_i are the number of components in i^{th} arm of K' and K respectively, there are possibly other ways of blowing up to get the i^{th} arm of K from that of K'. Let C' be a configuration of strands containing $K' \le K$ as in Definition 4.1. Assume furthermore that $n'_i < n_i$ for some i. Let $\widetilde{C}'$ be a curve configuration obtained from C' by standard blowups. Then we get the following three fundamental lemmas.

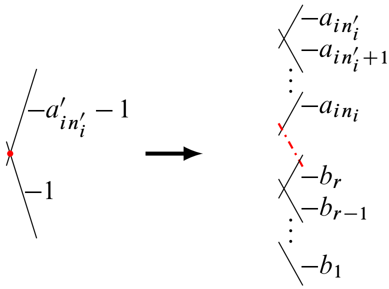


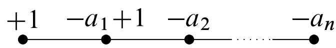
Figure 10: Finding an embedding of L.

Lemma 4.2 *Let C be a curve configuration for K, and let W be the minimal symplectic filling of Y corresponding to C. Suppose C' is a configuration for $K' \leq K$ such that the standard blowups $\widetilde{C}'$ of C' differs from C only in the components C^i_j for $n'_i \leq j \leq n_i$. Let $\widetilde{W}$ denote the minimal symplectic filling of Y corresponding to $\widetilde{C}'$. Then there is a symplectically embedded linear plumbing L of 2–spheres determined by $[b_1, b_2, \ldots, b_r]$ in $\widetilde{W}$ such that W is obtained by $\widetilde{W}$ by replacing the plumbing L with some minimal filling W_L of the lens space boundary of the linear plumbing L. Further, $[b_1, b_2, \ldots, b_r]$ is the dual of $[(a_{in'_i} - a'_{in'_i}), a_{in'_i+1}, a_{in'_i+2}, \ldots, a_{in_i}]$, where $-a_{ij}$ and $-a'_{ij}$ are the weights of the j^{th} component in the i^{th} arm of K and K', respectively.*

Proof We can assume that $a_{in'_i} - a'_{in'_i} \geq 2$ because the way of blowing up from the i^{th} arm of K' to that of K remains the same when we replace K' with K'', where K'' is obtained from K' by blowing up the last component of the i^{th} arm.

First we show that there is a symplectic linear embedding L in $\widetilde{W}$. Let S be a configuration of strands containing K obtained as follows. We blow up the last component in the i^{th} arm of K' in C' at a nonintersection point so that we have two consecutive strands of degree $-a'_{in'_i} - 1$ and -1. Since the continued fraction $[b_1, b_2, \ldots, b_r]$ is dual to $[(a_{in'_i} - a'_{in'_i}), a_{in'_i+1}, a_{in'_i+2}, \ldots a_{in_i}]$ by the definition of L, we obtain a linear chain of strands containing the rest of the i^{th} arm in K and L from the two strands by blowing up consecutively at intersection points as in Figure 10, so that there is an embedding L in the complement of K in a rational surface X. Furthermore, since we started from the same homological data of K' in C' and since a blowup for C' to S either increases the number of components or decreases the degree of an irreducible component of K, the homological data of K for both $\widetilde{C}'$ and S are the same, so that there is a symplectic embedding L in $\widetilde{W}$.

Before we examine the effect of replacing L with its minimal symplectic filling W_L, we briefly review the classification of minimal symplectic fillings of lens space, which can be found in [1] and [8]. For notational convenience, we denote a linear plumbing graph and a lens space determined the plumbing graph by the same L. For a lens space L given by $[b_1, b_2, \dots, b_r]$, we can choose a concave cap K_L of the form



where $[a_1, a_2, \dots, a_n]$ is a dual continued fraction of $[b_1, b_2, \dots, b_r]$. Suppose $X_L \cong \mathbb{CP}^2 \sharp N_0 \overline{\mathbb{CP}^2}$ is a rational symplectic 4–manifold obtained by gluing two plumbings according to L and K_L whose second homology class is generated by $\{l\} \cup E = \{E_1, \dots, E_{N_0}\}$. Then, for a given minimal symplectic filling W_L of L, we get a rational symplectic 4–manifold $X_{W_L} \cong \mathbb{CP}^2 \sharp N \overline{\mathbb{CP}^2}$ by gluing W_L and K_L and the image of K_L under blowing down is isotopic to two complex lines in $\mathbb{CP}^2$, which means that a minimal symplectic filling of L is determined by a choice of homological data of K_L in $\mathbb{CP}^2 \sharp N \overline{\mathbb{CP}^2}$ for some N. Hence, we draw a curve configuration C_{W_L} for W_L starting from a configuration of two $(+1)$ strands in $\mathbb{CP}^2$ by blowing-ups with only one $(+1)$ strand. This observation shows that the effect of replacing L in X_L with W_L is the following: We have another rational symplectic 4–manifold $X_{W_L} \cong \mathbb{CP}^2 \sharp N \overline{\mathbb{CP}^2}$ and the second homology classes in the complement of L are changed so that

$$l \to l \qquad \text{and} \qquad [L_i]^E \to [L_i]^e \qquad \text{for } 1 \le i \le n,$$

where $[L_i]^E$ and $[L_i]^e$ are homology classes of irreducible components of K_L in terms of $\{l\} \cup E = \{E_1, \dots, E_{N_0}\}$ and $\{l\} \cup e = \{e_1, \dots, e_N\}$ respectively.

Let $[C_j^i]^C$ and $[C_j^i]^{C'}$ be homology classes of C_j^i in C and C' respectively. Note that C is a curve configuration completed from the last $(-a'_{in'_i})$ strand in the i^{th} arm of K' by blowups without using any other strand in C'. If we blow up in the same ways starting with a single $(+1)$ strand instead of $(-a'_{in'_i})$ strand, we get a curve configuration C_{W_L} containing K_L. Hence there is a minimal symplectic filling W_L of L whose homological data of K_L in $X_{W_L} (= W_L \cup K_L) \cong \mathbb{CP}^2 \sharp N \overline{\mathbb{CP}^2}$ are given by $[L_j] = [C^i_{n'_i+j-1}]^C$ except for $[L_0] = l$ and $[L_1] = l + [C^i_{n'_i}]^C - [C^i_{n'_i}]^{C'}$, where $e = \{e_1, \dots, e_N\}$ is homology classes of exceptional spheres coming from the blowups from C' to C.

Finally, we show that the minimal symplectic filling W corresponding to C is given by $(\widetilde{W} \setminus L) \cup W_L$. Suppose X' is a rational symplectic 4–manifold obtained by blowups from a complex line arrangement so that it contains C'. We take a small Darboux neighborhood B' of a disk D in $C^i_{n'_i}$ of K' so that B' is disjoint from any other irreducible components of K'. Now we arrange all the blowups from C' to C inside B' and let B be blowups of B'. Then we have a symplectic embedding of K in $X = (X' \setminus B') \cup B$ and homological data of K that agrees with C. Furthermore, $B \setminus K$ is symplectic deformation equivalent to W_L: consider two complex lines in $\mathbb{CP}^2$ and a symplectic embedding of B' such that the image of D in $C^i_{n'_i}$ is a disk in one complex line and B' is disjoint from the other complex line. By the construction of B, there is a symplectic embedding of K_L in $(\mathbb{CP}^2 \setminus B') \cup B$, where the first component of K_L is the complex line in $(\mathbb{CP}^2 \setminus B')$ and the complement of K_L in $(\mathbb{CP}^2 \setminus B') \cup B$ is symplectic deformation equivalent to W_L. Since the complement of a neighborhood of $\mathbb{CP}^1$ in $\mathbb{CP}^2$ is a ball, $B \setminus K = B \setminus K_L$ is also symplectic deformation equivalent to W_L. Note that $K = (K \cap (X' \setminus B')) \cup (K \cap B) = (K' \setminus B') \cup (K \cap B)$. Hence

$$W = X \setminus K = ((X' \setminus B') \setminus K) \cup (B \setminus K) \cong (X' \setminus (K' \cup B')) \cup W_L.$$

By a similar argument, $\widetilde{W} \cong (X' \setminus (K' \cup B')) \cup L$, so that W is obtained from $\widetilde{W}$ by replacing L by W_L. □

Assume furthermore that there is a (-1) curve intersecting both $C^i_{n'_i}$ and another irreducible component C^k_l of K' in C'. Then there is a slight modification of the Lemma 4.2, involving two arms of K.

Lemma 4.3 *Suppose that there is a (-1) curve E intersecting $C^i_{n'_i}$ and C^k_l of K' in C' with $a'_{kl} < a_{kl}$. If the standard blowup $\widetilde{C}'$ of C' differs from C only in C^k_l and components C^i_j for $n'_i \le j \le n_i$, then there is a symplectically embedded linear plumbing $L \subset W$, described in Figure 11, such that W is obtained by $\widetilde{W}$ by replacing the plumbing L with some minimal filling W_L. Furthermore, $[b_1, b_2, \dots, b_r]$ is the dual of $[(a_{in'_i} - a'_{in'_i}) + 1, a_{in'_i+1}, a_{in'_i+2}, \dots, a_{in_i}]$, where $-a_{ij}$ and $-a'_{ij}$ are the weights of the j^{th} component in the i^{th} arm of K and K', respectively.*

Proof The proof is similar to that of Lemma 4.2 except for blowups at two intersection points of E in C' to find an embedding L. That is, we construct a configuration S of strands containing K, as in Figure 12, whose homological data is equal to that of $\widetilde{C}'$, so that there is a symplectic embedding of L in $\widetilde{W}$.

Next, by viewing L as a two-legged plumbing graph with a degree $(-b_1 - 1)$ of a central vertex, we get a concave cap K_L as in Figure 11: starting from the zero section

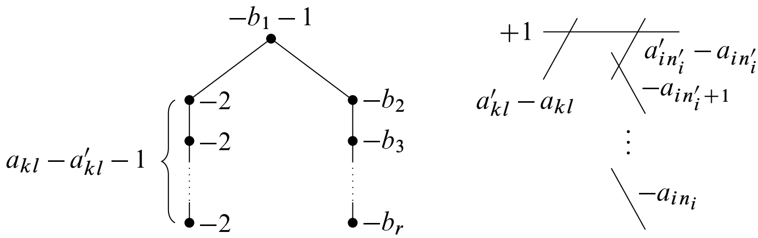


Figure 11: A plumbing graph of L and its concave cap K_L.

and infinity section together with two generic fibers of $\mathbb{F}_{b_1-1}$, we construct arms corresponding to $[-2,\dots,-2]$ and $[-b_2,\dots,-b_r]$. Then, by consecutive blowups at intersection points of the proper transform of zero section and the arm corresponding to $[-b_2,\dots,-b_r]$, we get a concave cap K_L for L. As before, for a given minimal symplectic filling of L, we get a rational symplectic 4–manifold by gluing K_L along L and the image of K_L in $\mathbb{CP}^2$ under blowing down is three complex lines in $\mathbb{CP}^2$ intersecting generically, implying that any curve configuration for K_L is obtained from blowing up at an intersection point between two complex lines in $\mathbb{CP}^2$. Therefore, using blowup data from C' to C (Figure 13), we get a minimal symplectic filling W_L of L.

Suppose that X' is a rational symplectic 4–manifold containing C'', obtained from C' by blowing down E, and let B' be a small Darboux neighborhood of the intersection point coming from the blowing down. Then a similar argument as in Lemma 4.2 above shows that the minimal symplectic filling W corresponding to C is obtained from $\widetilde{W}$ by replacing L with W_L. □

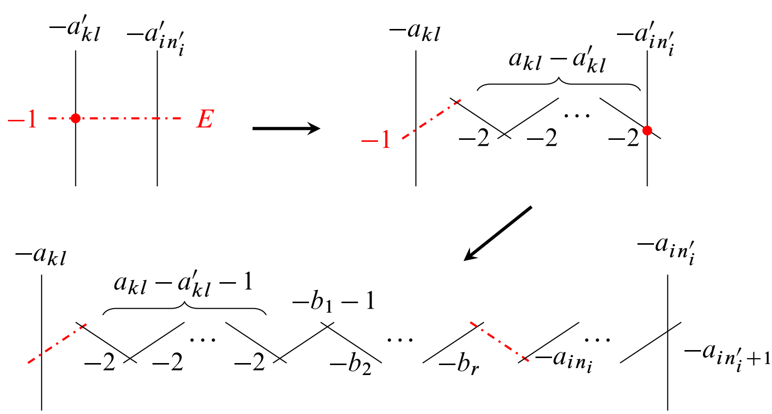


Figure 12: Embedding of L to $\widetilde{W}$.

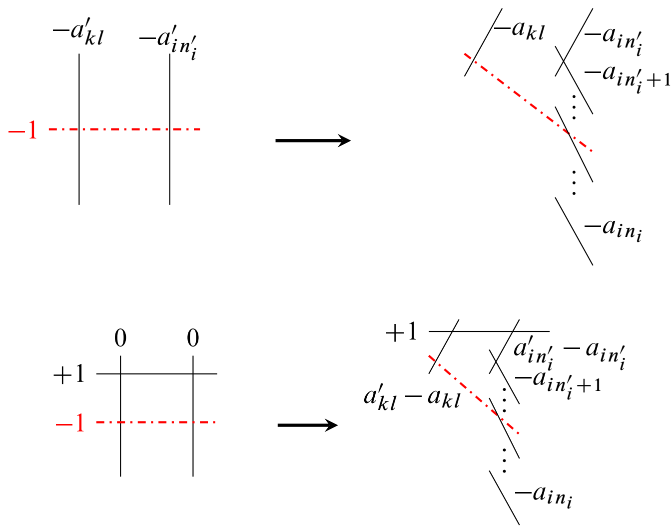


Figure 13: Top: blowups from C' to C. Bottom: curve configuration for W_L.

Assume that C' is a curve configuration containing $K' \leq K$ corresponding to a minimal symplectic filling W' of another small Seifert 3–manifold Y' and $\widetilde{C}'$ is a curve configuration obtained from C' by standard blowups. Then we can describe a minimal symplectic filling $\widetilde{W}$ of Y corresponding to $\widetilde{C}'$ explicitly.

Lemma 4.4 *Under the assumption above, there is a symplectically embedded plumbing of* 2*–spheres* Γ' *in the minimal resolution* $\widetilde{M}$ *such that a minimal symplectic filling* $\widetilde{W}$ *of* Y *corresponding to* $\widetilde{C}'$ *is obtained from* $\widetilde{M}$ *by replacing* Γ' *with* W'*.*

Proof Let K_0 be a plumbing graph determined by black strands in the left-hand side of Figure 8. Clearly, $K_0 \leq K$ so that there is a curve configuration $C_{\widetilde{M}}$ obtained by standard blowups from Figure 8, left. We first show that the curve configuration $C_{\widetilde{M}}$ corresponds to the minimal resolution $\widetilde{M}$. Recall that a concave cap K in Figure 4 can be found in [18; 16]: Starting from the zero and infinity sections with $b-1$ generic fibers of a Hirzebruch surface $\mathbb{F}_1$ which can be drawn as the left-hand side in Figure 8, we blow up intersection points of generic fibers and the infinity section so that we have a $(-b)$ rational curve which corresponds to the central vertex of the minimal resolution graph Γ. Then, we obtain a linear chain of strands containing both i^{th} arm of K and Γ from two (-1) strands by blowups, as in Figure 14, bottom. As a result, we have a configuration $S_{\widetilde{M}}$ containing both Γ and K disjointly, so that the complement of K in a rational surface X_Y is the minimal resolution $\widetilde{M}$ and K is a concave cap for Y. By

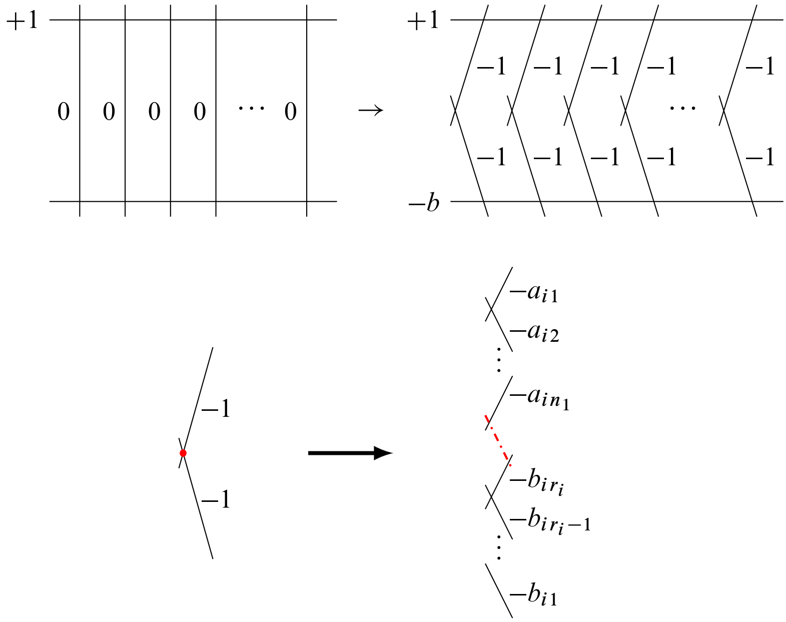


Figure 14: Top: blowing up Hirzebruch surface $\mathbb{F}_1$. Bottom: construction of each arm in K and Γ.

using the same argument as in the proof of Lemma 4.2 above, we conclude that $C_{\widetilde{M}}$ is a curve configuration for $\widetilde{M}$.

In the same way, we could get a configuration $S_{\Gamma'}$ of strands containing both K' and a plumbing graph Γ' so that the complement of K' in the resulting rational symplectic 4–manifold $X_{Y'} \cong \mathbb{CP}^2 \sharp M\overline{\mathbb{CP}^2}$ is a plumbing of 2–spheres according to Γ'. Note that $M-1$ is the number of blowups in the standard blowups from the left-hand side of Figure 8 to K'. Since $K' \leq K$, we obtain a configuration $S'_{\widetilde{M}}$ of strands containing Γ' and K disjointly from $S_{\Gamma'}$ by standard blowups at nonintersection points in the last component of each i^{th} arm of K'. Let $X = X_{Y'} \sharp N\overline{\mathbb{CP}^2}$ be a resulting rational symplectic 4–manifold. Then $X \cong X_Y = \widetilde{M} \cup K$, since the number of blowups in the standard blowups from the left-hand side of Figure 8 to K is equal to the sum of numbers of blowups for the left-hand side of Figure 8 to K', and K' to K. Furthermore, the homological data of K in $S'_{\widetilde{M}}$ is also equal to that of $C_{\widetilde{M}}$. Hence a plumbing graph Γ' is symplectically embedded in $\widetilde{M}$.

If there is a sequence of blowups from a configuration of strands representing K' to K, then we have a corresponding symplectic cobordism Z from Y' to Y because the total transform of K' is still a concave cap for Y' while K is a concave cap for Y.

In particular, if C' is a curve configuration for a minimal symplectic filling W', then we get a curve configuration C for a minimal filling W by the sequence of blowups from K' to K and $W = W' \cup Z$. In the case of standard blowups from K' to K, we can deduce from the construction of $S'_{\widetilde{M}}$ that the corresponding cobordism is equal to $\widetilde{M} \setminus \Gamma'$. Hence we have $\widetilde{W} = W' \cup (\widetilde{M} \setminus \Gamma')$, so that $\widetilde{W}$ is obtained from $\widetilde{M}$ by replacing Γ' with W'. □

4.1 Proof for type A

For a curve configuration C of type A, we want to show that the corresponding minimal symplectic filling W is obtained from the minimal resolution $\widetilde{M}$ by replacing each arm in the resolution graph Γ with its minimal symplectic filling. Since we already know in the proof of Lemma 4.4 above that a curve configuration $C_{\widetilde{M}}$, which is obtained from the left-hand side of Figure 8 by standard blowups, corresponds to $\widetilde{M}$ by repeatedly applying Lemma 4.2 with K' as in the left-hand side of Figure 8 so that the corresponding L is one of three arms in Γ, we conclude that all minimal symplectic fillings corresponding to a curve configuration C of type A are obtained by a sequence of rational blowdowns from the minimal resolution $\widetilde{M}$.

The following example illustrates this case.

Example 4.1 Let Y be a small Seifert 3–manifold whose associated plumbing graph and concave cap are shown in Figure 15, top. Then, there are two curve configurations of type A as in Figure 15, bottom. Of course, there exist other curve configurations of type B and C for minimal symplectic fillings of Y, which will be treated in Example 4.2 and Example 4.3 later. Note that each dash-dotted strand represents an exceptional 2–sphere, that is, a 2–sphere with self-intersection -1. We omit the degree of irreducible components of the concave cap for the sake of convenience in the figure. The bottom left curve configuration in Figure 15 is obtained by standard blowups from that of Figure 8, which means that the corresponding minimal filling is the minimal resolution. Note that only the third arm in the plumbing graph Γ has a nontrivial minimal symplectic filling that is obtained by rationally blowing down the (-4) 2–sphere. Using Lisca's description of the minimal symplectic fillings of lens spaces, we obtain the bottom right curve configuration in Figure 15, which represents a minimal symplectic filling obtained from the minimal resolution by rationally blowing down the (-4) 2–sphere in the third arm.

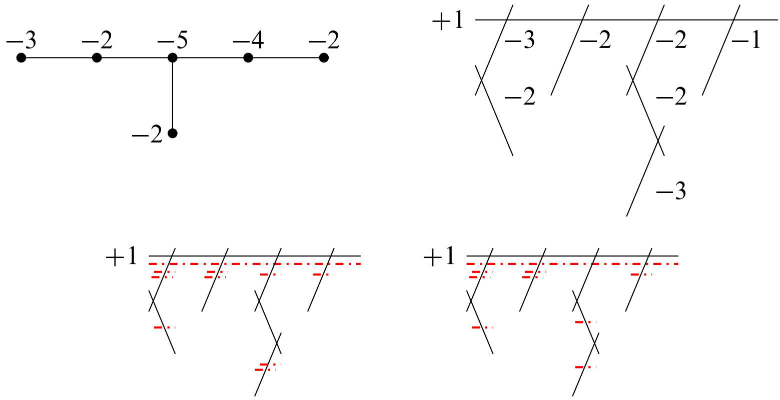


Figure 15: Top: plumbing graph Γ, and its concave cap K. Bottom: two curve configurations in Example 4.1.

4.2 Proof for type B

For a curve configuration C of type B, we want to show that the corresponding minimal symplectic filling W is obtained from the minimal resolution $\widetilde{M}$ by replacing disjoint subgraphs in the resolution graph Γ with their minimal symplectic filling. By reindexing if needed, we assume that the first and the second arms of the configurations in Figure 9 become the first and the second arm of K in C, respectively, and the proper transform of e_2 is not an irreducible component of K. Since we do not use e_2 during the blowups, we can get the first and the second arm of K, so that the homological data for the irreducible components in these arms agrees with that of C, from the configurations in Figure 9 leaving the third single (-1) arm unchanged. Hence we arrange the order of blowups from a configuration in Figure 9 to C so that we have an intermediate configuration C' of strands containing $K' \le K$ as in Figure 16. Note that the degree of strands in $C' \setminus K'$ is all -1. If we choose a linear plumbing graph

$$L' = \overset{-b_{1r_1}}{\bullet}\!-\cdots-\overset{-b_{11}}{\bullet}-\overset{-b}{\bullet}-\overset{-b_{21}}{\bullet}-\cdots-\overset{-b_{2r_2}}{\bullet}$$

to be a subgraph of Γ as a two-legged plumbing graph with the $(-b)$ central vertex, then K' gives a concave cap of L' and C' is a curve configuration for a minimal symplectic filling $W_{L'}$ of L'.

Let C_1 be a curve configuration obtained by standard blowups from C'. Then, by Lemma 4.4, the curve configuration C_1 corresponds to a minimal symplectic filling W_1, which is obtained from the minimal resolution $\widetilde{M}$ by replacing L' with $W_{L'}$. Furthermore, since $[a_{31}, a_{32}, \dots, a_{3n_3}] = [2, \dots, 2, c_1 + 1, c_2, \dots, c_k]$, where $[c_1, c_2, \dots, c_k]$

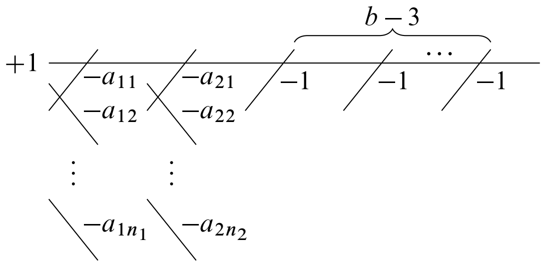


Figure 16: Concave cap K' for linear subgraph of Γ.

is the dual of $[b_{32}, b_{33}, \dots, b_{3r_3}]$, by Lemma 4.2 with L as a linear chain determined by $[b_{32}, b_{33}, \dots, b_{3r_3}]$, we conclude that the minimal symplectic filling W corresponding to C is obtained from W_1 by replacing L with its minimal symplectic filling. Hence the desired minimal symplectic filling W is obtained from $\widetilde{M}$ by replacing disjoint linear subgraphs

$$-b_{1r_1} \cdots -b_{11} \;\; -b \;\; -b_{21} \cdots -b_{2r_2} \qquad \text{and} \qquad -b_{32} \;\; -b_{33} \cdots -b_{3r_3}$$

of Γ with their minimal symplectic fillings, so that there is a sequence of rational blowdowns from $\widetilde{M}$ to W.

The following example illustrates the curve configurations of type B.

Example 4.2 We again consider a small Seifert 3–manifold Y used in Example 4.1. Since the left-hand configuration without exceptional 2–spheres in Figure 17 gives a concave cap of a lens space determined by a subgraph

$$-3 \;\; -2 \;\; -5 \;\; -2$$

of Γ, it gives a minimal symplectic filling W_L of the lens space $L(39, 16)$. Then, by blowups at points lying on the third arm different from the intersection point with the exceptional curve e, we get an embedding of a concave cap K of Y as in the right-hand curve configuration C_1 of Figure 17, which gives a minimal symplectic filling W_1 of Y. Furthermore, since there is a unique minimal symplectic filling of lens space $L(2, 1)$ corresponding to the (-2) 2–sphere in the third arm of Γ, W_1 is obtained from the minimal symplectic filling W_L. In fact, there are three more minimal symplectic fillings of Y which are of type B — see Figure 17 for the corresponding curve configurations. Note that the curve configuration C_1 for W_1 in Figure 17 comes from the right-hand configuration in Figure 9 and the curve c becomes a component of the first arm of K in

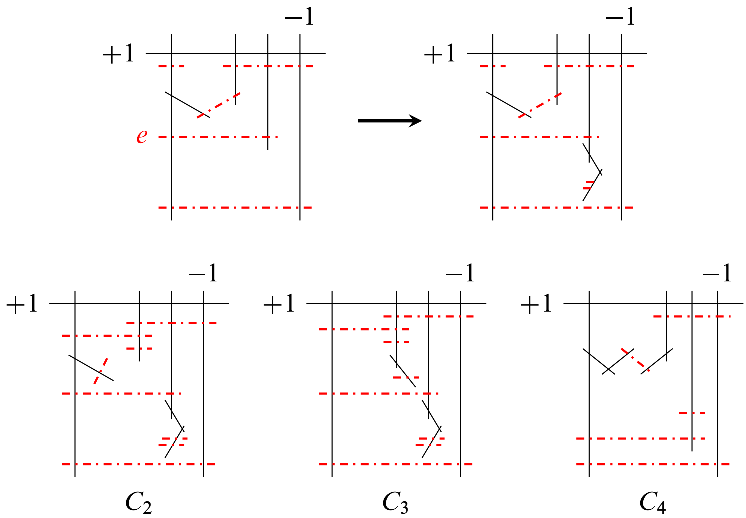


Figure 17: Top: curve configuration C_1 for W_1. Bottom: curve configurations for other symplectic fillings of Y.

the top right of Figure 15. Similarly, the curve configuration C_i for W_i $(2 \leq i \leq 4)$ is also obtained from the right-hand configuration in Figure 9. One can easily check that each W_i is obtained from the minimal resolution of Y by a linear rational blowdown surgery: explicitly, W_2, W_3 and W_4 are obtained by rationally blowing down along subgraphs

$$\overset{-2}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-5}{\bullet}\,, \qquad \overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet} \qquad \text{and} \qquad \overset{-3}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-4}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}$$

in Γ, respectively. And W_1 is also obtained by rationally blowing down along

$$\overset{-3}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}\,,$$

which is embedded in another plumbing

$$\overset{-3}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}\,.$$

4.3 Proof for type C

For a minimal symplectic filling W corresponding to a curve configuration C of type C, we want to find a curve configuration C_1 of type B such that there is a symplectically embedded linear chain L of 2–spheres (that is not visible in Γ) in W_1 corresponding to C_1 so that W is obtained from W_1 by replacing L with its minimal symplectic filling W_L.

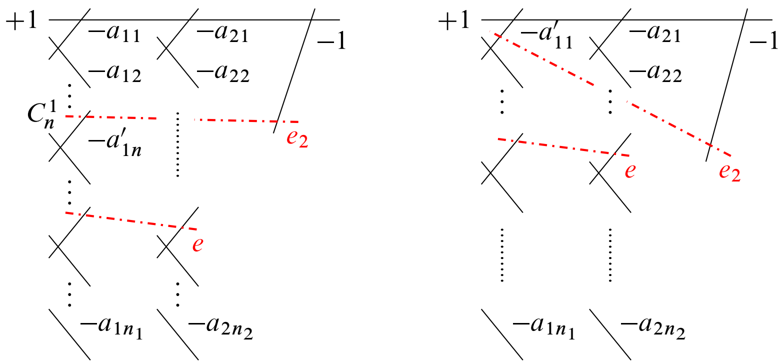


Figure 18: Part of intermediate configuration C'.

By reindexing if needed, we may assume that the first and the second arm of configurations in Figure 9 become that of K respectively, and the proper transform of e_2 becomes an irreducible component in the third arm of K after blowups. For convenience, we omit in figures all exceptional (-1) strands that intersect only one irreducible component of the corresponding concave cap K.

Now, by blowups at intersection points of e_1 consecutively, we get the first and the second arm so that the homological data for the irreducible components in these arms agrees with that of C except for one irreducible component, say C_n^1, of the first arm of K leaving the third single (-1) arm unchanged. Note that there is only one exceptional strand e connecting the first and the second arm as in Figure 18 because we blow up at intersection points of e_1 to get the first and the second arm of K. Hence we can arrange a sequence of blowups from a configuration in Figure 9 to a curve configuration C of type C so that we have an intermediate configuration C' of strands as in Figure 18: The left-hand/right-hand figures comes from the left-hand/right-hand figures in Figure 9, respectively. For simplicity, we only explain a curve configuration coming from the left-hand side in Figure 9. In contrast to the type B case, we have a $(-a'_{1n})$ strand with $a_{1n} > a'_{1n}$ in C' because we need to blow up at the intersection point of e_2 and c in Figure 9, which becomes the $(-a_{1n})$ strand in the curve configuration C at the top of Figure 19. Let C_1 be a curve configuration obtained from C' by standard blowups and W_1 be a minimal symplectic filling of Y corresponding to C_1. Then, by Lemma 4.3, there is a symplectic embedding L in W_1 so that W is obtained from W_1 by replacing L with its minimal symplectic filling W_L, where L is a plumbing graph at the bottom of Figure 19. Since a curve configuration C_1 for W_1 is of type B, there is a sequence of rational blowdowns from $\widetilde{M}$ to W as desired.

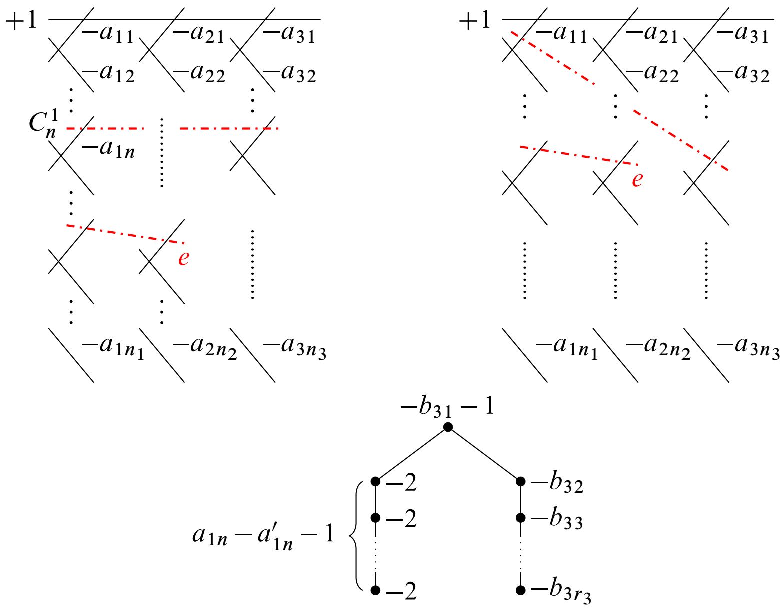


Figure 19: Top: part of curve configuration C for W. Bottom: a plumbing graph of L.

The following example illustrates this case.

Example 4.3 We consider a minimal symplectic filling W_5 of Y in Example 4.1, represented by a curve configuration C_5 in Figure 20. The curve configuration C_5 is obtained from the right-hand configuration in Figure 9, and the proper transforms of e_1 and e_2 are irreducible components of the concave cap K. Thus, as in the proof, we can find an intermediate configuration C' between the right-hand configuration in Figure 9 and C_5. Then it is easy to check that the homological data for standard blowups $\widetilde{C}'$ of C' and that of C_1 is equal; see Figures 17 and 20. From the proof of Lemma 4.3, we

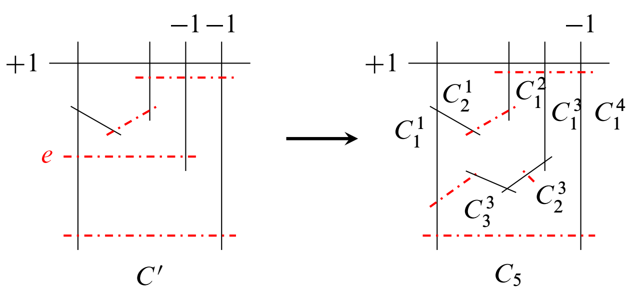


Figure 20: Curve configuration C_5 for symplectic filling W_5 of Y.

can explicitly check that there is a symplectic embedding of

$$L_1 = \overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}$$

to W_1 in Example 4.2, and W_5 is obtained by rationally blowing it down: let C_i^j be the i^{th} component of the j^{th} arm in K. Then the homological data of K for W_1 in $X = W_1 \cup K \cong \mathbb{CP}^2 \sharp 10\overline{\mathbb{CP}^2}$ is given by

$$\begin{aligned}
&[C_0] = l, \\
&[C_1^1] = l - e_2 - e_3 - e_4 - e_5, \quad [C_2^1] = e_2 - e_6, \\
&[C_1^2] = l - e_1 - e_2 - e_6, \\
&[C_1^3] = l - e_1 - e_3 - e_7, \qquad\quad [C_2^3] = e_7 - e_8, \qquad [C_3^3] = e_8 - e_9 - e_{10}, \\
&[C_1^4] = l - e_1 - e_4,
\end{aligned}$$

where C_0 is the central $(+1)$ 2–sphere of K, l is the homology class representing the complex line in $\mathbb{CP}^2$, and e_i is the homology class of each exceptional 2–sphere. As in the proof of Lemma 4.3, we can find a symplectic embedding of

$$L = \overset{-5}{\bullet}\!\!-\!\!\!-\!\!\!-\!\!\overset{-2}{\bullet}$$

to $W_1 \subset X$ whose homological data is given by $e_3 - e_5 - e_7 - e_8 - e_9$ and $e_9 - e_{10}$; refer to Figure 21, top. There are two minimal symplectic fillings of L whose corresponding curve configurations are as in Figure 21, bottom. Note that the first figure represents a linear plumbing while the second figure represents a rational homology 4–ball.

Hence, if we rationally blow down L from $X_L = L \cup K_L \cong \mathbb{CP}^2 \sharp 6\overline{\mathbb{CP}^2}$, then we get a new rational symplectic 4–manifold $X'_L \cong \mathbb{CP}^2 \sharp 4\overline{\mathbb{CP}^2}$, and the homological data of K_L changes as follows:

$$\begin{aligned}
l &\to l, \\
l - e_1 - e_2 &\to l - E_1 - E_2, \\
e_2 - e_3 &\to E_2 - E_3, \\
e_3 - e_4 &\to E_3 - E_4, \\
e_4 - e_5 - e_6 &\to E_1 - E_2 - E_3.
\end{aligned}$$

Here e_i and E_i denote the homology classes of exceptional spheres in X_L and X'_L. Note that the homological data of L in X_L is given by $e_1 - e_2 - e_3 - e_4 - e_5$ and

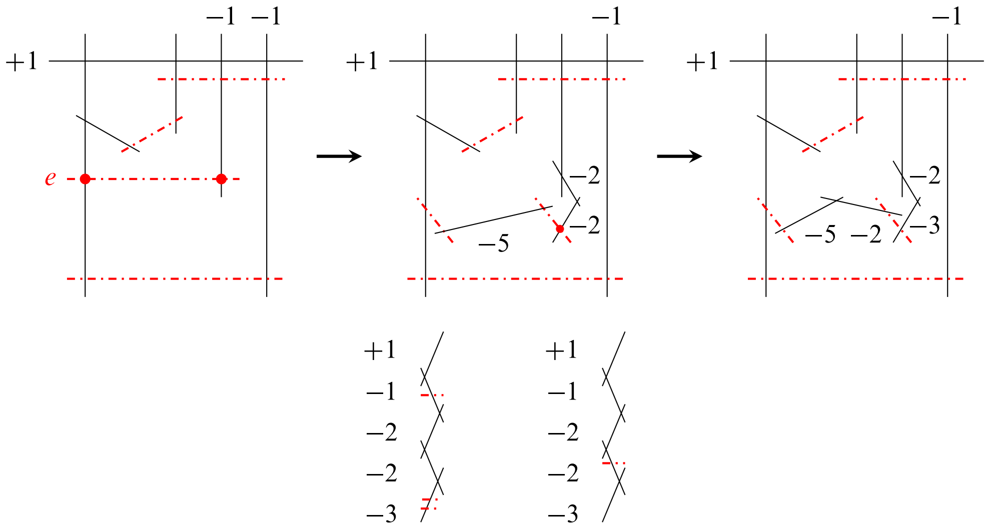


Figure 21: Top: embedding of L_1 in W_1. Bottom: two curve configurations for Y_L.

$e_5 - e_6$. Therefore, if we see X as $X_L \sharp 4\overline{\mathbb{CP}^2}$, we get $X' \cong \mathbb{CP}^2 \sharp 8\overline{\mathbb{CP}^2}$ by rationally blowing down L from X, and the homological data of K_L is changed by

$$l \to l,$$

$$l - e_3 - e_5 \to l - E_1 - E_2,$$

$$e_5 - e_7 \to E_2 - E_3,$$

$$e_7 - e_8 \to E_3 - E_4,$$

$$e_8 - e_9 - e_{10} \to E_1 - E_2 - E_3,$$

where e_1, e_2, e_4, e_6 and E_1, E_2, E_3, E_4 represent the standard exceptional 2–spheres in $X' \cong \mathbb{CP}^2 \sharp 8\overline{\mathbb{CP}^2}$. Therefore, the new homological data for the concave cap K, which give the right-hand curve configuration in Figure 20, are as follows:

$$[C_0] = l,$$
$$[C_1^1] = l - e_2 - e_4 - E_1 - E_2, \quad [C_2^1] = e_2 - e_6,$$
$$[C_1^2] = l - e_1 - e_2 - e_6,$$
$$[C_1^3] = l - e_1 - E_1 - E_3, \qquad [C_2^3] = E_3 - E_4, \quad [C_3^3] = E_1 - E_2 - E_3,$$
$$[C_1^4] = l - e_1 - e_4.$$

Remark We investigated all possible curve configurations for a small Seifert 3–manifold Y with $b \geq 5$ in the proof of the main theorem. As a result, we can find all minimal symplectic fillings of Y via corresponding curve configurations. For example, a complete list of minimal symplectic fillings of Y in Example 4.1 are given by Examples 4.1–4.3.

4.4 Proof for type D

We start to prove this case for a curve configuration coming from $C_{0,0,0}$. Note that $C_{0,0,0}$ itself is a curve configuration containing $K_{0,0,0}$ corresponding to rational homology ball filling of $\Gamma_{0,0,0}$ in Figure 3. By repeatedly blowing up at intersection points between exceptional strands and the first component of each arm, we can get a curve configuration $C_{p,q,r}$ containing $K_{p,q,r}$ corresponding to a rational homology ball filling of $\Gamma_{p,q,r}$ as in Figure 22. For notational convenience, we denote three exceptional strands in each $C_{p,q,r}$ by the same e_i with $i \in \mathbb{Z}_3$ so that e_i intersects the last component of i^{th} arm and the first component of $(i+1)^{\text{st}}$ arm of $K_{p,q,r}$. Let $C_{-1,-1,-1}$ be the right-hand figure of Figure 9 and $C_{p,q,-1}$ be a configuration of strands obtained from $C_{p,q,0}$ by blowing down e_2 in Figure 22. Then $C_{p,q,-1}$ contains $K_{p,q,-1}$, which is the proper transform of $K_{p,q,0}$ under blowing down.

Proposition 4.5 *For a curve configuration C coming from $C_{0,0,0}$, there is a curve configuration $C_{a,b,c}$ containing $K_{a,b,c}$ with $a,b,c \geq -1$ such that*

(i) *there is a sequence of blowups from $C_{a,b,c}$ to C,*

(ii) *there is either no blowup at e_i or blowups at both intersection points on e_i during the sequence of blowups,*

(iii) *there is no blowup at intersection points of $K_{a,b,c}$.*

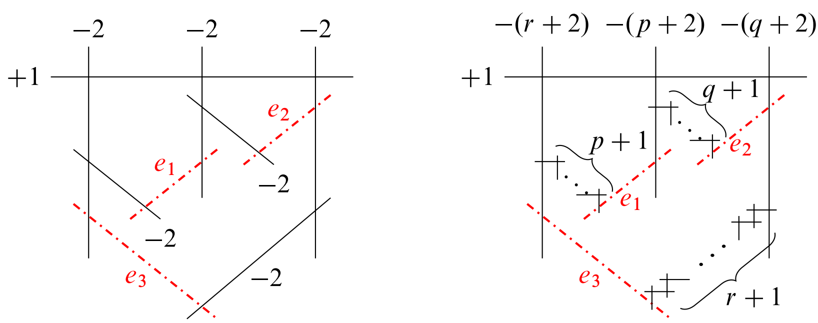


Figure 22: Curve configurations $C_{0,0,0}$ and $C_{p,q,r}$.

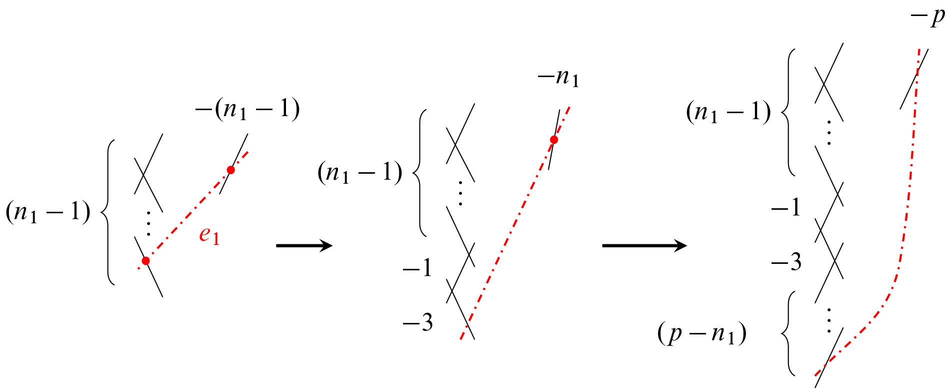


Figure 23: Part of the blowups from C_{n_1-3,n_2-3,n_3-3} to C'.

Proof Since there are no strands with degree ≤ -2 in C except for irreducible components of K, each irreducible component of $K_{0,0,0}$ in $C_{0,0,0}$ should become an irreducible component of K under blowups from $C_{0,0,0}$ to C. Hence, in order to get C from $C_{0,0,0}$ by blowups e_i, we should blow up at either two intersection points of e_i with arms or an intersection point of e_i with the $(i+1)^{\text{st}}$ arm only. Note that we get $C_{p,q,r}$ containing $K_{p,q,r}$ by blowups the latter case repeatedly. Hence, by rearranging the order of blowups from $C_{0,0,0}$ to a curve configuration C, we may assume that C is obtained from $C_{p,q,r}$ with $p,q,r \geq 0$ and there are no more blowups at an intersection point of e_i with the $(i+1)^{\text{st}}$ arm only. Since the configuration $C_{p,q,r}$ clearly satisfies conditions (i) and (ii), we are done if there is no blowup at intersection points of $K_{p,q,r}$ in $C_{p,q,r}$.

If there are blowups at intersection points of $K_{p,q,r}$ in $C_{p,q,r}$ to C, then we will find another $C_{a,b,c}$ with $a \geq p$, $b \geq q$ and $c \geq r$ satisfying conditions (i)–(iii) as follows. Let x_i be the first intersection point in the i^{th} arm of $K_{p,q,r}$ among the intersection points to be blown up, and C' be a configuration of strands obtained by blowing up at x_i for $1 \leq i \leq 3$. For notational convenience, we denote exceptional strands in $C_{p,q,r}$ and the proper transform of e_i in C' by the same e_i. There is a unique (-1) exceptional strand in each i^{th} arm of K' in C', which is the ${n_i}^{\text{th}}$ component of the i^{th} arm with $n_i \geq 2$, where K' is the total transform of $K_{p,q,r}$. Then we claim that there is a sequence of blowups from C_{n_1-3,n_2-3,n_3-3} to C': we blow up two intersection points of e_i simultaneously, and then we blow up at the intersection point between the exceptional (-1) strand and the first component of the $(i+1)^{\text{st}}$ arm consecutively to get C'; for example, see Figure 23 for the first arm. We see from the construction that a configuration C_{n_1-3,n_2-3,n_3-3} satisfies conditions (i) and (ii). Moreover, since x_i is

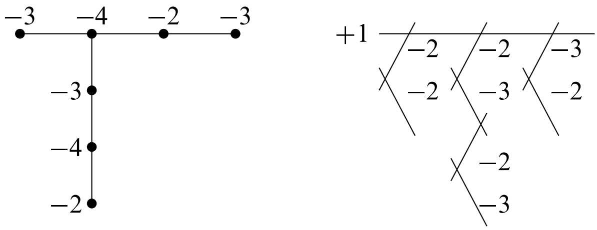


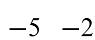
Figure 24: Plumbing graph Γ and its concave cap K.

the uppermost point among the intersection points to be blown up, there is no blowup at intersection points of K_{n_1-3,n_2-3,n_3-3} during the blowups from C_{n_1-3,n_2-3,n_3-3} to C. Therefore C_{n_1-3,n_2-3,n_3-3} is a desired curve configuration $C_{a,b,c}$. $\square$

Since $K_{a,b,c} \leq K$ (guaranteed by condition (ii) in Proposition 4.5), there is a curve configuration C_1 of Y obtained from $C_{a,b,c}$ by standard blowups. If one of a, b, c is -1, then the curve configuration C_1 is of type B or type C, so that there is a sequence of rational blowdowns from $\widetilde{M}$ to the minimal symplectic filling W_1 corresponding to C_1. If all $a, b, c \geq 0$, then W_1 is obtained from $\widetilde{M}$ by replacing $\Gamma_{a,b,c}$ with its rational homology ball filling by Lemma 4.4. On the other hand, conditions (ii) and (iii) in Proposition 4.5 guarantee that there is a sequence of rational blowdowns from W_1 to the minimal symplectic filling W corresponding to C by using Lemma 4.2 or Lemma 4.3 repeatedly.

We end this section by giving an example of minimal symplectic fillings involving 3–legged rational blowdown surgery.

Example 4.4 Let Y be a small Seifert 3–manifold whose minimal resolution graph Γ and concave cap K are given by Figure 24. We consider two minimal symplectic fillings W_1 and W_2 of Y whose respective curve configurations are given by Figure 25, top and bottom. Note that the curve configuration in Figure 25, top, is obtained from $C_{0,0,0}$ by standard blowups. Thus, as in the proof, W_1 is obtained from the minimal resolution by rationally blowing down $\Gamma_{0,0,0}$. Let us denote v_0 by a central vertex and v_i^j by i^{th} vertex of the j^{th} arm in Γ. Then, v_0, v_1^1, v_1^2 and $v_1^3 + v_2^3$ give a symplectic embedding of $\Gamma_{0,0,0}$ to the minimal resolution. A computation similar to that of Example 4.3 shows that there is a symplectic embedding L of



to W_1, and W_2 is obtained from W_1 by rationally blowing down L.

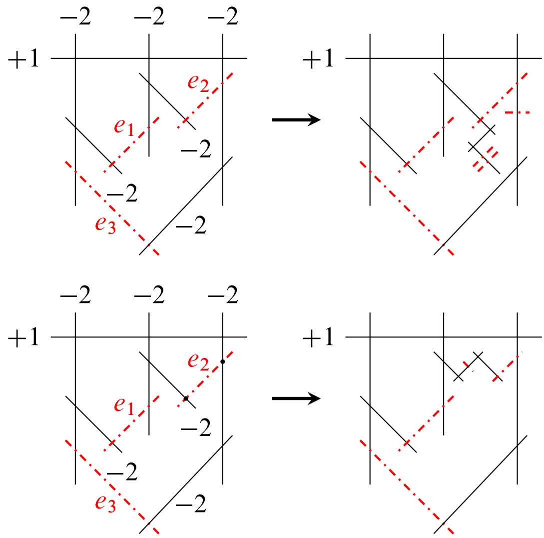


Figure 25: Top: curve configuration for W_1. Bottom: curve configuration for W_2.

References

[1] **M Bhupal**, **K Ono**, *Symplectic fillings of links of quotient surface singularities*, Nagoya Math. J. 207 (2012) 1–45 MR Zbl

[2] **M Bhupal**, **B Ozbagci**, *Symplectic fillings of lens spaces as Lefschetz fibrations*, J. Eur. Math. Soc. 18 (2016) 1515–1535 MR Zbl

[3] **M Bhupal**, **A I Stipsicz**, *Weighted homogeneous singularities and rational homology disk smoothings*, Amer. J. Math. 133 (2011) 1259–1297 MR Zbl

[4] **H Choi**, **J Park**, *A Lefschetz fibration on minimal symplectic fillings of a quotient surface singularity*, Math. Z. 295 (2020) 1183–1204 MR Zbl

[5] **R Fintushel**, **R J Stern**, *Rational blowdowns of smooth 4–manifolds*, J. Differential Geom. 46 (1997) 181–235 MR Zbl

[6] **D T Gay**, **A I Stipsicz**, *Symplectic rational blow-down along Seifert fibered 3–manifolds*, Int. Math. Res. Not. 2007 (2007) art. id. rnm084 MR Zbl

[7] **D T Gay**, **A I Stipsicz**, *Symplectic surgeries and normal surface singularities*, Algebr. Geom. Topol. 9 (2009) 2203–2223 MR Zbl

[8] **P Lisca**, *On symplectic fillings of lens spaces*, Trans. Amer. Math. Soc. 360 (2008) 765–799 MR Zbl

[9] **D McDuff**, *The structure of rational and ruled symplectic 4–manifolds*, J. Amer. Math. Soc. 3 (1990) 679–712 MR Zbl

[10] **P Orlik**, *Seifert manifolds*, Lecture Notes in Math. 291, Springer (1972) MR Zbl

[11] **P Orlik**, **P Wagreich**, *Isolated singularities of algebraic surfaces with C^* action*, Ann. of Math. 93 (1971) 205–228 MR Zbl

[12] **H Park**, **J Park**, **D Shin**, **G Urzúa**, *Milnor fibers and symplectic fillings of quotient surface singularities*, Adv. Math. 329 (2018) 1156–1230 MR Zbl

[13] **H Park**, **A I Stipsicz**, *Smoothings of singularities and symplectic surgery*, J. Symplectic Geom. 12 (2014) 585–597 MR Zbl

[14] **J Park**, *Seiberg–Witten invariants of generalised rational blow-downs*, Bull. Austral. Math. Soc. 56 (1997) 363–384 MR Zbl

[15] **H Pinkham**, *Normal surface singularities with C^* action*, Math. Ann. 227 (1977) 183–193 MR Zbl

[16] **L Starkston**, *Symplectic fillings of Seifert fibered spaces*, Trans. Amer. Math. Soc. 367 (2015) 5971–6016 MR Zbl

[17] **L Starkston**, *Comparing star surgery to rational blow-down*, J. Gökova Geom. Topol. GGT 10 (2016) 60–79 MR Zbl

[18] **A I Stipsicz**, **Z Szabó**, **J Wahl**, *Rational blowdowns and smoothings of surface singularities*, J. Topol. 1 (2008) 477–517 MR Zbl

[19] **M Symington**, *Symplectic rational blowdowns*, J. Differential Geom. 50 (1998) 505–518 MR Zbl

[20] **M Symington**, *Generalized symplectic rational blowdowns*, Algebr. Geom. Topol. 1 (2001) 503–518 MR Zbl

School of Mathematics, Korea Institute for Advanced Study
Seoul, South Korea

Department of Mathematical Sciences, Seoul National University
Seoul, South Korea

hakho@kias.re.kr, jipark@snu.ac.kr

Received: 26 August 2020 Revised: 12 July 2022

Algebraic & Geometric Topology **23**:8 (2023) 3531–3552
DOI: 10.2140/agt.2023.23.3531
Published: 5 November 2023

Milnor–Witt motivic cohomology of complements of hyperplane arrangements

KEYAO PENG

We compute the (total) Milnor–Witt motivic cohomology of the complement of a hyperplane arrangement in an affine space as an algebra with given generators and relations. We also obtain some corollaries by realization to classical cohomology.

14C25, 14F42, 19E15

1 Introduction

Let K be a perfect field of characteristic different from 2, and let $U \subset \mathbb{A}_K^N$ be the complement of a finite union of hyperplanes. For $K = \mathbb{R}$, the cohomology ring $H^*_{\mathrm{sing}}(U(\mathbb{R}), \mathbb{Z})$ is just the direct sum of $\mathbb{Z}$ corresponding to each regions (connected components), and those regions form a poset. In the special case when the hyperplanes arise from a root system, the resulting poset is the corresponding Weyl group with the weak Bruhat order. In general, the poset of regions is ranked by the number of separating hyperplanes and its Möbius function has been computed; see Edelman [8].

For any essentially smooth scheme X over K and any integers $p, q \in \mathbb{Z}$, one can define the Milnor–Witt (MW) motivic cohomology groups $H^{p,q}_{\mathrm{MW}}(X, \mathbb{Z})$ introduced by Bachmann, Calmès, Déglise, Fasel and Østvær [1]. There are homomorphisms (functorial in X), for any $p, q \in \mathbb{Z}$,

$$H^{p,q}_{\mathrm{MW}}(X, \mathbb{Z}) \to H^{p,q}_{M}(X, \mathbb{Z}),$$

where the right-hand side denotes the ordinary motivic cohomology of Voevodsky.

As illustrated by the list of properties in the following section, the Milnor–Witt motivic cohomology groups behave in a fashion similar to ordinary motivic cohomology groups, but there are crucial differences (for instance, there are no reasonable Chern classes).

In this paper, we compute the total Milnor–Witt cohomology ring of the complement of a hyperplane arrangement in affine spaces $H_{\mathrm{MW}}(U)$ using methods very similar to Chatzistamatiou [4], with some necessary modifications. To state our main result, we first recall a few facts.

Let R be a commutative ring. The Milnor–Witt K–theory of R is defined to be the graded algebra freely generated by elements of degree 1 of the form $[a]$ with $a \in R^\times$ and an element η in degree -1, subject to the relations

(1) $[a][1-a] = 0$ for any a such that $a, 1-a \in R^\times \setminus \{1\}$;

(2) $[ab] = [a] + [b] + \eta[a][b]$ for any $a, b \in R^\times$;

(3) $\eta[a] = [a]\eta$ for any $a \in R^\times$;

(4) $\eta(2 + \eta[-1]) = 0$.

It defines a presheaf on the category of schemes over a perfect field K via $X \mapsto K_*^{\mathrm{MW}}(\mathcal{O}(X))$. On the other hand, one can also consider the Milnor–Witt motivic cohomology (bigraded) presheaf

$$X \mapsto H_{\mathrm{MW}}(X) := \bigoplus_{p,q} H_{\mathrm{MW}}^{p,q}(X, \mathbb{Z}).$$

By Déglise and Fasel [7, Theorem 4.2.2], there is a morphism of presheaves

$$s \colon \bigoplus_{n\in\mathbb{Z}} K_n^{\mathrm{MW}}(-) \to \bigoplus_{n\in\mathbb{Z}} H_{\mathrm{MW}}^{n,n}(-, \mathbb{Z}) \subset H_{\mathrm{MW}}(X),$$

which specializes to the above isomorphism if $X = \mathrm{Spec}(F)$, where F is a finitely generated field extension of K; see Calmès and Fasel [3].

Theorem 1.1 *Let K be a perfect field of characteristic different from 2 and let $U \subset \mathbb{A}_K^N$ be the complement of a finite union of hyperplanes. There is an isomorphism of $H_{\mathrm{MW}}(K)$–algebras*

$$H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U \cong H_{\mathrm{MW}}(U)$$

defined by mapping $(f) \in \mathbb{G}_m(U)$ to the class $[f]$ in $H_{\mathrm{MW}}^{1,1}(U, \mathbb{Z})$ corresponding to f under s. Here, $H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$ is the free (associative) graded $H_{\mathrm{MW}}(K)$–algebra generated by $\mathbb{G}_m(U)$ in degree $(1,1)$ and J_U is the ideal generated by the elements

(1) $(f) - [f]$ *if* $f \in K^\times \subset \mathbb{G}_m(U)$;

(2) $(f) + (g) + \eta(f)(g) - (fg)$ *if* $f, g \in \mathbb{G}_m(U)$;

(3) $(f_1)(f_2)\cdots(f_t)$ *for any* $f_1, \dots, f_t \in \mathbb{G}_m(U)$ *such that* $\sum_{i=1}^t f_i = 1$;

(4) $(f)^2 - [-1](f)$ *if* $f \in \mathbb{G}_m(U)$.

As indicated above, this theorem and its proof are inspired by the computation of the (ordinary) motivic cohomology of U in [4]. We can recover the main theorem [4, Theorem 3.5] of the motivic cohomology case by taking $\eta = 0$. As a corollary, we obtain the following result:

Corollary 1.2 *Let $U \subset \mathbb{A}^N_K$ be the complement of a finite union of hyperplanes. The isomorphism of Theorem 1.1 induces an isomorphism*

$$\bigoplus_{n\in\mathbb{Z}} K_n^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U \to \bigoplus_{n\in\mathbb{Z}} H^{n,n}_{\mathrm{MW}}(U,\mathbb{Z}).$$

We do not know if the left-hand side coincides with $K_*^{\mathrm{MW}}(U)$. To conclude, we spend a few lines on the real realization homomorphism

$$H_{\mathrm{MW}}(U,\mathbb{Z}) \to H^*_{\mathrm{sing}}(U(\mathbb{R}),\mathbb{Z})$$

when U is over $K = \mathbb{R}$. We prove in particular that both sides have essentially the same generators, and that the map is surjective.

Conventions The base field K is assumed to be perfect and of characteristic not 2. For a scheme X over K, we write $H_{\mathrm{MW}}(X)$ for the total MW motivic cohomology ring $\bigoplus_{p,q\in\mathbb{Z}} H^{p,q}_{\mathrm{MW}}(X,\mathbb{Z})$.

For each $f \in \mathbb{G}_m(U)$, we use (f) to indicate the corresponding generator in the corresponding free algebra (eg $K_n^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$) and $[f]$ to indicate the corresponding element in the cohomology group (eg $H^{1,1}_{\mathrm{MW}}(U,\mathbb{Z})$).

2 Milnor–Witt motivic cohomology

In this section, we define Milnor–Witt motivic cohomology and state some properties that will be used in the proof of Theorem 1.1. We start with the (big) category of motives $\widetilde{\mathrm{DM}}(K) := \widetilde{\mathrm{DM}}_{\mathrm{Nis}}(K,\mathbb{Z})$ defined in [7, Definition 3.3.2] and the functor

$$\widetilde{M} : \mathrm{Sm}/K \to \widetilde{\mathrm{DM}}(K).$$

The category $\widetilde{\mathrm{DM}}(K)$ is symmetric monoidal [7, Proposition 3.3.4] with unit $\mathbb{1} = \widetilde{M}(\mathrm{Spec}(K))$. For any integers $p, q \in \mathbb{Z}$, we obtain MW motivic cohomology groups

$$H^{p,q}_{\mathrm{MW}}(X,\mathbb{Z}) := \mathrm{Hom}_{\widetilde{\mathrm{DM}}(K)}(\widetilde{M}(X), \mathbb{1}(q)[p]).$$

By [7, Proposition 4.1.2], motivic cohomology groups can be computed as the Zariski hypercohomology groups of explicit complexes of sheaves.

We will make use of the following property of $\widetilde{\mathrm{DM}}(K)$. First, we note that $\widetilde{\mathrm{DM}}(K)$ is also a triangulated category.

Proposition 2.1 (Gysin triangle) *Let X be a smooth K–scheme, let $Z \subset X$ be a smooth closed subscheme of codimension c and let $U = X \setminus Z$. Suppose that the normal cone $N_X Z$ admits a trivialization $\phi\colon N_X Z \cong Z \times \mathbb{A}^c$. Then there is a Gysin triangle*

$$\widetilde{M}(U) \to \widetilde{M}(X) \to \widetilde{M}(Z)(c)[2c] \xrightarrow{+1},$$

where the last two arrows depend on the choice of ϕ.

Proof We have an adjunction of triangulated categories

$$\mathrm{SH}(K) \leftrightarrows \widetilde{\mathrm{DM}}(K)$$

obtained by combining the adjunction of [6, Section 4.1] and the classical Dold–Kan correspondence (eg [5, 5.3.35]). Here, $\mathrm{SH}(K)$ is the stable homotopy category of smooth schemes over K. The functor $\mathrm{SH}(K) \to \widetilde{\mathrm{DM}}(K)$ being exact, the statement follows for instance from [13, Chapter 3, Theorem 2.23]. □

Furthermore, the Milnor–Witt motivic cohomology groups satisfy most of the formal properties of ordinary motivic cohomology and were computed in a few situations:

(1) If $q \leq 1$, there are canonical isomorphisms

$$H^{p,q}_{\mathrm{MW}}(X, \mathbb{Z}) \cong H^{p-q}_{\mathrm{Nis}}(X, \boldsymbol{K}^{\mathrm{MW}}_q) \cong H^{p-q}_{\mathrm{Zar}}(X, \boldsymbol{K}^{\mathrm{MW}}_q)$$

where $\boldsymbol{K}^{\mathrm{MW}}_q$ is the unramified Milnor–Witt K–theory sheaf (in weight q) introduced in [12].

(2) If L/K is a finitely generated field extension there are isomorphisms $H^{n,n}_{\mathrm{MW}}(L, \mathbb{Z}) \cong K^{\mathrm{MW}}_n(L)$ fitting in a commutative diagram, for any $n \in \mathbb{Z}$,

$$\begin{tikzcd} H^{n,n}_{\mathrm{MW}}(L,\mathbb{Z}) \arrow[r, "\sim"] \arrow[d] & K^{\mathrm{MW}}_n(L) \arrow[d] \\ H^{n,n}_{M}(L,\mathbb{Z}) \arrow[r, "\sim"] & K^{M}_n(L) \end{tikzcd}$$

where $K^M_n(L)$ is the (n^{th}) Milnor K–theory group of L, the bottom horizontal map is the isomorphism of Suslin, Nesterenko and Totaro, and the right-hand vertical map is the natural homomorphism from Milnor–Witt K–theory to Milnor K–theory. This result has the following consequence: the Milnor–Witt motivic cohomology groups are computed via an explicit complex of Nisnevich sheaves $\widetilde{\mathbb{Z}}(q)$ for any integer $q \in \mathbb{Z}$. The above result shows that there is a morphism of complexes of sheaves

$$\widetilde{\mathbb{Z}}(q) \to K_q^{\mathrm{MW}}[-q],$$

where the right-hand side is the complex whose only nontrivial sheaf is K_q^{MW} in degree $-q$. For any essentially smooth scheme X over K, this yields group homomorphisms

$$H_{\mathrm{MW}}^{p,q}(X, \mathbb{Z}) \to H^{p-q}(X, \boldsymbol{K}_q^{\mathrm{MW}}),$$

which are compatible with the ring structure on both sides. In the particular case $p = 2n$ and $q = n$ for some $n \in \mathbb{Z}$, we obtain isomorphisms (functorial in X)

$$H_{\mathrm{MW}}^{2n,n}(X, \mathbb{Z}) \xrightarrow{\sim} \widetilde{\mathrm{CH}}^n(X),$$

where the right-hand term is the n^{th} Chow–Witt group of X (defined in [2; 9]). Again, these isomorphisms fit into commutative diagrams

$$\begin{tikzcd}
H_{\mathrm{MW}}^{2n,n}(X, \mathbb{Z}) \arrow[r, "\sim"] \arrow[d] & \widetilde{\mathrm{CH}}^n(X) \arrow[d] \\
H_{M}^{2n,n}(X, \mathbb{Z}) \arrow[r, "\sim"] & \mathrm{CH}^n(X)
\end{tikzcd}$$

where the right-hand vertical homomorphism is the natural map from Chow–Witt groups to Chow groups.

(3) The total Milnor–Witt motivic cohomology has Borel classes for symplectic bundles [15] but in general the projective bundle theorem fails [14].

(4) If X is a smooth scheme over $\mathbb{R}$, there are two interesting realization maps. On the one hand, one may consider the composite

$$H_{\mathrm{MW}}^{p,q}(X, \mathbb{Z}) \to H_{M}^{p,q}(X, \mathbb{Z}) \to H_{\mathrm{sing}}^{p}(X(\mathbb{C}), \mathbb{Z}),$$

where the right-hand map is the complex realization map. On the other hand, one may also consider the composite

$$H_{\mathrm{MW}}^{p,q}(X, \mathbb{Z}) \to H^{p-q}(X, \boldsymbol{K}_q^{\mathrm{MW}}) \to H^{p-q}(X, \boldsymbol{I}^q) \to H_{\mathrm{sing}}^{p-q}(X(\mathbb{R}), \mathbb{Z}),$$

where $\boldsymbol{I}^q$ is the unramified sheaf associated to the q^{th} power of the fundamental ideal in the Witt ring, $\boldsymbol{K}_q^{\mathrm{MW}} \to \boldsymbol{I}^q$ is the canonical projection and $H^{p-q}(X, \boldsymbol{I}^q) \to H_{\mathrm{sing}}^{p-q}(X(\mathbb{R}), \mathbb{Z})$ is Jacobson's signature map [11].

We note here that these two realization maps show that Milnor–Witt motivic cohomology is in some sense the analogue of both the singular cohomology of the complex and the real points of X.

3 Basic structure of the cohomology ring

Let V be an affine space, ie $V \cong \mathbb{A}_K^N$ for some $N \in \mathbb{N}$. We consider finite families I of hyperplanes in V (which we suppose are distinct). We denote by $|I|$ the cardinality of I and set $U_I^V := V \setminus \left(\bigcup_{Y \in I} Y\right)$, and simply write U_I^N when $V = \mathbb{A}_K^N$. For any hyperplane Y, we put $I_Y := \{Y_i \cap Y \mid Y_i \in I, Y_i \neq Y\}$.

Proposition 3.1 *Let V and I be as above. We have*

$$\widetilde{M}(U_I^V) \cong \bigoplus_{j \in J} \mathbb{1}(n_j)[n_j]$$

for some set J and integers $n_j \geq 0$.

Proof We proceed by induction on the dimension N of V and $|I|$. If $|I| = 0$, then $\widetilde{M}(U_I^V) = \widetilde{M}(V) \cong \mathbb{1}$ and we are done. So let $|I| \geq 1$ and $Y \in I$. The Gysin triangle reads as

$$\widetilde{M}(U_I^V) \to \widetilde{M}(U_{I \setminus \{Y\}}^V) \xrightarrow{\phi} \widetilde{M}(U_{I_Y}^Y)(1)[2] \xrightarrow{+1}. \tag{3-1}$$

If $\phi = 0$, then the triangle is split and consequently we obtain an isomorphism

$$\widetilde{M}(U_I^V) \cong \widetilde{M}(U_{I \setminus \{Y\}}^V) \oplus \widetilde{M}(U_{I_Y}^Y)(1)[1]. \tag{3-2}$$

Since $|I \setminus \{Y\}| < |I|$ and $\dim(Y) = \dim(V) - 1$, we conclude by induction that the right-hand side has the correct form. We are then reduced to showing that $\phi = 0$.

By induction,

$$\begin{aligned}\phi \in \mathrm{Hom}_{\widetilde{\mathrm{DM}}(K)}(\widetilde{M}(U_{I \setminus \{Y\}}^V), \widetilde{M}(U_{I_Y}^Y)(1)[2]) \\ \cong \bigoplus_{j,k} \mathrm{Hom}_{\widetilde{\mathrm{DM}}(K)}(\mathbb{1}(n_j)[n_j], \mathbb{1}(m_k)[m_k + 1])\end{aligned}$$

for some integers $n_j, m_k \geq 0$, so it suffices to prove that $\mathrm{Hom}_{\widetilde{\mathrm{DM}}(K)}(\mathbb{1}, \mathbb{1}(m)[m+1]) = 0$ for any $m \in \mathbb{Z}$ to conclude. Now,

$$\mathrm{Hom}_{\widetilde{\mathrm{DM}}(K)}(\mathbb{1}, \mathbb{1}(m)[m + 1]) = H_{\mathrm{MW}}^{m+1,m}(K, \mathbb{Z})$$

and the latter is trivial by [7, Proposition 4.1.2 and proof of Theorem 4.2.4]. □

As an immediate corollary, we obtain the following result:

Corollary 3.2 *The motivic cohomology $H_{\mathrm{MW}}(U_I^V)$ is a finitely generated, free $H_{\mathrm{MW}}(K)$–module.*

To obtain more precise results, we now study the Gysin (split) triangle (3-1) in more detail. We can rewrite it as

$$\widetilde{M}(U_{I_Y}^Y)(1)[1] \xrightarrow{\beta^Y} \widetilde{M}(U_I^V) \xrightarrow{\alpha^Y} \widetilde{M}(U_{I\setminus\{Y\}}^V) \xrightarrow{0}$$

and therefore we obtain the short (split) exact sequence, in which the morphisms are induced by the first two morphisms in the triangle,

$$(3\text{-}3)\quad 0 \to \bigoplus_{p,q} H_{\mathrm{MW}}^{p,q}(U_{I\setminus\{Y\}}^V, \mathbb{Z}) \xrightarrow{\alpha_*^Y} \bigoplus_{p,q} H_{\mathrm{MW}}^{p,q}(U_I^V, \mathbb{Z}) \xrightarrow{\beta_*^Y} \bigoplus_{p,q} H_{\mathrm{MW}}^{p-1,q-1}(U_{I_Y}^Y, \mathbb{Z}) \to 0.$$

The inclusion $Y \subset V$ yields a morphism $U_{I_Y}^Y \to U_{I\setminus\{Y\}}^V$ and therefore a morphism $\iota\colon \widetilde{M}(U_{I_Y}^Y) \to \widetilde{M}(U_{I\setminus\{Y\}}^V)$. The global section f of V corresponding to the equation of Y becomes invertible in U_I^V and therefore yields a morphism $[f]\colon \widetilde{M}(U_I^V) \to \mathbb{1}(1)[1]$ corresponding to the class $[f] \in H_{\mathrm{MW}}^{1,1}(U_I^V, \mathbb{Z})$ given by the morphism

$$s\colon \bigoplus_{n\in\mathbb{Z}} K_n^{\mathrm{MW}}(-) \to \bigoplus_{n\in\mathbb{Z}} H_{\mathrm{MW}}^{n,n}(-, \mathbb{Z}).$$

Lemma 3.3 *The following diagram commutes*:

$$\begin{tikzcd}
\widetilde{M}(U_{I_Y}^Y)(1)[1] \arrow[r, "\iota(1)[1]"] \arrow[d, "\beta^Y"'] & \widetilde{M}(U_{I\setminus\{Y\}}^V)(1)[1] \\
\widetilde{M}(U_I^V) \arrow[r, "\Delta"'] & \widetilde{M}(U_I^V) \otimes \widetilde{M}(U_I^V) \arrow[u, "\alpha^Y \otimes [f]"']
\end{tikzcd}$$

Proof The commutative diagram of schemes

$$\begin{tikzcd}
U_I^V \arrow[r] \arrow[d] & U_{I\setminus\{Y\}}^V \arrow[d, "(\mathrm{Id}, f)"] \\
U_{I\setminus\{Y\}}^V \times \mathbb{G}_m \arrow[r] & U_{I\setminus\{Y\}}^V \times \mathbb{A}_K^1
\end{tikzcd}$$

yields a morphism of Gysin triangles and thus a commutative diagram

$$\begin{tikzcd}
\widetilde{M}(U^Y_{I_Y})(1)[1] \arrow[r, "\beta^Y"] \arrow[d, "\iota(1)[1]"'] & \widetilde{M}(U^V_I) \arrow[r, "\alpha^Y"] \arrow[d] & \widetilde{M}(U^V_{I\setminus\{Y\}}) \arrow[r] \arrow[d] & \cdots \\
\widetilde{M}(U^V_{I\setminus\{Y\}})(1)[1] \arrow[r] \arrow[dr, equal] & \widetilde{M}(U^V_{I\setminus\{Y\}} \times \mathbb{G}_m) \arrow[r] \arrow[d] & \widetilde{M}(U^V_{I\setminus\{Y\}} \times \mathbb{A}^1_K) \arrow[r] & \cdots \\
& \widetilde{M}(U^V_{I\setminus\{Y\}})(1)[1] & &
\end{tikzcd}$$

in which the map $\widetilde{M}(U^V_{I\setminus\{Y\}} \times \mathbb{G}_m) \to \widetilde{M}(U^V_{I\setminus\{Y\}})(1)[1]$ is just the projection. We conclude by observing that the middle vertical composite is just $(\alpha^Y \otimes [f]) \circ \Delta$. $\square$

We may now prove the main result of this section.

Proposition 3.4 *The cohomology ring $H_{\mathrm{MW}}(U)$ is generated by the classes of units in U as an $H_{\mathrm{MW}}(K)$–algebra. In particular, the homomorphism*

$$s\colon \bigoplus_{n\in\mathbb{Z}} K^{\mathrm{MW}}_n(U) \to \bigoplus_{n\in\mathbb{Z}} H^{n,n}_{\mathrm{MW}}(U,\mathbb{Z})$$

is surjective.

Proof We again prove the result by induction on $|I|$ and the dimension of V, the case $|I| = 0$ being obvious. Suppose then that the result holds for $U^Y_{I_Y}$ and $U^V_{I\setminus\{Y\}}$ and consider the split sequence (3-3). For any $x \in H_{\mathrm{MW}}(U) = H_{\mathrm{MW}}(U^V_I)$, we have that $\beta^Y_*(x) \in H_{\mathrm{MW}}(U^Y_{I_Y})$ is in the subalgebra generated by $\{[f] \mid f \in \mathbb{G}_m(U^Y_{I_Y})\}$ and η. For any $f_1,\dots,f_n \in \mathbb{G}_m(U^V_{I\setminus\{Y\}})$, Lemma 3.3 yields

$$\beta^Y_*\big([(f_1)|_{U^V_I}]\cdots[(f_n)|_{U^V_I}]\cdot[t]\big) = [(f_1)|_{U^Y_{I_Y}}]\cdots[(f_n)|_{U^Y_{I_Y}}].$$

The map $\mathbb{G}_m(U^V_{I\setminus\{Y\}}) \to \mathbb{G}_m(U^Y_{I_Y})$ being surjective, it follows that there exists $x' \in H_{\mathrm{MW}}(U^V_I)$ in the subalgebra generated by units such that $\beta^Y_*(x - x') = 0$. Thus, $x - x' = \alpha_*(y)$ for some $y \in H_{\mathrm{MW}}(U^V_{I\setminus\{Y\}})$ and the result follows from the fact that α_* is just induced by the inclusion $U^V_I \subset U^V_{I\setminus\{Y\}}$. $\square$

4 Relations in the cohomology ring

The purpose of this section is to prove that the relations of Theorem 1.1 hold in $H_{\mathrm{MW}}(U)$. The first two relations are obviously satisfied since the homomorphism is induced by the ring homomorphism

$$s\colon \bigoplus_{n\in\mathbb{Z}} K_n^{\mathrm{MW}}(U) \to \bigoplus_{n\in\mathbb{Z}} H_{\mathrm{MW}}^{n,n}(U,\mathbb{Z}).$$

Recall now that the last two relations are

(3) $[f_1][f_2]\cdots[f_t]$ if $f_i \in \mathbb{G}_m(U)$ for any i and $\sum_{i=1}^t f_i = 1$;

(4) $[f]^2 - [-1][f]$ if $f \in \mathbb{G}_m(U)$.

We will prove that they are equal to 0 in $H_{\mathrm{MW}}(U)$. Actually, it will be more convenient to work with the following relations, where $\epsilon := -\langle -1\rangle = -1-\eta[-1]$:

(3′) $R(f_0,\dots,f_t)$, defined by

$$\sum_{i=0}^{t} \epsilon^{t+i}[f_0]\cdots\widehat{[f_i]}\cdots[f_t] + \sum_{0\le i_0<\cdots<i_k\le t} (-1)^k[-1]^k[f_0]\cdots\widehat{[f_{i_0}]}\cdots\widehat{[f_{i_k}]}\cdots[f_t]$$

for $f_i \in \mathbb{G}_m(U)$ such that $\sum_{i=0}^t f_i = 0$.

(4′) **Anticommutativity** $[f][g]-\epsilon[g][f]$.

Lemma 4.1 *The two groups of relations are equivalent in $H_{\mathrm{MW}}(U)$.*

Proof We first assume that (3) and (4) are satisfied. Since (1) and (2) are satisfied, we have $[-f] = [-1]+\langle -1\rangle[f]$. As (4) is satisfied and $[-1]=\epsilon[-1]$ in $K_*^{\mathrm{MW}}(K)$,

$$[-f][f] = [-1][f] + \langle -1\rangle [f]^2 = \epsilon([-1][f]-[f]^2) = 0$$

and then $[fg][-fg] = [f][g]+\epsilon[g][f]$ for any $g,f \in \mathbb{G}_m(U)$ by [12, proof of Lemma 3.7]. Suppose next that $\sum_{i=0}^t f_i = 0$, so that $\sum_{i=1}^t f_i/(-f_0) = 1$. Combining (3) and the anticommutativity law, we obtain

$$0 = [1] = [f_j^{-1}] + \langle f_j^{-1}\rangle [f_j] \qquad \text{(by (2))}, \tag{4-1}$$

$$\begin{aligned}\left[\frac{-f_i}{f_j}\right] &= \langle f_j^{-1}\rangle[-f_i] + [f_j^{-1}] \\ &= \langle f_j^{-1}\rangle([-f_i]-[f_j]) \qquad \text{(by (4-1))} \\ &= \langle f_j^{-1}\rangle(\langle -1\rangle[f_i]+[-1]-[f_j]),\end{aligned} \tag{4-2}$$

$$
\begin{aligned}
\text{(4-3)}\qquad ([f_0]-[-1])^k &= \sum_{i=0}^{k}\binom{k}{i}[-1]^{k-i}[f_0]^i \\
&= \Bigl(\sum_{i=0}^{k-1}\binom{k}{i}\Bigr)[-1]^{k-1}[f_0]+(-1)^k[-1]^k \qquad \text{(by (4))} \\
&= (-1)^{k-1}[-1]^{k-1}[f_0]+(-1)^k[-1]^k
\end{aligned}
$$

and

$$
\begin{aligned}
0 &= (-\langle f_0\rangle)^t\Bigl[\frac{-f_1}{f_0}\Bigr]\Bigl[\frac{-f_2}{f_0}\Bigr]\cdots\Bigl[\frac{-f_t}{f_0}\Bigr] && \text{(by (3))} \\
&= \bigl([f_0]-[-1]-\langle -1\rangle[f_1]\bigr)\cdots\bigl([f_0]-[-1]-\langle -1\rangle[f_t]\bigr) && \text{(by (4-2))} \\
&= \epsilon^t\widehat{[f_0]}[f_1]\cdots[f_t]+\sum_{i=1}^{t}\epsilon^{t-1}[f_1]\cdots\widehat{[f_i]}([f_0]-[-1])\cdots[f_t] \\
&\qquad +\sum_{i<j}([f_0]-[-1])^2[f_1]\cdots\widehat{[f_i]}\cdots\widehat{[f_j]}\cdots[f_t]+\cdots \\
&= \sum_{i=0}^{t}\epsilon^{t+i}[f_0]\cdots\widehat{[f_i]}\cdots[f_t] \\
&\qquad +\sum_{0\le i_0<\cdots<i_k\le t}(-1)^k[-1]^k[f_0]\cdots\widehat{[f_{i_0}]}\cdots\widehat{[f_{i_k}]}\cdots[f_t] && \text{(by (4-3))} \\
&= R(f_0,\dots,f_t).
\end{aligned}
$$

Conversely, suppose that $(3')$ and $(4')$ hold. A direct calculation shows that

$$R(-1,f_1,\dots,f_t)=(-\langle -1\rangle)^t[f_1]\cdots[f_t]=\epsilon^t[f_1]\cdots[f_t],$$

and consequently that (3) also holds. For every field $K\neq\mathbb{F}_2$, we have $1+a+b=0$ for some $a,b\neq 0$ and it follows from $[-a][-b]=0$ in $K_*^{\mathrm{MW}}(K)$ that

$$
\begin{aligned}
R(f,af,bf) &= R(f,af,bf)-[-a][-b]=R(f,af,bf)-\Bigl[-\frac{af}{f}\Bigr]\Bigl[-\frac{bf}{f}\Bigr] \\
&= R(f,af,bf)-(\langle -1\rangle[af]+[-1]-[f])(\langle -1\rangle[bf]+[-1]-[f]) \\
&= -[-1][f]+[-1]^2-([f]-[-1])^2=[-1][f]-[f]^2. \qquad \square
\end{aligned}
$$

Remark 4.2 The following properties of the relations R and anticommutativity hold:

(1) For any $a,b\in\mathbb{G}_m(U)$, we have $[a/b]=-\langle b^{-1}\rangle R(b,-a)$.

(2) For any $f_0, \dots, f_t \in \mathbb{G}_m(U)$, by direct computation, we have
$$R(f_0, \dots, f_t) - \epsilon^i [f_i] R(f_0, \dots, \hat{f_i}, \dots, f_t) = P(f_0, \dots, \hat{f_i}, \dots, f_t)$$
for some polynomial P. This uses the anticommutativity and the fact that $[-1] = \epsilon^j [-1]$ for any $j \geq 0$ in the computation.

(3) For any $f_0, \dots, f_t \in K^\times$ such that $\sum_{i=0}^t f_i = 0$, we have $R(f_0, \dots, f_t) = 0$ in $K_*^{\mathrm{MW}}(K)$.

The following lemma will prove useful in the proof of the main theorem:

Lemma 4.3 *Any morphism $\phi\colon \widetilde{M}(U_I^V) \to T$ in $\widetilde{\mathrm{DM}}(K)$ such that*
$$\widetilde{M}(U_{I_Y}^Y)(1)[1] \xrightarrow{\beta^Y} \widetilde{M}(U_I^V) \xrightarrow{\phi} T$$
is trivial for every $Y \in I$ factors through $\widetilde{M}(K)$, ie there is a morphism $\psi\colon \widetilde{M}(K) \to T$ such that the diagram
$$\begin{tikzcd} \widetilde{M}(U_I^V) \arrow[r, "\phi"] \arrow[d] & T \\ \widetilde{M}(K) \arrow[ur, "\psi"'] & \end{tikzcd}$$
is commutative.

Proof We prove as usual the result by induction on $|I|$, the result being trivial if $|I| = 0$, ie if $U_I^V \cong \mathbb{A}_K^N$. By assumption, ϕ factors through $\widetilde{M}(U_{I\setminus\{Y\}}^V)$, ie we have a commutative diagram
$$\begin{tikzcd} \widetilde{M}(U_I^V) \arrow[r, "\alpha^Y"] \arrow[dr, "\phi"'] & \widetilde{M}(U_{I\setminus\{Y\}}^V) \arrow[d, "\phi_0"] \\ & T \end{tikzcd}$$
For $H \in I' = I \setminus \{Y\}$, we have an associated Gysin morphism $\beta^H : \widetilde{M}(U_{I_H}^H)(1)[1] \to \widetilde{M}(U_I^V)$ which induces a commutative diagram
$$\begin{tikzcd} \widetilde{M}(U_{I_H}^H)(1)[1] \arrow[r, "\beta^H"] \arrow[d, "\alpha^Y(1)[1]"'] & \widetilde{M}(U_I^V) \arrow[r, "\phi"] \arrow[d, "\alpha^Y"] & T \\ \widetilde{M}(U_{I'_H}^H)(1)[1] \arrow[r, "\beta^H"] & \widetilde{M}(U_{I'}^V) \arrow[ur, "\phi_0"'] & \end{tikzcd}$$

in which the morphism $\alpha^Y(1)[1]$ on the left is split surjective. It follows that

$$\phi_0 \circ \beta^H \circ \alpha^Y(1)[1] = \phi \circ \beta^H = 0$$

implies $\phi_0 \circ \beta^H = 0$. We conclude by induction. $\square$

Proposition 4.4 *Let S be an essentially smooth K–scheme and let $f_i \in \mathbb{G}_m(S)$ be such that $\sum_{i=0}^t f_i = 0$. Then*

$$R(f_0, \dots, f_t) = 0 \quad \textit{in } H_{\mathrm{MW}}(S).$$

Proof The global sections $f_0, \dots, f_t$ yield a morphism $j = (f_0, \dots, f_t)\colon S \to \mathbb{A}_K^{t+1}$ which restricts to a morphism $j : S \to U_I^H$, where $H \subset \mathbb{A}_K^{t+1}$ is given by $\sum_{i=0}^t x_i = 0$ and $I = \{\{x_1 = 0\}, \dots, \{x_t = 0\}\}$. Since $R(f_0, \dots, f_t) = j^*(R(x_0, \dots, x_t))$, we can reduce the proposition to the case $S = U_I^H$.

For any x_j, we set $Y_j := \{x_j = 0\} \subset H$ and we obtain a Gysin morphism

$$\beta_j : \widetilde{M}(U_{I_{Y_j}}^{Y_j})(1)[1] \to \widetilde{M}(U_I^H)$$

and a composite

$$\widetilde{M}(U_{I_{Y_j}}^{Y_j})(1)[1] \xrightarrow{\beta_j} \widetilde{M}(U_I^H) \xrightarrow{R(x_0,\dots,x_t)} \mathbb{1}(t)[t].$$

By Remark 4.2 and Lemma 3.3,

$$\begin{aligned} R(x_0, \dots, x_t) \circ \beta_j \\ &= (\epsilon^j [x_j] R(x_0, \dots, \hat{x}_j, \dots, x_t) + P(x_0, \dots, \hat{x}_j, \dots, x_t)) \circ \beta_j \\ &= \epsilon^j ([x_j] R(x_0, \dots, \hat{x}_j, \dots, x_t)) \circ \beta_j + P(x_0, \dots, \hat{x}_j, \dots, x_t) \circ \alpha_j \circ \beta_j \\ &= \epsilon^j R(x_0|_{U_{I_{Y_j}}^{Y_j}}, \dots, \hat{x}_j, \dots, x_t|_{U_{I_{Y_j}}^{Y_j}}). \end{aligned}$$

As $R(f, -f) = 0$ for $f \in \mathbb{G}_m(S)$ by Remark 4.2, we obtain by induction that $R(x_0, \dots, x_t) \circ \beta_j = 0$ for any $j = 0, \dots, t$. Applying Lemma 4.3, we obtain a commutative diagram

$$\begin{tikzcd} \widetilde{M}(U_I^H) \arrow[r, "R(x_0,\dots,x_t)"] \arrow[d] & \mathbb{1}(t)[t] \\ \widetilde{M}(K) \arrow[ur, "\psi"'] & \end{tikzcd}$$

As $\mathrm{char}(K) \neq 2$, U_I^H has a K–rational point $(\lambda_0, \dots, \lambda_t) \in \mathbb{A}_K^{t+1}$, and we obtain a diagram

$$\begin{tikzcd}
\widetilde{M}(K) \arrow[d, "u"] \arrow[drr, "R(\lambda_0,\dots,\lambda_t)"] & & \\
\widetilde{M}(U_I^H) \arrow[rr, "R(x_0,\dots,x_t)"] \arrow[d] & & \mathbb{1}(t)[t] \\
\widetilde{M}(K) \arrow[urr, "\psi"'] & &
\end{tikzcd}$$

The vertical composite being the identity, $\psi = R(\lambda_0, \dots, \lambda_t)$, and the latter is trivial by the relations in Milnor–Witt K–theory. □

Applying Lemma 4.1, we obtain the following corollary:

Corollary 4.5 *Let S be an essentially smooth K–scheme.*

(1) *For any $f_1, \dots, f_t \in \mathbb{G}_m(S)$ such that $\sum_{i=1}^t f_i = 1$, we have*

$$[f_1][f_2]\cdots[f_t] = 0 \in H_{\mathrm{MW}}(S).$$

(2) *For any $f \in \mathbb{G}_m(S)$, we have $[f]^2 - [-1][f] = 0$ in $H_{\mathrm{MW}}(S)$.*

5 Proof of the main theorem

In this section, we prove Theorem 1.1. We denote by $J_U \subset H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$ the ideal generated by the relations

(1) $(f) - [f]$ for $f \in K^\times \subset \mathbb{G}_m(U)$;

(2) $(f) + (g) + \eta(f)(g) - (fg)$ for $f, g \in \mathbb{G}_m(U)$;

(3) $(f_1)(f_2)\cdots(f_t)$ for any $f_1, \dots, f_t \in \mathbb{G}_m(U)$ such that $\sum_{i=1}^t f_i = 1$;

(4) $(f)^2 - [-1](f)$ for $f \in \mathbb{G}_m(U)$.

By Lemma 4.1, $J_U \subset H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$ is in fact generated by

(1) $(f) - [f]$ for $f \in K^\times \subset \mathbb{G}_m(U)$;

(2) $(f) + (g) + \eta(f)(g) - (fg)$ for $f, g \in \mathbb{G}_m(U)$;

(3′) **Anticommutativity** $(f)(g) - \epsilon(g)(f)$ for any $f, g \in \mathbb{G}_m(U)$;

(4′) $R(f_0,\dots,f_t)$, given by

$$\sum_{i=0}^{t}\epsilon^{t+i}(f_0)\cdots\widehat{(f_i)}\cdots(f_t) + \sum_{0\le i_0<\cdots<i_k\le t}(-1)^k[-1]^k(f_0)\cdots\widehat{(f_{i_0})}\cdots\widehat{(f_{i_k})}\cdots(f_t)$$

for any $f_0,\dots,f_t\in\mathbb{G}_m(U)$ such that $\sum_{i=0}^{t} f_i=0$.

In view of Corollary 4.5, the morphism $H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}\to H_{\mathrm{MW}}(U)$ defined by $(f)\mapsto[f]$ induces a morphism of $H_{\mathrm{MW}}(K)$–algebras

$$\rho\colon H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U\to H_{\mathrm{MW}}(U).$$

Now, choose linear polynomials $\phi_1,\dots,\phi_s$ that define the hyperplanes $Y_i\in I$ and let $J'_U\subset H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$ be the ideal generated by the relations (1), (2), (3′) and (4′) for elements of the form $f_j=\lambda_j\phi_{i_j}$ or $f_j=\lambda_j$ for $\lambda_j\in K^\times$ and $\phi_{i_j}\in\{\phi_1,\dots,\phi_s\}$. We have a string of surjective morphisms of $H_{\mathrm{MW}}(K)$–algebras

$$H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J'_U\to H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U\xrightarrow{\rho}H_{\mathrm{MW}}(U),$$

whose composite we denote by ρ'.

Theorem 5.1 *The morphism of $H_{\mathrm{MW}}(K)$–algebras*

$$H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U\xrightarrow{\rho}H_{\mathrm{MW}}(U)$$

is an isomorphism.

Proof It suffices to prove that ρ' is an isomorphism. To see this, we work again by induction on $|I|$. If $|I|=0$, then $U\cong\mathbb{A}^N_K$ for some $N\in\mathbb{N}$. By homotopy invariance, we have to prove that the map

$$\rho'\colon H_{\mathrm{MW}}(K)\{\mathbb{G}_m(K)\}/J'_K\to H_{\mathrm{MW}}(K)$$

is an isomorphism. Now, the morphism of $H_{\mathrm{MW}}(K)$–algebras

$$H_{\mathrm{MW}}(K)\to H_{\mathrm{MW}}(K)\{\mathbb{G}_m(K)\}/J'_K$$

is surjective by relation (1). Its composite with ρ' is the identity and we conclude in that case.

Assume now that $Y\in I$ is defined by $\phi_1=0$ and that we have isomorphisms

$$H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U^V_{I'})\}/J'_{U^V_{I'}}\xrightarrow{\sim}H_{\mathrm{MW}}(U^V_{I'}),$$
$$H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U^Y_{I_Y})\}/J'_{U^Y_{I_Y}}\xrightarrow{\sim}H_{\mathrm{MW}}(U^Y_{I_Y}).$$

The morphism $U_I^V \to U_{I'}^V$ induces a morphism $\mathbb{G}_m(U_{I'}^V) \to \mathbb{G}_m(U_I^V)$ and then a commutative diagram

$$\begin{tikzcd}
 & 0 \arrow[d] \\
H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_{I'}^V)\}/J'_{U_{I'}^V} \arrow[r, "\cong"'] \arrow[d, "\widetilde{\alpha}"'] & H_{\mathrm{MW}}(U_{I'}^V) \arrow[d, "\alpha_*^Y"] \\
H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_I^V)\}/J'_{U_I^V} \arrow[r, "\rho'"] \arrow[d, "\widetilde{\beta}"'] & H_{\mathrm{MW}}(U_I^V) \arrow[d, "\beta_*^Y"] \\
H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_{I_Y}^Y)\}/J'_{U_{I_Y}^Y} \arrow[r, "\cong"'] & H_{\mathrm{MW}}(U_{I_Y}^Y) \arrow[d] \\
 & 0
\end{tikzcd}$$

in which $\widetilde{\beta}$ is the unique lift of $\beta_*^Y \circ \rho$ and the right column is exact. We are thus reduced to proving that the left vertical sequence is short exact to conclude. It is straightforward to check that $\widetilde{\alpha}$ is injective and $\widetilde{\beta}$ is surjective. Moreover, the commutativity of the diagram and the fact that $\beta_*^Y \circ \alpha_*^Y = 0$ imply that $\widetilde{\beta} \circ \widetilde{\alpha} = 0$, so we are left to prove exactness in the middle.

Let $x \in H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_I^V)\}/J'_{U_I^V}$. The group $\mathbb{G}_m(U_I^V)$ being generated by $\mathbb{G}_m(U_{I'}^V)$ and ϕ_1, we may use relations (2) and (4) to see that $x = (\phi_1)\widetilde{\alpha}(x_1) + \widetilde{\alpha}(x_0)$ in $H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_I^V)\}/J'_{U_I^V}$. By Lemma 3.3, we get $\widetilde{\beta}(x) = \tilde{\imath}(x_1)$, where $\tilde{\imath}$ is induced by the restriction $\mathbb{G}_m(U_{I'}^V) \to \mathbb{G}_m(U_{I_Y}^Y)$. Consequently, we need to prove that, if $\tilde{\imath}(x_1) = 0$, then $(\phi_1)\widetilde{\alpha}(x_1)$ is in the image of $\widetilde{\alpha}$. With this in mind, we now prove that the kernel of $\tilde{\imath}$ is generated by elements of the form

$$R(f_0, \dots, f_t),$$

where $f_j = \lambda\phi_{i_j}$ with $i_j > 1$ or $f_j = \lambda$ and $\sum_{i=0}^t f_i|_{U_{I_Y}^Y} = 0$. Denote by L' the ideal of $H_{\mathrm{MW}}(K)\{\mathbb{G}_m(U_{I'}^V)\}$ generated by such elements. By construction, the restriction induces a homomorphism

$$L' + J'_{U_{I'}^V} \to J'_{U_{I_Y}^Y},$$

which is surjective. Indeed, relations (1), (2) and (3′) can be lifted using the fact that the map $\mathbb{G}_m(U_{I'}^V) \to \mathbb{G}_m(U_{I_Y}^Y)$ is surjective, while an element satisfying relation (4) with every f_j of the form $f_j = \lambda_j\phi_{i_j}$ or $f_j = \lambda_j$ for $\lambda_j \in K^\times$ (with $i_j \neq 1$) lifts to an element in L'. As in [4, proof of Theorem 3.5], we see that the kernel of the group homomorphism $\mathbb{G}_m(U_{I'}^V) \to \mathbb{G}_m(U_{I_Y}^Y)$ is generated by elements of the form

(1) $\lambda\phi_i/\phi_j$ with i and j such that $Y_1 \cap Y_i = Y_1 \cap Y_j$ and $\lambda = (\phi_j)|_{Y_1}/(\phi_i)|_{Y_1}$;

(2) $\lambda\phi_i$, where i is such that $Y_1 \cap Y_i = \varnothing$ and $\lambda = 1/(\phi_i)|_{Y_1}$.

Remark 4.2 yields

$$\left[\frac{\lambda\cdot\phi_i}{\phi_j}\right] = -\langle\phi_j^{-1}\rangle R(\phi_j, -\lambda\cdot\phi_i) \subset L',$$

while $[\lambda\phi_i] = \epsilon R(-1, \lambda\cdot\phi_i) \subset L'$ showing that $\ker(\mathbb{G}_m(U_{I'}^V) \to \mathbb{G}_m(U_{I_Y}^Y)) \subset L' + J'_{U_{I'}^V}$. We deduce that $\ker(\tilde{\iota}) = L'$.

We now conclude. If $\tilde{\iota}(x_1) = 0$, then $x_1 \in L'$ and we may suppose that $x_1 = R(f_0, \dots, f_t)$ for $f_0, \dots, f_t$ such that $\sum_{i=0}^t f_i|_{U_{I_Y}^Y} = 0$. It follows that $\sum_{i=0}^t f_i = -\mu\phi_1$ for $\mu \in K$. If $\mu = 0$, there is nothing to do. Otherwise, use $R(\mu\phi_1, f_0, \dots, f_t) = 0$ and Remark 4.2 to get

$$\begin{aligned}(\phi_1)\tilde{\alpha}(x_1) &= (\mu\phi_1)\tilde{\alpha}(x_1) - \langle\phi_1\rangle(\mu)\tilde{\alpha}(x_1)\\ &= (\mu\phi_1)\tilde{\alpha}(x_1) + R(\mu\phi_1, f_0, \dots, f_t) - \langle\phi_1\rangle(\mu)\tilde{\alpha}(x_1)\\ &= \tilde{\alpha}(P(f_0, \dots, f_t)) - \tilde{\alpha}(\langle\phi_1\rangle(\mu)x_1) \in \mathrm{image}(\tilde{\alpha}).\end{aligned}$$

□

Corollary 5.2 *The graded ring isomorphism of Theorem 5.1 induces an isomorphism*

$$\bigoplus_{n\in\mathbb{Z}} K_n^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/J_U \to \bigoplus_{n\in\mathbb{Z}} H_{\mathrm{MW}}^{n,n}(U, \mathbb{Z}).$$

Proof Notice that the ideal J_U of Theorem 5.1 is homogeneous, and it follows that $\bigoplus_{n\in\mathbb{Z}} H_{\mathrm{MW}}^{n,n}(U, \mathbb{Z})$ can be computed as $H_{\mathrm{MW}}^{*,*}(K)\{\mathbb{G}_m(U)\}/J_U$, where $H_{\mathrm{MW}}^{*,*}(K)$ is the diagonal of $H_{\mathrm{MW}}(K)$. □

6 Combinatorial description

In this section, we fix an affine space $V = \mathbb{A}_K^N$, a family of hyperplanes I and we set $U := U_I^N$. We let $Q(U)$ be the cokernel of the group homomorphism $\mathbb{G}_m(K) \to \mathbb{G}_m(U)$, and we observe that the divisor map

$$\mathbb{G}_m(U) \xrightarrow{\mathrm{div}} \bigoplus_{Y_i\in I} \mathbb{Z}\cdot Y_i$$

in $\mathbb{A}_K^N$ induces an isomorphism $Q(U) \cong \bigoplus_{Y_i\in I} \mathbb{Z}\cdot Y_i$. We consider the exterior algebra $\Lambda_{\mathbb{Z}}Q(U)$ and write $\Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U) := \mathbb{Z}[\eta]/2\eta \otimes_{\mathbb{Z}} \Lambda_{\mathbb{Z}}Q(U)$. The abelian group $Q(U)$ being free, the $\mathbb{Z}[\eta]/2\eta$–module $\Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U)$ is also free, with usual basis. To provide a combinatorial description of $H_{\mathrm{MW}}(U)$, we will have to slightly modify the definition of the divisor map above, in order to incorporate the action of η. We then define a map

$$\mathbb{G}_m(U) \xrightarrow{\widetilde{\mathrm{div}}} \Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U)$$

as follows:

(1) If $f = \lambda\phi$ or $f = \lambda$, where $\lambda \in \mathbb{G}_m(K)$ and ϕ is a linear polynomial as above, then $\widetilde{\mathrm{div}}(f) = \mathrm{div}(f)$.

(2) If $f, g \in \mathbb{G}_m(U)$, then $\widetilde{\mathrm{div}}(fg) = \widetilde{\mathrm{div}}(f) + \widetilde{\mathrm{div}}(g) + \eta \cdot \widetilde{\mathrm{div}}(f) \wedge \widetilde{\mathrm{div}}(g)$.

Lemma 6.1 *The map* $\widetilde{\mathrm{div}}$ *is well defined.*

Proof We first notice that $\widetilde{\mathrm{div}}(fg) = \widetilde{\mathrm{div}}(gf)$, since

$$\widetilde{\mathrm{div}}(fg)-\widetilde{\mathrm{div}}(gf) = \eta\cdot\widetilde{\mathrm{div}}(f)\wedge\widetilde{\mathrm{div}}(g)-\eta\cdot\widetilde{\mathrm{div}}(g)\wedge\widetilde{\mathrm{div}}(f) = 2\eta\cdot\widetilde{\mathrm{div}}(f)\wedge\widetilde{\mathrm{div}}(g) = 0.$$

Let $f_1, f_2, g_1, g_2 \in \mathbb{G}_m(U)$ be such that $f_1 g_1 = f_2 g_2$. Let $Y \in I$ be such that $f_i = Y^{n_i} \cdot f_i'$ with $\mathrm{div}_Y(f_i') = 0$ and $g_i = Y^{m_i} \cdot g_i'$ with $\mathrm{div}_Y(g_i') = 0$ for $i = 1, 2$ and $m_i, n_i \in \mathbb{Z}$. We get

$$\widetilde{\mathrm{div}}(f_1 g_1) = (m_1 + n_1) \cdot Y + \widetilde{\mathrm{div}}(f_1' g_1') + (m_1 + n_1)\eta(Y \wedge \widetilde{\mathrm{div}}(f_1' g_1')),$$
$$\widetilde{\mathrm{div}}(f_2 g_2) = (m_2 + n_2) \cdot Y + \widetilde{\mathrm{div}}(f_2' g_2') + (m_2 + n_2)\eta(Y \wedge \widetilde{\mathrm{div}}(f_2' g_2')).$$

As $\Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U)$ is free with usual basis, we deduce that $\widetilde{\mathrm{div}}(f_2' g_2') = \widetilde{\mathrm{div}}(f_1' g_1')$, which allows us to conclude by induction on the number of nontrivial factors in the decomposition of $f_1 g_1$. □

Now let $L_U \subset \Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U)$ be the ideal generated by the elements

(1) $Y_1 \wedge \cdots \wedge Y_s$ for $Y_i \in I$ such that $Y_1 \cap \cdots \cap Y_s = \varnothing$;

(2) $\sum_{j=1}^{s}(-1)^k Y_1 \wedge \cdots \wedge \widehat{Y}_j \wedge \cdots \wedge Y_s$ for $Y_i \in I$ such that $Y_1 \cap \cdots \cap Y_s \neq \varnothing$ and $\mathrm{codim}(Y_1 \cap \cdots \cap Y_s) < s$.

As a consequence of Lemma 6.1, the map $\widetilde{\mathrm{div}}$ induces a morphism of $\mathbb{Z}[\eta]/2\eta$–algebras

$$\psi : (\mathbb{Z}[\eta]/2\eta)\{\mathbb{G}_m(U)\} \to \Lambda_{\mathbb{Z}[\eta]/2\eta}Q(U)/L_U.$$

It is now time to introduce the ring

$$A_0(U) := K_*^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}/\bigl(J_U + K^\times \cdot K_*^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}\bigr).$$

As $\epsilon = -1 - [-1]\eta \sim -1$ in $A_0(U)$, it follows that $A_0(U)$ is an exterior algebra. Moreover, the coefficient ring $K_*^{\mathrm{MW}}(K)$ can be reduced to $K_*^{\mathrm{MW}}(K)/(K^\times \cdot K_*^{\mathrm{MW}}(K)) \cong \mathbb{Z}[\eta]/2\eta$.

Proposition 6.2 *The morphism of $\mathbb{Z}[\eta]/2\eta$–algebras*

$$\psi\colon \mathbb{Z}[\eta]/2\eta\{\mathbb{G}_m(U)\} \to \Lambda_{\mathbb{Z}[\eta]/2\eta} Q(U)/L_U$$

induces an isomorphism

$$\Psi\colon A_0(U) \to \Lambda_{\mathbb{Z}[\eta]/2\eta} Q(U)/L_U.$$

Proof We first prove that Ψ is well defined, which amounts to showing that the image of J_U is contained in L_U. For $f \in K^\times$, we have $[f] \in K^\times \cdot K_*^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$ and $\widetilde{\mathrm{div}}(f) = 0$, showing that the first relation is satisfied. The second relation is satisfied by definition of $\widetilde{\mathrm{div}}$, while relation ($3'$) is satisfied as $\Lambda_{\mathbb{Z}[\eta]/2\eta} Q(U)/L_U$ is an exterior algebra. As in the proof of Theorem 5.1, we are then left with elements of J'_U, ie elements of the form $R(f_0, \dots, f_t)$ for $\sum_{i=0}^t f_i = 0$, where $f_j = \lambda_j \phi_j$ or $f_j = \lambda_j$. Modulo $K^\times \cdot K_*^{\mathrm{MW}}(K)\{\mathbb{G}_m(U)\}$, we have $R(f_0, \dots, f_t) \sim \sum_{i=0}^t (-1)^{t+i} [f_0] \cdots \widehat{[f_i]} \cdots [f_t]$ and we just need to prove that

$$\alpha := (-1)^t \psi(R(f_0, \dots, f_t)) = \sum_{i=0}^t (-1)^i \widetilde{\mathrm{div}}(f_0) \wedge \cdots \wedge \widehat{\widetilde{\mathrm{div}}(f_i)} \wedge \cdots \wedge \widetilde{\mathrm{div}}(f_t)$$

is an element of L_U. Note that, if there are more than two constant functions among the f_j, α would be trivial. Suppose that $f_0 = \lambda_0$ is the only constant, and let $f_j = \lambda_j \phi_j$ with kernel $Y_j \in I$, so that $\alpha = Y_1 \wedge \cdots \wedge Y_t$. Since $\sum_{j=1}^t \lambda_j \phi_j = -\lambda_0 \neq 0$, we can easily get that $Y_1 \cap \cdots \cap Y_t = \varnothing$ and $\alpha = Y_1 \wedge \cdots \wedge Y_t \in L_U$. In the case where none of the f_j is constant, $\alpha = \sum_{i=0}^t (-1)^i Y_0 \wedge \cdots \wedge \widehat{Y_i} \wedge \cdots \wedge Y_t$. And, for every i, we have $\sum_{j=0, j\neq i}^t \lambda_j \phi_j = -\lambda_i \phi_i$, which means $Y_i \subseteq Y_0 \cap \cdots \cap \widehat{Y_i} \cap \cdots \cap Y_t = Y_0 \cap \cdots \cap Y_t$. If $Y_0 \cap \cdots \cap Y_t = \varnothing$, so is $Y_0 \cap \cdots \cap \widehat{Y_i} \cap \cdots \cap Y_t$, thus $Y_0 \wedge \cdots \wedge \widehat{Y_i} \wedge \cdots \wedge Y_t \in L_U$; otherwise, $\mathrm{codim}(Y_0 \cap \cdots \cap Y_t) = \mathrm{codim}(Y_0 \cap \cdots \cap \widehat{Y_i} \cap \cdots \cap Y_t) \leq t < t+1$, which just fits the condition (2) of L_U. This proves that Ψ is well defined.

To prove that Ψ is an isomorphism, we construct the inverse map by

$$\Phi\colon \Lambda_{\mathbb{Z}[\eta]/2\eta} Q(U)/L_U \to A_0(U),\ Y_i \mapsto (\phi_i)$$

and prove that it is well defined. As above, we just need to discuss elements of L_U. If $Y_1 \cap \cdots \cap Y_s = \varnothing$, then we can find $\lambda_i \in K^\times$ such that $\sum_i \lambda_i \phi_i = 1$, and thus $(\phi_1) \cdots (\phi_s) \sim (\lambda_1 \phi_1) \cdots (\lambda_s \phi_s) = 0$ in $A_0(U)$. In the case $\mathrm{codim}(Y_1 \cap \cdots \cap Y_s) < s$, we have $\sum_i \lambda_i \phi_i = 0$ for some $\lambda_i \in K^\times$. Then $\sum_{i=1}^s (-1)^i (\phi_1) \cdots \widehat{(\phi_i)} \cdots (\phi_s) \sim (-1)^{s-1} R(\lambda_1 \phi_1, \dots, \lambda_s \phi_s) = 0$ in $A_0(U)$. This shows that the inverse map is well defined. □

The following corollary shows that the rank of the free $H_{\mathrm{MW}}(K)$–module $H_{\mathrm{MW}}(U)$ is exactly the same as the rank of the free $H_M(K)$–module $H_M(U)$ [4, Proposition 3.11]:

Corollary 6.3 *The rank of the free $H_{\mathrm{MW}}(K)$–module $H_{\mathrm{MW}}(U)$ is equal to the rank of the free module $\Lambda_{\mathbb{Z}} Q(U)/L_U$.*

Proof It is clear that $\mathrm{rk}_{\mathbb{Z}[\eta]/2\eta}(\Lambda_{\mathbb{Z}[\eta]/2\eta} Q(U)/L_U) = \mathrm{rk}_{\mathbb{Z}}(\Lambda_{\mathbb{Z}} Q(U)/L_U)$. As all generators in $H_{\mathrm{MW}}(U)$ are from $H^{p,p}_{\mathrm{MW}}(U,\mathbb{Z})$, we have

$$\mathrm{rk}_{H_{\mathrm{MW}}(K)}(H_{\mathrm{MW}}(U)) = \mathrm{rk}_{K^{\mathrm{MW}}_*(K)}\left(\bigoplus_{n\in\mathbb{Z}} H^{n,n}_{\mathrm{MW}}(U,\mathbb{Z})\right) = \mathrm{rk}_{\mathbb{Z}[\eta]/2\eta}(A_0(U)). \qquad \square$$

7 I–cohomology and singular cohomology

In ordinary motivic cohomology theory, we have a realization functor to the topological cohomology of complex points. This yields the following comparative result:

Proposition 7.1 [4, Proposition 3.9] *In the case $K=\mathbb{C}$, there is an isomorphism of rings*

$$\bigoplus_n H^{n,n}_M(U,\mathbb{Q}) \otimes_{H_M(K)} K^M_*(K)/K^\times \cdot K^M_*(K) \xrightarrow{\cong} \bigoplus_n H^n_{\mathrm{sing}}(U(\mathbb{C}),\mathbb{Q}).$$

In this section, we provide an analogue for the singular cohomology of the real points of the complement of a hyperplane arrangement defined over $\mathbb{R}$. We start with some results about the $\boldsymbol{I}$–cohomology [9].

As recalled in Section 2, we have natural homomorphisms from Milnor–Witt motivic cohomology to $\boldsymbol{I}^*$–cohomology

$$H^{p,q}_{\mathrm{MW}}(X,\mathbb{Z}) \to H^{p-q}(X,\boldsymbol{K}^{\mathrm{MW}}_q) \to H^{p-q}(X,\boldsymbol{I}^q)$$

which induce a ring homomorphism $H_{\mathrm{MW}}(X) \to \bigoplus_{r,q} H^r(X,\boldsymbol{I}^q)$ (where $\boldsymbol{I}^q = \boldsymbol{K}^{\mathrm{MW}}_q = \boldsymbol{W}$ for $q<0$). In case $X=\mathrm{Spec}(K)$, we obtain in particular a ring homomorphism $H_{\mathrm{MW}}(K) \to \bigoplus_{r,q} H^r(K,\boldsymbol{I}^q) = \bigoplus_{q\in\mathbb{Z}} I^q(K)$.

Proposition 7.2 *The morphism of $\bigoplus_{q\in\mathbb{Z}} I^q(K)$–algebras*

$$j : H_{\mathrm{MW}}(U) \otimes_{H_{\mathrm{MW}}(K)} \left(\bigoplus_{q\in\mathbb{Z}} I^q(K)\right) \to \bigoplus_{r,q} H^r(U,\boldsymbol{I}^q)$$

is an isomorphism. Moreover, $H^r(U,\boldsymbol{I}^q)=0$ for $r\neq 0$.

Proof We write $H_{\mathrm{MW}}(U)\otimes \boldsymbol{I}$ for the graded ring $H_{\mathrm{MW}}(U)\otimes_{H_{\mathrm{MW}}(K)}\big(\bigoplus_q I^q(K)\big)$. We follow the same induction process as in the proof of the main theorem. When $|I|=0$, we only need to consider $\mathrm{Spec}(K)$ by homotopy invariance, and the result is trivial.

Assume now that $Y\in I$ and that we have isomorphisms for $U^V_{I'}$ and $U^Y_{I_Y}$. Notice that, for $\boldsymbol{I}$–cohomology, we still have a Gysin long exact sequence [9, remarque 9.3.5]. The proof of the main theorem yields the commutative diagram

$$\begin{tikzcd}
0 \arrow[d] & \bigoplus_q H^{-1}(U^Y_{I_Y},\boldsymbol{I}^{q-1}) \arrow[d] \\
H_{\mathrm{MW}}(U^V_{I'})\otimes \boldsymbol{I} \arrow[r, "\cong"] \arrow[d] & \bigoplus_q H^0(U^V_{I'},\boldsymbol{I}^q) \arrow[d] \\
H_{\mathrm{MW}}(U^V_{I})\otimes \boldsymbol{I} \arrow[r, "j"] \arrow[d] & \bigoplus_q H^0(U^V_{I},\boldsymbol{I}^q) \arrow[d] \\
H_{\mathrm{MW}}(U^Y_{I_Y})\otimes \boldsymbol{I} \arrow[r, "\cong"] \arrow[d] & \bigoplus_q H^0(U^Y_{I_Y},\boldsymbol{I}^{q-1}) \arrow[d] \\
0 & \bigoplus_q H^1(U^V_{I'},\boldsymbol{I}^q)
\end{tikzcd}$$

By our assumption, $H^{-1}(U^Y_{I_Y},\boldsymbol{I}^{q-1})$ and $H^1(U^V_{I'},\boldsymbol{I}^q)$ are both 0, so the right column is also short exact. We conclude that j is an isomorphism as well. The same argument implies that $H^r(U^V_I,\boldsymbol{I}^q)=0$ for $r\neq 0$. □

The analogue of Corollary 3.2 in this setting then reads as follows:

Corollary 7.3 *There is a finite set J and integers $n_j\geq 0$ for any $j\in J$ such that*

$$H^0(U^V_I,\boldsymbol{I}^q)\cong\bigoplus_{j\in J} I^{q-n_j}(K)b_j$$

as a free $\bigoplus_q I^q(K)$–module with basis elements $b_j\in H^0(U^V_I,\boldsymbol{I}^{n_j})$.

Proof Every step is the same as in Proposition 3.1, except the splitting, which comes from the identification with $H_{\mathrm{MW}}(U^V_I)\otimes\boldsymbol{I}$. □

As in [11; 10], we can compute the cohomology of the real spectrum using $\boldsymbol{I}$–cohomology.

Proposition 7.4 [10, Proposition 3.6] *The signature map induces an isomorphism*

$$H^r\big(X, \operatorname*{Colim}_{q\geq 0} \boldsymbol{I}^q\big) \xrightarrow{\text{sign}_\infty} H^r_{\text{sing}}(\text{Sper}(X), \mathbb{Z}),$$

where $\text{Sper}(X)$ *is the real spectrum. In particular*,

$$\operatorname*{Colim}_{q\geq 0} I^q(K) \cong H^0_{\text{sing}}(\text{Sper}(K), \mathbb{Z}).$$

In our case, since U is always noetherian and $\text{Colim}_{q\geq 0}$ is filtered, we have a canonical isomorphism

$$H^r\big(U, \operatorname*{Colim}_{q\geq 0} \boldsymbol{I}^q\big) \cong \operatorname*{Colim}_{q\geq 0} H^r(U, \boldsymbol{I}^q). \tag{7-1}$$

Combining with Corollary 7.3, we obtain the following proposition:

Proposition 7.5 *There exists an integer* $N > 0$ *such that*

$$H^0(U_I^V, \boldsymbol{I}^N) \otimes_{\bigoplus_{q\geq 0} I^q(K)} H^0_{\text{sing}}(\text{Sper}(K), \mathbb{Z}) \xrightarrow{2^{-N}\text{sign}} H^0_{\text{sing}}(\text{Sper}(U_I^V), \mathbb{Z})$$

is an isomorphism. Moreover, $H^r_{\text{sing}}(\text{Sper}(U_I^V), \mathbb{Z}) = 0$ *for* $r \neq 0$.

Proof By (7-1), we can rewrite the right-hand side as $\text{Colim}_{q\geq 0} H^0(U_I^V, \boldsymbol{I}^q)$. Applying Corollary 7.3, we get

$$\operatorname*{Colim}_{q\geq 0}\Big(\bigoplus_{j\in J} I^{q-n_j}(K)b_j\Big) \cong \bigoplus_{j\in J}\big(\operatorname*{Colim}_{q\geq 0} I^{q-n_j}(K)b_j\big) \cong \bigoplus_{j\in J} H^0_{\text{sing}}(\text{Sper}(K), \mathbb{Z})b_j.$$

Let $N \in \mathbb{N}$ be such that $N \geq n_j$ for all $j \in J$. Using again Corollary 7.3,

$$H^0(U_I^V, \boldsymbol{I}^N) \cong \bigoplus_{j\in J} I^{N-n_j}(K)b_j,$$

which implies

$$\bigoplus_{j\in J} I^{N-n_j}(K)b_j \otimes_{\bigoplus_{q\geq 0} I^q(K)} H^0_{\text{sing}}(\text{Sper}(K), \mathbb{Z}) \cong \bigoplus_{j\in J} H^0_{\text{sing}}(\text{Sper}(K), \mathbb{Z})b_j$$

since, for every j, we have $N - n_j \geq 0$. That proves the first part, while the second part is trivial. □

Taking $K = \mathbb{R}$, we have $H^0_{\text{sing}}(\mathbb{R}, \mathbb{Z}) = \mathbb{Z}$ and we recover the classical result for complements of hyperplane arrangements

$$H^0(U_I^V, \boldsymbol{I}^N) \xrightarrow[\cong]{2^{-N}\text{sign}} H^0_{\text{sing}}(U_I^V(\mathbb{R}), \mathbb{Z}) \cong \bigoplus_{R_i \in \text{connected components}} \mathbb{Z}\{R_i\}.$$

References

[1] **T Bachmann**, **B Calmès**, **F Déglise**, **J Fasel**, **P A Østvær**, *Milnor–Witt motives*, preprint (2020) arXiv 2004.06634

[2] **J Barge**, **F Morel**, *Groupe de Chow des cycles orientés et classe d'Euler des fibrés vectoriels*, C. R. Acad. Sci. Paris Sér. I Math. 330 (2000) 287–290 MR Zbl

[3] **B Calmès**, **J Fasel**, *A comparison theorem for MW-motivic cohomology*, preprint (2017) arXiv 1708.06100

[4] **A Chatzistamatiou**, *Motivic cohomology of the complement of hyperplane arrangements*, Duke Math. J. 138 (2007) 375–389 MR Zbl

[5] **D-C Cisinski**, **F Déglise**, *Triangulated categories of mixed motives*, Springer (2019) MR Zbl

[6] **F Déglise**, **J Fasel**, *The Milnor–Witt motivic ring spectrum and its associated theories*, preprint (2017) arXiv 1708.06102

[7] **F Déglise**, **J Fasel**, *MW-motivic complexes*, preprint (2017) arXiv 1708.06095

[8] **P H Edelman**, *A partial order on the regions of $\mathbb{R}^n$ dissected by hyperplanes*, Trans. Amer. Math. Soc. 283 (1984) 617–631 MR Zbl

[9] **J Fasel**, *Groupes de Chow–Witt*, Mém. Soc. Math. Fr. 113, Soc. Math. France, Paris (2008) MR Zbl

[10] **J Hornbostel**, **M Wendt**, **H Xie**, **M Zibrowius**, *The real cycle class map*, Ann. K–Theory 6 (2021) 239–317 MR Zbl

[11] **J A Jacobson**, *Real cohomology and the powers of the fundamental ideal in the Witt ring*, Ann. K–Theory 2 (2017) 357–385 MR Zbl

[12] **F Morel**, *$\mathbb{A}^1$–algebraic topology over a field*, Lecture Notes in Math. 2052, Springer (2012) MR Zbl

[13] **F Morel**, **V Voevodsky**, *$\mathbb{A}^1$–homotopy theory of schemes*, Inst. Hautes Études Sci. Publ. Math. 90 (1999) 45–143 MR Zbl

[14] **N Yang**, *Projective bundle theorem in MW-motivic cohomology*, Doc. Math. 26 (2021) 1045–1083 MR Zbl

[15] **N Yang**, *Quaternionic projective bundle theorem and Gysin triangle in MW-motivic cohomology*, Manuscripta Math. 164 (2021) 39–65 MR Zbl

Institut Fourier, Université Grenoble Alpes
Grenoble, France

keyao.peng@univ-grenoble-alpes.fr

Received: 19 November 2020 Revised: 12 December 2021

Algebraic & Geometric Topology **23**:8 (2023) 3553–3586
DOI: 10.2140/agt.2023.23.3553
Published: 5 November 2023

Connective models for topological modular forms of level *n*

LENNART MEIER

We construct and study connective versions of topological modular forms of higher level like $\mathrm{tmf}_1(n)$. In particular, we use them to realize Hirzebruch's level-n genus as a map of ring spectra.

55N34; 55N22, 55P91

1 Introduction

The basic tenet of Waldhausen's philosophy of *brave new algebra* is to replace known notions for commutative rings by corresponding notions for E_∞–ring spectra. These days replacing the integers by the sphere spectrum is no longer so brave and new, but rather a well-established principle. In extension, we might want to find and study E_∞–analogues of other prominent rings as well. The aim of the present paper is to do this for rings of holomorphic modular forms with respect to congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Topological analogues of modular forms for $\mathrm{SL}_2(\mathbb{Z})$ itself were already introduced about twenty years ago. Indeed, Goerss, Hopkins and Miller introduced three spectra TMF, Tmf and tmf of topological modular forms. Recall that the rings $M_*(\mathrm{SL}_2(\mathbb{Z}); \mathbb{Z})$

and $\widetilde{M}_*(\mathrm{SL}_2(\mathbb{Z});\mathbb{Z})$ of holomorphic and meromorphic integral modular forms can be defined as the global sections $H^0(\overline{\mathcal{M}}_{\mathrm{ell}};\omega^{\otimes *})$ and $H^0(\mathcal{M}_{\mathrm{ell}};\omega^{\otimes *})$ of powers of a certain line bundle ω on the compactified and uncompactified moduli stack of elliptic curves, respectively.[1] In analogy, TMF is defined as the global sections of a sheaf $\mathcal{O}^{\mathrm{top}}$ of E_∞–ring spectra on $\mathcal{M}_{\mathrm{ell}}$ with $\pi_{2k}\mathcal{O}^{\mathrm{top}} \cong \omega^{\otimes k}$ and Tmf as the global sections of an analogous sheaf on $\overline{\mathcal{M}}_{\mathrm{ell}}$. The edge maps of the resulting descent spectral sequences take the form of homomorphisms

$$\pi_{2*}\mathrm{TMF} \to \widetilde{M}_*(\mathrm{SL}_2(\mathbb{Z});\mathbb{Z}) \quad \text{and} \quad \pi_{2*}\mathrm{Tmf} \to M_*(\mathrm{SL}_2(\mathbb{Z});\mathbb{Z}).$$

The former morphism is an isomorphism after base change to $\mathbb{Z}\bigl[\frac{1}{6}\bigr]$ (while taking higher cohomology of $\omega^{\otimes *}$ into account at the primes 2 and 3) and thus TMF can be really seen as the rightful analogue of $\widetilde{M}(\mathrm{SL}_2(\mathbb{Z});\mathbb{Z})$. In contrast, $\pi_*\mathrm{Tmf}$ has torsion-free summands in negative degree, whereas $M_*(\mathrm{SL}_2(\mathbb{Z}),\mathbb{Z})$ is concentrated in nonnegative degrees. The solution is to define tmf simply as the connective cover $\tau_{\geq 0}\mathrm{Tmf}$, and one can show that indeed $\pi_{2*}\mathrm{tmf}\bigl[\frac{1}{6}\bigr]$ is isomorphic to $M_*(\mathrm{SL}_2(\mathbb{Z}),\mathbb{Z}\bigl[\frac{1}{6}\bigr])$. We mention that one of the motivations for constructing tmf was lifting the Witten genus to a map of E_∞–ring spectra $M\mathrm{String} \to \mathrm{tmf}$ as achieved in Ando, Hopkins and Rezk [1]. For applications to the stable homotopy groups of spheres and exotic spheres, see for instance Hopkins and Mahowald [23], Behrens, Hill, Hopkins and Mahowald [3], Wang and Xu [46] and Isaksen, Wang and Xu [25].

In number theory, it is very common not only to consider modular forms with respect to $\mathrm{SL}_2(\mathbb{Z})$, but also to congruence subgroups of these; the most important being $\Gamma = \Gamma_0(n)$, $\Gamma_1(n)$ or $\Gamma(n)$. Algebrogeometrically, such modular forms can be defined as sections of the pullback of $\omega^{\otimes *}$ to compactifications $\overline{\mathcal{M}}(\Gamma)$ of stacks classifying generalized elliptic curves with certain level structures (see eg Deligne and Rapoport [6], Diamond and Im [7], Conrad [5] and the author's [36]); for example, $\overline{\mathcal{M}}(\Gamma_1(n))$ classifies generalized elliptic curves with a chosen point of order n whose multiples intersect every irreducible component of every geometric fiber. Hill and Lawson [17] defined sheaves of E_∞–ring spectra on these stacks and obtained spectra $\mathrm{Tmf}(\Gamma)$, as their global sections, and $\mathrm{TMF}(\Gamma)$, by restriction to the loci of smooth elliptic curves. The latter spectra are good topological analogues of the rings $\widetilde{M}(\Gamma;\mathbb{Z}[1/n])$ of meromorphic modular forms in

[1]The terms *meromorphic* and *holomorphic* come from the corresponding analytic definitions, where one demands that the given function on the upper half-plane can be continued meromorphically and holomorphically, respectively, to the cusp(s). The former kind of modular form is also sometimes called *weakly holomorphic*.

the sense that $\pi_*\mathrm{TMF}(\Gamma)$ is isomorphic to this ring if Γ is $\Gamma_1(n)$ or $\Gamma(n)$ (with $n \geq 2$) and, if we invert 6, also in the case $\Gamma = \Gamma_0(n)$.

In contrast, neither $\mathrm{Tmf}(\Gamma)$ nor its connective cover $\tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ are in general good analogues of the ring of holomorphic modular forms $M(\Gamma; \mathbb{Z}[1/n])$, even in the nice case of $\Gamma = \Gamma_1(n)$ and $n \geq 2$. Writing $\mathrm{Tmf}_1(n)$ for $\mathrm{Tmf}(\Gamma_1(n))$, the reason is that $H^1(\overline{\mathcal{M}}(\Gamma_1(n)); \omega)$ and thus $\pi_1\mathrm{Tmf}_1(n)$ is nontrivial in general (with $n = 23$ being the first example), while this contribution does not occur in $M(\Gamma; \mathbb{Z}[1/n])$. Following an idea of Lawson, we define a connective version $\mathrm{tmf}_1(n)$ by "artificially" removing π_1, while still retaining the E_∞–structure on $\mathrm{tmf}_1(n)$. The following will be proven as Theorems 2.12 and 2.22.

Theorem 1.1 *There is an essentially unique connective E_∞–ring spectrum $\mathrm{tmf}_1(n)$ with an E_∞–ring map $\mathrm{tmf}_1(n) \to \mathrm{Tmf}_1(n)$ that identifies the homotopy groups of the source with $M(\Gamma_1(n); \mathbb{Z}[1/n])$.*

Moreover, the involution of $\overline{\mathcal{M}}(\Gamma_1(n))$ sending a point of order n on the universal elliptic curve to its negative defines on $\mathrm{tmf}_1(n)$ the structure of a genuine C_2–spectrum. Its slices in the sense of Hill, Hopkins and Ravenel [16] are trivial in odd degrees and can be explicitly identified in even degrees.

The analogous theorem also works to define $\mathrm{tmf}(n)$, but $\mathrm{tmf}_0(n)$ we define only in certain cases since in the general case it is not yet clear what the "correct" definition is. The spectrum $\mathrm{tmf}(n)$ has been further investigated in [21, Theorem 3.14], where a criterion for the nonvanishing of its Tate spectrum is proven.

One of the principal motivations for the consideration of $\mathrm{tmf}_1(n)$ is its connection to the Hirzebruch level-n genera $\mathrm{MU}_* \to M(\Gamma_1(n); \mathbb{Z}[1/n])$. They specialize for $n = 2$ to the classic Ochanine elliptic genus and have similar rigidity properties in general; see Hirzebruch, Berger and Jung [20]. We will prove the following as Theorem 3.6.

Theorem 1.2 *For every $n \geq 2$, there is a ring map $\mathrm{MU} \to \mathrm{tmf}_1(n)$ realizing on homotopy groups the Hirzebruch level-n–genus. Moreover, this map refines to a map $\mathrm{MU}_{\mathbb{R}} \to \mathrm{tmf}_1(n)$ of C_2–spectra.*

We have two further classes of results on the spectra $\mathrm{tmf}_1(n)$ and their cousins. The first is the following compactness result, contained in Theorem 4.4 and Corollary 4.6.

Theorem 1.3 *The* $\mathrm{tmf}[1/n]$*–modules* $\mathrm{tmf}_0(n)$, $\mathrm{tmf}_1(n)$ *and* $\mathrm{tmf}(n)$ *are perfect*, *ie they are compact objects in the module category*, *in the cases they are defined. In particular*, *their* $\mathbb{F}_p$*–cohomologies are finitely presented over the Steenrod algebra and thus their* p*–completions are fp-spectra in the sense of Mahowald and Rezk* [33].

By a result of Kuhn [28, Theorem 1.7] this implies, for example, that the Hurewicz image of $\pi_*\mathrm{tmf}(\Gamma) \cong \pi_*\Omega^\infty\mathrm{tmf}(\Gamma)$ in $H_*(\Omega^\infty\mathrm{tmf}(\Gamma); \mathbb{F}_p)$ is finite-dimensional, where $\mathrm{tmf}(\Gamma)$ denotes either $\mathrm{tmf}_0(n)$, $\mathrm{tmf}_1(n)$ or $\mathrm{tmf}(n)$. We also note that in contrast to the theorem, $\mathrm{tmf}_1(n)$ will not be a perfect $\mathrm{tmf}_0(n)$–module in general. We also show that $\mathrm{tmf}_0(n)$, $\mathrm{tmf}_1(n)$ and $\mathrm{tmf}(n)$ are faithful as $\mathrm{tmf}[1/n]$–modules, answering a question of Höning and Richter [21, page 21].

The second result is a variant of the decomposition results of the author [37], which we state in this introduction only at the prime 2 and for $\mathrm{tmf}_1(n)$, and which will be proven as Theorem 5.6.

Theorem 1.4 *Let* $n > 1$ *be odd. If one can lift every weight-*1 *modular form for* $\Gamma_1(n)$ *over* $\mathbb{F}_2$ *to a form of the same weight and level over* $\mathbb{Z}_{(2)}$, *we have a* C_2*–equivariant splitting*

$$\mathrm{tmf}_1(n)_{(2)} \simeq \bigoplus_i \Sigma^{n_i\rho}\mathrm{tmf}_1(3)_{(2)},$$

where ρ *denotes the real regular representation of* C_2.

In the author's earlier work [36, Appendix C], it is shown that for $1 < n < 65$ odd indeed every weight-1 modular form for $\Gamma_1(n)$ over $\mathbb{F}_2$ lifts to a form of the same weight and level over $\mathbb{Z}_{(2)}$, while for $n = 65$ it does not. See also [36, Remark 3.14] for a further discussion of this condition.

Conventions and notation

All notions are to be understood suitably derived or ∞–categorical. This means that pushout means either a pushout in the respective ∞–category or a homotopy pushout in the underlying model category. We will use $\otimes$ for the (derived) smash product. Note that this coincides with the coproduct in the ∞–category CAlg of E_∞–ring spectra.

When we use G–spectra, we will always mean genuine G–spectra. The notations $\tau_{\le k}$ and $\tau_{\ge k}$ denote the k–(co)connective cover of a spectrum and we use the same notation for the slice-(co)connective covers of a G–spectrum. Furthermore, we denote by $\mathbb{S}$

the sphere (G–)spectrum. In some parts of this article, we have the opportunity to use $\mathrm{RO}(C_2)$–graded homotopy groups of C_2–spectra. We will use the notation σ for the sign representation and ρ or $\mathbb{C}$ for the regular representation of C_2.

We will use the notations $\mathrm{TMF}_1(n)$ and $\mathrm{TMF}(\Gamma_1(n))$ interchangeably and similarly in related contexts.

Acknowledgements

I want to thank Tyler Lawson for explaining to me the idea of how to construct a connective model for $\mathrm{TMF}_1(n)$, and for the sketch of an argument that $\mathrm{tmf}_1(3)$ is not perfect as a $\mathrm{tmf}_0(3)$–module. It is also a pleasure to acknowledge the influence and encouragement of Mike Hill. Furthermore I want to thank Eva Höning and Birgit Richter for their interest and remarks on a preliminary version, and the referee for their extensive comments.

Finally, I want to thank the Hausdorff Institute for hospitality in 2015 when part of this work was undertaken. Apologies for the subsequent delay in publication.

2 The construction of connective topological modular forms

The aim of this section is to construct connective spectra $\mathrm{tmf}(\Gamma)$ of topological modular forms and thereby prove Theorem 1.1. Here Γ denotes a congruence subgroup Γ in the following sense, which is a bit more restrictive than the standard definition.

Definition 2.1 We call $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ a *congruence subgroup of level n* if $\Gamma = \Gamma(n)$ or $\Gamma_1(n) \subset \Gamma \subset \Gamma_0(n)$.[2]

As explained in [17; 37, Section 2.1], we can associate with every such Γ a (nonconnective and nonperiodic) E_∞–ring spectrum $\mathrm{Tmf}(\Gamma)$. (See also [44, Theorem 5.2] for the case of $\Gamma(n)$.) These arise as global sections of sheaves of E_∞–ring spectra $\mathcal{O}^{\mathrm{top}}$ on stacks $\overline{\mathcal{M}}(\Gamma)$ classifying generalized elliptic curves with certain level structures; the details will not be important for the purposes of this article, but see for instance [6; 5; 45; 36]. Our goal in this section is to construct a nice connective version $\mathrm{tmf}(\Gamma)$ for $\mathrm{Tmf}(\Gamma)$. For this, we will fix a localization $\mathbb{Z}_S$ of the integers and restrict mostly to tame congruence subgroups.

[2]We refer to [8] for background on the congruence subgroups $\Gamma_1(n)$, $\Gamma(n)$ and $\Gamma_0(n)$ and their relationship to moduli of elliptic curves. This material though is barely necessary for the present paper, as we use the congruence subgroups primarily as notation.

Definition 2.2 We say that a congruence subgroup Γ of level n is *tame* with respect to $\mathbb{Z}_S$ if $n \geq 2$ and n is invertible in $\mathbb{Z}_S$; in the case $\Gamma_1(n) \subset \Gamma \subset \Gamma_0(n)$ we demand additionally that $\gcd(6, [\Gamma : \Gamma_1(n)])$ is invertible in $\mathbb{Z}_S$.[3]

The definition ensures that the order of every automorphism of a point in $\overline{\mathcal{M}}(\Gamma)$ is invertible and thus the stack is of cohomological dimension one. As explained in [37, Section 2.1], in this case $\pi_*\tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is concentrated in even degrees except for $\pi_1\mathrm{Tmf}(\Gamma)$, which might be nonzero. (The smallest n for which this happens is 23.) Moreover, the even homotopy groups of $\mathrm{Tmf}(\Gamma)$ are precisely isomorphic to the ring of holomorphic modular forms $M(\Gamma; \mathbb{Z}[1/n])$.

Following the lead of [29, Proposition 11.1] (and additional explanations by its author), we will first describe a general procedure to kill π_1 for E_∞–rings that applies to $\tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ for Γ tame. We will then present a C_2–equivariant refinement that helps to define a nice version of $\mathrm{tmf}(\Gamma)$ also in some nontame cases; see Construction 2.24. We note that our techniques are only necessary if $\pi_1\mathrm{Tmf}(\Gamma)$ is nontrivial as otherwise the usual connective cover defines a perfectly good version of $\mathrm{tmf}(\Gamma)$.

2.1 The nonequivariant argument

Let R be a connective E_∞–ring spectrum with $\pi_0 R$ an étale extension of $\mathbb{Z}_S$, a localization of $\mathbb{Z}$, and $\eta \cdot 1 = 0$; here, $\eta \in \pi_1\mathbb{S}$ is the Hopf element and $1 \in \pi_0 R$ the unit. (The relevant example for us is $R = \tau_{\geq 0}\mathrm{Tmf}(\Gamma)_S$ with $\pi_0 R = \mathbb{Z}_S$ if $\Gamma_1(n) \subset \Gamma \subset \Gamma_0(n)$ and $\pi_0 R = \mathbb{Z}_S[\zeta_n]$ if $\Gamma = \Gamma(n)$.) We want to construct a map $R' \to R$ of E_∞–ring spectra which is injective on π_* and with cokernel $\pi_1 R$. In the following, we localize everything implicitly at the set S — so $\mathbb{Z}$ really means $\mathbb{Z}_S$, etc.

Let A first be a general E_∞–ring spectrum. For an A–module M, we denote by

$$\mathbb{P}_A(M) \simeq A \oplus M \oplus (M^{\otimes_A 2})_{h\Sigma_2} \oplus \cdots$$

the free unital E_∞–A–algebra on M; cf [32, Example 3.1.3.14].

Definition 2.3 Let $x\colon \Sigma^k A \to A$ be an A–linear map. We define its *E_∞–cone* $C^A(x)$ as the pushout $A \otimes_{\mathbb{P}_A(\Sigma^k A)} A$ of E_∞–ring spectra. Here, the first map $\mathbb{P}_A(\Sigma^k A) \to A$ is the free E_∞–map on x, while the second arises from applying $\mathbb{P}_A$ to the unique map $\Sigma^k A \to 0$.

[3] As the quotient $\Gamma_0(n)/\Gamma_1(n)$ is $(\mathbb{Z}/n)^\times$, the latter condition reduces to $\gcd(6, \varphi(n))$ being invertible in the case $\Gamma = \Gamma_0(n)$. Thus we require that 2 is invertible and also 3 if n is divisible by a prime of the form $3k+1$ or by 9.

Note that if B is an E_∞–A–algebra, we have $C^A(x)\otimes_A B \simeq C^B(x)$. Writing the usual cone $C(x)$ as the pushout $A\sqcup_{\Sigma^k A\oplus A} A$ in A–modules produces a map $C(x)\to C^A(x)$ via the inclusion $A\oplus\Sigma^k A\to \mathbb{P}^A(\Sigma^k A)$ of the first two summands and the identity id_A.

Lemma 2.4 *If $x=0$, the canonical map $C(x)\to C^A(x)$ is split as a map of A–modules.*

Proof The pushout square

(2.5)
$$\begin{tikzcd} A\oplus\Sigma^k A \arrow[r] \arrow[d] & A \arrow[d] \\ A \arrow[r] & C(0)\simeq A\oplus\Sigma^{k+1}A \end{tikzcd}$$

arises from the pushout square

(2.6)
$$\begin{tikzcd} \Sigma^k A \arrow[r] \arrow[d] & 0 \arrow[d] \\ 0 \arrow[r] & \Sigma^{k+1}A \end{tikzcd}$$

via the functor $\mathrm{Mod}_A\to\mathrm{CAlg}_A$ of square-zero extension. In particular, it is a diagram of E_∞–A–algebras. As the E_∞–pushout square (P) defining $C^A(0)$ arises from (2.6) as well, but via $\mathbb{P}_A$, we see that the square (2.5) receives a map from the square (P). The resulting map $C^A(0)\to C(0)$ defines a splitting of $C(0)\to C^A(0)$ by the universal property of the pushout square (2.5). □

We will apply our general consideration to the connective E_∞–ring spectrum R we have fixed. As η is zero in π_*R, we obtain an E_∞–map $C^{\mathbb{S}}(\eta)\to R$. This induces an E_∞–map $\tau_{\le 1}C^{\mathbb{S}}(\eta)\to\tau_{\le 1}R$; see [16, Proposition 4.35].

Lemma 2.7 *The 1–coconnective cover $\tau_{\le 1}C^{\mathbb{S}}(\eta)$ is equivalent to $H\mathbb{Z}$.*

Proof We claim that the canonical map $C(\eta)\to C^{\mathbb{S}}(\eta)$ is 2–connected. By the Hurewicz theorem, we can test this after tensoring with $H\mathbb{Z}$ and thus it suffices to show that the resulting map $C(\eta\otimes H\mathbb{Z})\to C^{H\mathbb{Z}}(\eta\otimes H\mathbb{Z})$ is 2–connected. But $\eta\otimes H\mathbb{Z}$ agrees with the 0–map $\Sigma H\mathbb{Z}\to H\mathbb{Z}$. Thus, we have to show that

$$H\mathbb{Z}\oplus\Sigma^2 H\mathbb{Z}\to C^{H\mathbb{Z}}(\eta\otimes H\mathbb{Z})\simeq \mathbb{P}^{H\mathbb{Z}}\Sigma^2 H\mathbb{Z}\simeq H\mathbb{Z}\oplus\Sigma^2H\mathbb{Z}\oplus(\Sigma^4H\mathbb{Z})_{hC_2}\oplus\cdots$$

is 2–connected. As noted above, the map is split injective and thus must be indeed an isomorphism on π_i even for $i\le 3$. □

The inclusion of 1–truncated connective E_∞–ring spectra into all connective E_∞–ring spectra admits a left adjoint by [31, Proposition 5.5.6.18; 32, Proposition 7.1.3.14]. By [32, Proposition 7.1.3.15(3)], it agrees with $\tau_{\leq 1}$ on underlying spectra.

By [32, Theorem 7.5.0.6], we can extend the E_∞–map $H\mathbb{Z} = \tau_{\leq 1} C^{\mathbb{S}}(\eta) \to \tau_{\leq 1} R$ to an E_∞–map $H\pi_0 R \to \tau_{\leq 1} R$ since the map $\mathbb{Z} \to \pi_0 R$ is étale. Define now R' via the homotopy pullback square

(2.8)
$$\begin{tikzcd} R' \arrow[r] \arrow[d] & H\pi_0 R \arrow[d] \\ R \arrow[r] & \tau_{\leq 1} R \end{tikzcd}$$

This construction provides the existence part of the following proposition.

Proposition 2.9 *Let R be a connective E_∞–ring spectrum such that $\pi_0 R$ is an étale extension of a localization $\mathbb{Z}_S$ of the integers and $\eta \cdot 1 = 0$ in $\pi_1 R$. Then there exists a morphism $R' \to R$ of E_∞–ring spectra inducing an isomorphism on π_i for $i \neq 1$ and satisfying $\pi_1 R' = 0$. Moreover, for every other $R'' \to R$ with these properties, there is an equivalence $R'' \to R'$ of E_∞–ring spectra over R.*

Proof It remains to show uniqueness. We localize again everything implicitly at S. We first note that the map $H\mathbb{Z} \to \tau_{\leq 1} R$ constructed above is actually the unique E_∞–map with this source and target. Indeed, for connectivity reasons, we have an equivalence of mapping spaces $\mathrm{Map}_{\mathrm{CAlg}}(H\mathbb{Z}, \tau_{\leq 1} R) \simeq \mathrm{Map}_{\mathrm{CAlg}}(C^{\mathbb{S}}(\eta), \tau_{\leq 1} R)$. The latter is equivalent to the space of nullhomotopies of η in $\tau_{\leq 1} R$, ie $\mathrm{Map}_{\mathrm{Sp}}(\Sigma^2 \mathbb{S}, \tau_{\leq 1} R) \simeq *$. Using that thus $\tau_{\leq 1} R$ has an essentially unique structure of an $H\mathbb{Z}$–E_∞–algebra, we deduce again from [32, Theorem 7.5.0.6] that the space of E_∞–maps from $H\pi_0 R$ to $\tau_{\leq 1} R$ is equivalent to the set of ring homomorphisms $\pi_0 R \to \pi_0 R$.

Given now $R'' \to R$ as in the proposition, we obtain a map $R'' \to \tau_{\leq 1} R'' \simeq H\pi_0 R \to \tau_{\leq 1} R$. We see that R'' arises as a pullback of a diagram of the same shape as (2.8), but possibly with a map $H\pi_0 R \to \tau_{\leq 1} R$ inducing a different isomorphism f on π_0 than the identity. The paragraph above implies that using the map f on $H\pi_0 R$ we obtain an equivalence between the cospans constructing R' and R'' and thus between R' and R'' over R. ∎

To apply this to topological modular forms, we need the following two lemmas.

Lemma 2.10 *Let Γ be a tame congruence subgroup with respect to a localization $\mathbb{Z}_S$. Then η is zero in $\pi_1 \mathrm{Tmf}(\Gamma)_S$.*

Proof According to [37, Proposition 2.5], the descent spectral sequence

$$H^s(\overline{\mathcal{M}}(\Gamma); \omega^{\otimes t}) \Rightarrow \pi_{2t-s}\mathrm{Tmf}(\Gamma)$$

for $\mathrm{Tmf}(\Gamma)_S$ is concentrated in lines 0 and 1. Thus $\pi_1\mathrm{Tmf}(\Gamma)_S \cong H^1(\overline{\mathcal{M}}(\Gamma)_S; \omega)$ and it suffices to show that the image of η in $H^1(\overline{\mathcal{M}}(\Gamma)_S; \omega)$ is trivial. This is the content of [36, Proposition 2.16] unless $\Gamma_1(n) \subsetneq \Gamma \subsetneq \Gamma_0(n)$. For the remainder of the proof, assume that we are in this case and set $G = \Gamma / \Gamma_1(n)$.

We will argue that the map

$$H^1(\overline{\mathcal{M}}(\Gamma)_{(2)}; \omega) \to H^1(\overline{\mathcal{M}}(\Gamma_1(n)_{(2)}; \omega)$$

is isomorphic to the inclusion of G–invariants. As η vanishes in the target, this will imply the vanishing of η in the source.

The map $\overline{\mathcal{M}}(\Gamma_1(n))_{(2)} \to \overline{\mathcal{M}}(\Gamma)_{(2)}$ induces a map

$$c\colon \mathcal{X} = [\overline{\mathcal{M}}(\Gamma_1(n))_{(2)}/G] \to \overline{\mathcal{M}}(\Gamma)_{(2)}$$

from the stack quotient. Denote the pullback of ω to $\mathcal{X}$ also by ω. By [37, Lemma A.2], the induced map $H^1(\overline{\mathcal{M}}(\Gamma)_{(2)}; \omega) \to H^1(\mathcal{X}; \omega)$ is an isomorphism. Moreover, the descent spectral sequence

$$H^i(G; H^j(\overline{\mathcal{M}}(\Gamma_1(n))_{(2)}; \omega)) \Rightarrow H^{i+j}(\mathcal{X}; \omega)$$

is concentrated in the zero-line since the order of G is invertible in $\mathbb{Z}_{(2)}$ by the tameness of Γ. Thus,

$$H^1(\overline{\mathcal{M}}(\Gamma)_{(2)}; \omega) \cong H^1(\mathcal{X}; \omega) \to H^1(\overline{\mathcal{M}}(\Gamma_1(n)_{(2)}; \omega)$$

is indeed the inclusion of G–invariants. □

Lemma 2.11 *Let Γ be a tame congruence subgroup with respect to a localization $\mathbb{Z}_S$. Then $\pi_0\mathrm{Tmf}(\Gamma) \cong \mathbb{Z}_S$ if $\Gamma_1(n) \subset \Gamma \subset \Gamma_0(n)$ and $\pi_0\mathrm{Tmf}(\Gamma) \cong \mathbb{Z}_S[\zeta_n]$ if $\Gamma = \Gamma(n)$.*

Proof As recalled above, we have $\pi_0\mathrm{Tmf}(\Gamma) \cong H^0(\overline{\mathcal{M}}(\Gamma); \mathcal{O}_{\overline{\mathcal{M}}(\Gamma)})$. In the cases that $\Gamma = \Gamma_0(n)$, $\Gamma_1(n)$ or $\Gamma(n)$ the computation of this group is classical and can be found for instance in [36, Proposition 2.13]. The case of $\Gamma_1(n) \subsetneq \Gamma \subsetneq \Gamma_0(n)$ follows by identifying $H^0(\overline{\mathcal{M}}(\Gamma); \mathcal{O}_{\overline{\mathcal{M}}(\Gamma)})$ with $H^0(\overline{\mathcal{M}}(\Gamma); \mathcal{O}_{\overline{\mathcal{M}}(\Gamma)})^{\Gamma/\Gamma_1(n)}$ again using [37, Lemma A.2]. □

This allows us to use Proposition 2.9 to define $\mathrm{tmf}(\Gamma)_S$ in the tame case by killing π_1 from $\tau_{\geq 0}\mathrm{Tmf}(\Gamma)_S$. Summarizing we obtain:

Theorem 2.12 *For every set of primes S and every congruence subgroup Γ that is tame with respect to $\mathbb{Z}_S$, there is up to equivalence a unique connective E_∞–ring spectrum $\mathrm{tmf}(\Gamma)_S$ with an E_∞–ring map $\mathrm{tmf}(\Gamma)_S \to \mathrm{Tmf}(\Gamma)_S$ that identifies the homotopy groups of the source with the ring of holomorphic modular forms $M(\Gamma; \mathbb{Z}_S)$.*

Formally, we could also apply this procedure in some nontame cases (for instance if we localize away from 2), but the author knows of no reason to regard these constructions in these cases as "correct".

Notation 2.13 We will use the abbreviations

$$\mathrm{tmf}_1(n) = \mathrm{tmf}(\Gamma_1(n)), \quad \mathrm{tmf}_0(n) = \mathrm{tmf}(\Gamma_0(n)), \quad \mathrm{tmf}(n) = \mathrm{tmf}(\Gamma(n)),$$

when these make sense.

Remark 2.14 For every ring spectrum R, we can consider the stack $\mathcal{X}_R$ associated to the graded Hopf algebroid $(\mathrm{MU}_{2*}(R), (\mathrm{MU}\otimes\mathrm{MU})_{2*}(R))$. If R is complex orientable, this coincides with the stack quotient $[\mathrm{Spec}\,\pi_{2*}R/\mathbb{G}_m]$. In [38, Definition 5.5] we introduced cubic versions $\mathcal{M}_1(n)_{\mathrm{cub}}$ and $\mathcal{M}_0(n)_{\mathrm{cub}}$ of the moduli stacks $\mathcal{M}(\Gamma_1(n))$ and $\mathcal{M}(\Gamma_0(n))$. These come with a finite morphism to the moduli stack $\mathcal{M}_{\mathrm{cub}}$ of cubic curves, where we allow arbitrary Weierstraß equations. In [38, Theorem 5.19] we showed that $\mathcal{M}_1(n)_{\mathrm{cub}} \simeq [M(\Gamma_1(n); \mathbb{Z}[1/n])/\mathbb{G}_m]$ for $n \geq 2$. In combination, we see that $\mathcal{X}_{\mathrm{tmf}_1(n)} \simeq \mathcal{M}_1(n)_{\mathrm{cub}}$ for $n \geq 2$. In the case $n = 1$, the corresponding equivalence $\mathcal{X}_{\mathrm{tmf}} \simeq \mathcal{M}_{\mathrm{cub}}$ has a quite different character and was shown in [34]. Whether there are equivalences $\mathcal{X}_{\mathrm{tmf}_0(n)} \simeq \mathcal{M}_0(n)_{\mathrm{cub}}$ for a suitable definition of $\mathrm{tmf}_0(n)$ remains open, to the knowledge of the author, even for $n = 3$.

2.2 The C_2–equivariant argument

All the stacks $\mathcal{M}(\Gamma)$ come with an involution induced from postcomposing the level structure with the $[-1]$–automorphism of the elliptic curve. As explained in Remark 2.15, this induces a C_2–action on $\mathrm{Tmf}(\Gamma)$. Our goal in this subsection is to define suitable C_2–spectra $\mathrm{tmf}(\Gamma)$ in the tame case. This will allow us to construct an E_∞–ring spectrum $\mathrm{tmf}(\Gamma)$ also if there is just a tame subgroup $\Gamma' \subset \Gamma$ of index 2; see Construction 2.24.

Remark 2.15 The goal of this remark is to clarify the construction of the C_2–action on $\mathrm{Tmf}(\Gamma)$ sketched above.

Denote the automorphism of $\mathcal{M}(\Gamma)$ described above by t. As t commutes with the forgetful map $\mathrm{pr}\colon \mathcal{M}(\Gamma) \to \mathcal{M}_{\mathrm{ell}}$, this defines a C_2–action inside the slice category $(\mathrm{Stacks}/\mathcal{M}_{\mathrm{ell}})^{\text{ét,op}}$ of stacks étale over $\mathcal{M}_{\mathrm{ell}}$. We will use a lax commutative triangle

$$\begin{tikzcd}
(\mathrm{Stacks}/\mathcal{M}_{\mathrm{ell}})^{\text{ét,op}} \arrow[rr, "N"] \arrow[dr, "\mathcal{O}^{\mathrm{top}}"'] & \arrow[d, phantom, "\Longleftarrow"] & (\mathrm{Stacks}/\overline{\mathcal{M}}_{\mathrm{ell}})^{\text{log-ét,op}} \arrow[dl, "\mathcal{O}^{\mathrm{top}}"] \\
& \mathrm{CAlg}(\mathrm{Sp}) &
\end{tikzcd}$$

Here, the diagonal arrows are the Goerss–Hopkins–Miller and Hill–Lawson sheaves of ring spectra. The horizontal arrow N is a normalization construction; see for example [22, Proposition 2.27]. The canonical map $\mathcal{O}^{\mathrm{top}}(N(U)) \to \mathcal{O}^{\mathrm{top}}(U)$ for $U \to \mathcal{M}_{\mathrm{ell}}$ étale comes from the fact that $U \subset N(U)$ is an open substack and the Hill–Lawson sheaf restricts to the Goerss–Hopkins–Miller sheaf.

Applying the left diagonal arrow to $(\mathcal{M}(\Gamma), t)$ gives a C_2–action on $\mathrm{TMF}(\Gamma)$. Doing the same with the composite of the right diagonal arrow and the horizontal arrow produces the C_2–action on $\mathrm{Tmf}(\Gamma)$. Moreover, we obtain a C_2–map $\mathrm{Tmf}(\Gamma) \to \mathrm{TMF}(\Gamma)$.

As explained in [37, Example 6.12], the C_2–action induced by t on $\mathrm{TMF}(\Gamma)$ is equivalent to the one induced by the C_2–action in $(\mathrm{Stacks}/\mathcal{M}_{\mathrm{ell}})^{\text{ét,op}}$ given by $\mathrm{id}_{\mathcal{M}(\Gamma)}$ on $\mathcal{M}(\Gamma)$, but choosing the $[-1]$–isomorphism between the elliptic curves classified by pr and $\mathrm{pr}\,\mathrm{id}_{\mathcal{M}(\Gamma)}$. This C_2–action induces multiplication by -1 on the pullback of ω to $\mathcal{M}(\Gamma)$: indeed, the $[-1]$–automorphism of an elliptic curve induces multiplication by -1 on the sheaf of differentials. Moreover, pr classifies precisely the pullback of the universal elliptic curve $\mathcal{E}^{\mathrm{uni}}$ and ω is the restriction of $\Omega^1_{\mathcal{E}^{\mathrm{uni}}/\mathcal{M}_{\mathrm{ell}}}$ to $\mathcal{M}_{\mathrm{ell}}$ along the zero section.

Thus, if Γ is tame, it implies that C_2 acts by $(-1)^k$ on $\pi_{2k}\mathrm{TMF}(\Gamma) \cong H^0(\mathcal{M}(\Gamma); \omega^{\otimes k})$. Since $\pi_{2k}\mathrm{Tmf}(\Gamma)$ injects in the tame case into $\pi_{2k}\mathrm{TMF}(\Gamma)$, the same is true for $\pi_{2k}\mathrm{Tmf}(\Gamma)$.

The action t can be trivial, eg for $\Gamma = \Gamma_0(n)$ or $\Gamma(2)$. This forces $\pi_{2k}\mathrm{Tmf}(\Gamma) = 0$ for k odd in these cases (as t acts both by 1 and -1 and the groups are torsion-free). This corresponds to the classical fact that there are no modular forms of odd weight if $-\mathrm{id}$ is in Γ.

In the following we will use standard notation from equivariant homotopy theory. In particular, for an inner product space V with G–action, we denote by $S(V)$ the unit sphere and by S^V the 1–point compactification as G–spaces. We denote by $a = a_\sigma\colon S^0 \to S^\sigma$ the inclusion for σ the real sign representation of C_2.

The Hopf map defines a C_2–map $\overline{\eta}\colon S(\mathbb{C}^2) \to S^{\mathbb{C}}$, where C_2 acts on $\mathbb{C}$ via complex conjugation. This stabilizes to an element in $\pi_\sigma^{C_2}\mathbb{S}$, which restricts to $\eta \in \pi_1^e\mathbb{S}$.

Lemma 2.16 *The homotopy groups $\pi_\sigma^{C_2}(\mathbb{S})$ and $\pi_{-\sigma}^{C_2}\mathbb{S}$ are infinite cyclic and generated by $\overline{\eta}$ and a, respectively.*

Proof For $\pi_\sigma^{C_2}(\mathbb{S})$, this is proven as formula (8.1) in [2]. (Note that they use the notation $\pi_{p,q}^s$ for our $\pi_{p\sigma+q}^{C_2}(\mathbb{S})$.) Proposition 7.0 in op. cit. implies that the homomorphism $\pi_{-\sigma}^{C_2}\mathbb{S} \to \pi_0\mathbb{S}$, taking a map $\mathbb{S} \to \Sigma^\sigma\mathbb{S}$ to its geometric fixed points is an isomorphism. Taking fixed points of the map a clearly gives the identity map $S^0 \to S^0$, which yields the result. □

In the following, we denote by $\tau_{\le i}$ the slice coconnective cover, by $\tau_{\ge i}$ the slice connective cover and by $\tau_i = \tau_{\ge i}\tau_{\le i}$ the i^{th} slice for C_2–spectra. We refer to [16] for background about the slice filtration. We denote by $H\underline{\mathbb{Z}}$ the C_2–Eilenberg–Mac Lane spectrum for the constant Mackey functor $\underline{\mathbb{Z}}$.

Lemma 2.17 *We have an equivalence $\tau_{\le 1}C\overline{\eta} \simeq H\underline{\mathbb{Z}}$.*

Proof It suffices to show that the first slice of $C\overline{\eta}$ is null and the zeroth slice is $H\underline{\mathbb{Z}}$. As shown in [16] and summarized in [18, Section 2.4], this is implied by the calculations $\underline{\pi}_0 C\overline{\eta} \cong \underline{\mathbb{Z}}$ and $\underline{\pi}_\sigma C\overline{\eta} = 0$. These follows easily by the long exact sequence arising from the cofiber sequence

$$S^\sigma \xrightarrow{\overline{\eta}} S^0 \to C\overline{\eta}$$

and the computations of $\pi_{-\sigma}^{C_2}\mathbb{S}$, $\pi_0^{C_2}\mathbb{S}$ and $\pi_\sigma^{C_2}(\mathbb{S})$ above, using also that $\pi_{-1}^{C_2}S^\sigma = 0$. □

The following lemma is a C_2–slice analogue of Lewis's equivariant Hurewicz theorem [30, Theorem 2.1]. Recall that a C_2–spectrum is *k–slice connected* if and only if $\tau_{\le k}X = 0$.

Lemma 2.18 *A connective C_2–spectrum X is k–slice connected if and only if $H\underline{\mathbb{Z}} \otimes X$ is k–slice connected.*

Spelled out, the latter condition is equivalent to $\underline{H}_V(X;\underline{\mathbb{Z}}) = \underline{\pi}_V(H\underline{\mathbb{Z}} \otimes X) = 0$ for all C_2–representations V of the form $i\rho$ or $i\rho - 1$ with $|V| \le k$.

Proof If X is k–slice connected, the same is true for $H\underline{\mathbb{Z}} \otimes X$. For the converse, assume that $\underline{H}_V^{C_2}(X;\underline{\mathbb{Z}}) = 0$ for all C_2–representations V of the form $i\rho$ or $i\rho - 1$ with $|V| \leq k$. By induction on k, we can assume that X is $(k-1)$–slice connected and we need to show that $\tau_k X = 0$ to deduce that X is indeed k–slice connected. Let W be $\frac{1}{2}k\rho$ if k is even and $\frac{1}{2}(k+1)\rho - 1$ if k is odd. As $\tau_{\geq k+1} X$ and its suspension are k–slice connected, the direction discussed above shows $\underline{H}_W(\tau_{\geq k+1} X;\underline{\mathbb{Z}}) = \underline{H}_W(\Sigma\tau_{\geq k+1} X;\underline{\mathbb{Z}}) = 0$. Thus,

$$0 = \underline{H}_W(X;\underline{\mathbb{Z}}) \to \underline{H}_W(\tau_k X;\underline{\mathbb{Z}})$$

is an isomorphism. As summarized in [18, Section 2.4], the slice $\tau_k X$ is of the form $S^W \otimes HM$ for some Mackey functor M and we deduce that $\underline{H}_0(HM;\underline{\mathbb{Z}}) \cong \underline{H}_W(\tau_k X;\underline{\mathbb{Z}}) = 0$.

We know that $\tau_0 \mathbb{S} = H\underline{\mathbb{Z}}$. As HM is (slice) connective, a similar argument to before shows that

$$M \cong \underline{\pi}_0(\mathbb{S} \otimes HM) \cong \underline{\pi}_0(H\underline{\mathbb{Z}} \otimes HM) = \underline{H}_0(HM;\underline{\mathbb{Z}}) = 0.$$

Thus, $\tau_k X = 0$, as was to be shown. □

For an element $x \in \pi_k^{C_2}\mathbb{S}$, we can define a (naive) C_2–equivariant E_∞–cone $C^{\mathbb{S}}(x)$ as in the nonequivariant situation in the preceding subsection. The arguments for the following two results are quite analogous to those of the preceding section, so we allow ourselves to be brief.

Lemma 2.19 *The map $C(\overline{\eta}) \to C^{\mathbb{S}}(\overline{\eta})$ is slice-2–connected.*

Proof By Lemma 2.18 it suffices to check that $C(\overline{\eta}) \otimes H\underline{\mathbb{Z}} \to C^{\mathbb{S}}(\overline{\eta}) \otimes H\underline{\mathbb{Z}}$ is slice-2–connected. Since $\underline{\pi}_\sigma H\underline{\mathbb{Z}} = 0$ and thus $\overline{\eta}$ becomes zero in $H\underline{\mathbb{Z}}$, this agrees with

$$\begin{aligned} H\underline{\mathbb{Z}} \oplus \Sigma^\rho H\underline{\mathbb{Z}} \to C^{H\underline{\mathbb{Z}}}(\Sigma H\underline{\mathbb{Z}}) &\simeq \mathbb{P}^{H\underline{\mathbb{Z}}}\Sigma^\rho H\underline{\mathbb{Z}} \\ &\simeq H\underline{\mathbb{Z}} \oplus \Sigma^\rho H\underline{\mathbb{Z}} \oplus (\Sigma^{2\rho} H\underline{\mathbb{Z}})_{hC_2} \oplus \cdots. \end{aligned}$$

Analogously to Lemma 2.4, the map is split injective and thus indeed slice-2–connected (even slice-3–connected). □

Together with Lemma 2.17 this implies that $\tau_{\leq 1} C^{\mathbb{S}}(\overline{\eta}) \simeq H\underline{\mathbb{Z}}$. To deduce the analogue of Proposition 2.9, we will need one more categorical result.

Lemma 2.20 *Let G be a finite group and Sp_G be the ∞–category of G–spectra. Denote by $\mathrm{Sp}_G^{\geq 0}$ the full subcategory of connective G–spectra and by $\mathrm{Sp}_G^{[0,k]}$ that of connective and slice-k–truncated G–spectra. Then the inclusion*

$$\mathrm{CAlg}(\mathrm{Sp}_G^{[0,k]}) \to \mathrm{CAlg}(\mathrm{Sp}_G)$$

admits for every $k \geq 0$ a left adjoint, which agrees on the level of underlying G–spectra with the slice truncation $\tau_{\leq k}$.

Proof Connective G–spectra form a presentable ∞–category with compact generators the $\Sigma^\infty G/H_+$. We obtain $\mathrm{Sp}_G^{[0,k]}$ by localizing $\mathrm{Sp}_G^{\geq 0}$ at the collection S of maps $C \to 0$ for C a slice cell of dimension greater than k. By [31, Proposition 5.5.4.15], $\mathrm{Sp}_G^{[0,k]}$ is presentable again.

If X is connective and $Y \geq k+1$ in the slice filtration, then by [16, Proposition 4.26] $X \otimes Y \geq k+1$. Thus, $\tau_{\leq k}$ is compatible with $\otimes$ in the following sense: if $X \to Y$ in $\mathrm{Sp}_G^{\geq 0}$ induces an equivalence $\tau_{\leq k}X \to \tau_{\leq k}Y$, then $\tau_{\leq k}(X \otimes Z) \to \tau_{\leq k}(Y \otimes Z)$ is an equivalence for every $Z \in \mathrm{Sp}_G^{\geq 0}$. By [32, Proposition 2.2.1.9], $\mathrm{Sp}_G^{[0,k]}$ inherits the structure of a symmetric monoidal ∞–category from $\mathrm{Sp}_G^{\geq 0}$ and $\tau_{\leq k}$ is strong symmetric monoidal, while the inclusion $\mathrm{Sp}_G^{[0,k]} \to \mathrm{Sp}_G^{\geq 0}$ is lax symmetric monoidal. The same proposition gives that the resulting maps

$$(\mathrm{Sp}_G^{[0,k]})^\otimes \to (\mathrm{Sp}_G^{\geq 0})^\otimes \quad \text{and} \quad (\mathrm{Sp}_G^{\geq 0})^\otimes \to (\mathrm{Sp}_G^{[0,k]})^\otimes$$

of ∞–operads are adjoint. Since commutative algebras in such an ∞–operad $\mathcal{C}^\otimes$ are defined as sections of $\mathcal{C}^\otimes \to \mathrm{NFin}_*$ as maps of operads, we see that the resulting maps between $\mathrm{CAlg}(\mathrm{Sp}_G^{[0,k]})$ and $\mathrm{CAlg}(\mathrm{Sp}_G^{\geq 0})$ are indeed adjoint. Here, we use the characterization of an adjunction given by [42], namely the existence of a unit and counit, satisfying the triangle identities up to homotopy. □

Proposition 2.21 *Let R be a connective E_∞–ring C_2–spectrum with $\underline{\pi}_0^{C_2} = \mathbb{Z}_S$ being a localization of $\mathbb{Z}$ and $\overline{\eta} = 0 \in \pi_\sigma^{C_2} R$. Then there is an E_∞–ring C_2–spectrum R' with an E_∞–map $R' \to R$ inducing an equivalence on slices in degree 0 and degrees at least 2 and such that $\tau_1 R' = 0$. Moreover, for every other $R'' \to R$ with these properties, there is an equivalence $R'' \to R'$ of E_∞–ring C_2–spectra over R.*

Proof Since $\overline{\eta}$ is zero in R, we obtain a map $C^{\mathbb{S}}(\overline{\eta})_S \to R \to \tau_{\leq 1}R$ of E_∞–ring C_2–spectra, which factors over $H\underline{\mathbb{Z}}_S = \tau_{\leq 1}C^{\mathbb{S}}(\overline{\eta})_S$. Define now R' via the homotopy pullback square

$$\begin{tikzcd} R' \arrow[r] \arrow[d] & H\underline{\mathbb{Z}}_S \arrow[d] \\ R \arrow[r] & \tau_{\leq 1} R \end{tikzcd}$$

The proof of uniqueness is analogous to Proposition 2.9. □

To formulate the consequences for $\mathrm{tmf}(\Gamma)$, we want to recall from [18] that a C_2–spectrum E is strongly even if its odd slices vanish and its even slices are of the form $S^{k\rho} \otimes H\underline{A}$ or, equivalently, if $\underline{\pi}_{k\rho} E$ is constant and $\underline{\pi}_{k\rho-1} E = 0$.

Theorem 2.22 *For every set of primes S and every congruence subgroup $\Gamma_1(n) \subset \Gamma \subset \Gamma_0(n)$ that is tame with respect to $\mathbb{Z}_S$, we can define a strongly even connective E_∞–ring C_2–spectrum $\mathrm{tmf}(\Gamma)_S$ with an E_∞–ring C_2–map $\mathrm{tmf}(\Gamma)_S \to \mathrm{Tmf}(\Gamma)_S$ that identifies the underlying homotopy groups of the source with $M(\Gamma; \mathbb{Z}_S)$.*

Proof Equip $\mathrm{Tmf}(\Gamma) \simeq \mathrm{Tmf}(\Gamma)^{E(C_2)_+}$ with the cofree structure of a C_2–spectrum. We claim that

(1) $\overline{\eta} \in \pi_\sigma^{C_2} \tau_{\geq 0} \mathrm{Tmf}(\Gamma)$ is zero,

(2) the only odd slice of $\tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is τ_1, and

(3) the even slices of $\mathrm{Tmf}(\Gamma)$ are of the form $S^{k\rho} \otimes H\underline{A}$.

Given these claims, applying Proposition 2.21 to $R = \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ yields the C_2–spectrum $R' = \mathrm{tmf}(\Gamma)$ with the required properties: the first claim implies that we can apply Proposition 2.21, while the other two ensure that $\mathrm{tmf}(\Gamma)$ is strongly even.

For proving the claims, we will distinguish the (overlapping) cases that $\frac{1}{2} \in \mathbb{Z}_S$ and that Γ is tame for $\mathbb{Z}_{(2)}$.

For the first claim, note that $\pi_\sigma^{C_2} \tau_{\geq 0}\mathrm{Tmf}(\Gamma) \cong \pi_\sigma^{C_2}\mathrm{Tmf}(\Gamma)$ (eg since S^σ is a slice cell). The restriction map $\pi_\sigma^{C_2}\mathrm{Tmf}(\Gamma) \to \pi_1\mathrm{Tmf}(\Gamma)$ is an injection: if $\frac{1}{2} \in \mathbb{Z}_S$, this follows from the homotopy fixed points spectral sequence; else, use the line after (6.15) in [37]. Since $\overline{\eta}$ restricts to $\eta \in \pi_1\mathrm{Tmf}(\Gamma)$, Lemma 2.10 implies thus the vanishing of $\overline{\eta}$.

If Γ is tame for $\mathbb{Z}_{(2)}$, Theorem 6.16 of [37] yields the last two claims. If $\frac{1}{2} \in S$, we obtain $\pi_{k\rho-1}^{C_2}\mathrm{Tmf}(\Gamma) = 0$ by the homotopy fixed point spectral sequence since

$\pi_{2k-1}\mathrm{Tmf}(\Gamma) = 0$ for $k > 1$ by [37, Section 2.1]. This yields the second claim by [18, Proposition 2.9]. For the third claim it is enough to show that $\underline{\pi}_{k\rho}\mathrm{Tmf}(\Gamma)$ are constant Mackey functors; see [18, Proposition 2.13]. This follows again from the homotopy fixed point spectral sequence and the fact that the C_2–action on $\pi_{k\rho}\mathrm{Tmf}(\Gamma)$ is trivial: indeed, C_2 acts by $(-1)^k$ on $\pi_{2k}\mathrm{Tmf}(\Gamma)$ (see Remark 2.15) and the presence of $k\sigma$ twists the action by the same sign. □

Remark 2.23 The case that $\Gamma = \Gamma_0(n)$ is not excluded in the previous theorem, but one easily checks that $\Gamma_0(n)$ can only be tame if $\frac{1}{2} \in \mathbb{Z}_S$. In this case, we obtain simply the cofree C_2–spectrum of $\mathrm{tmf}_0(n)_S$ with the trivial action.

Construction 2.24 Given $\Gamma' \subset \Gamma \subset \Gamma_0(n)$ with Γ' tame with respect to $\mathbb{Z}_S$ and $\Gamma/\Gamma' \cong C_2$, we can extend our previous definition by defining $\mathrm{tmf}(\Gamma)_S$ as $\mathrm{tmf}(\Gamma')_S^{C_2}$ (so for example $\mathrm{tmf}_0(3) = \mathrm{tmf}_1(3)^{C_2}$ as in [18]). If Γ itself is already tame, then $\frac{1}{2} \in \mathbb{Z}_S$. One then easily computes (eg with the slice spectral sequence) that $\pi_*\mathrm{tmf}(\Gamma')_S^{C_2} \cong \pi_*\mathrm{tmf}(\Gamma)_S$ and one can use the uniqueness part of Theorem 2.12 to identify our new definition with the previous one.

Remark 2.25 In the setting of Construction 2.24, the map $\mathrm{tmf}(\Gamma) \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is an isomorphism in π_* for $* \geq 2$ even if Γ is not tame. Indeed, the cofiber of $\mathrm{tmf}(\Gamma') \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma')$ is the target's first slice and thus by [37, Theorem 6.16] equivalent to $\Sigma^\sigma HM$, where M is the constant Mackey functor on $H^1(\overline{\mathcal{M}}(\Gamma')_S; \omega) \cong \pi_1\mathrm{Tmf}(\Gamma')$. We directly observe that the nonequivariant homotopy groups of $\Sigma^\sigma HM$ vanish in degrees at least 2. Moreover the cofiber sequence $(C_2)_+ \to S^0 \to S^\sigma$ induces a long exact sequence

$$\pi_k^e HM \to \pi_k^{C_2} HM \to \pi_{k-\sigma}^{C_2} HM \to \pi_{k-1}^e HM \xrightarrow{\mathrm{tr}} \pi_{k-1}^{C_2} HM,$$

which implies that $\pi_k^{C_2}\Sigma^\sigma HM = \pi_{k-\sigma}^{C_2} HM = 0$ for $k \geq 2$ and actually also for $k = 1$ if tr is injective, ie if $\pi_1\mathrm{Tmf}(\Gamma')$ has no 2–torsion. Thus, $\mathrm{tmf}(\Gamma) \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is indeed an isomorphism in π_* for $* \geq 2$, and even for $* = 1$ if $\pi_1\mathrm{Tmf}(\Gamma')$ has no 2–torsion.

3 Realization of Hirzebruch's level-n genus

In the previous section we defined ring spectra $\mathrm{tmf}_1(n) = \mathrm{tmf}(\Gamma_1(n))$. The spectra $\mathrm{tmf}_1(n)$ are even for $n \geq 2$ and thus complex orientable. We want to show that there is a complex orientation for $\mathrm{tmf}_1(n)$ such that the corresponding map

$$\mathrm{MU}_{2*} \to \mathrm{tmf}_1(n)_{2*} \cong M(\Gamma_1(n); \mathbb{Z}[1/n])$$

agrees with the level-n genus introduced by Hirzebruch [19] and Witten [48] and studied for instance in [27; 11; 15; 47]. We recall its definition below. For this purpose it will be convenient to use algebrogeometric language, for which we recall first the following set of definitions.

Definition 3.1 A *formal group* over a base scheme S is a Zariski sheaf $F\colon \mathrm{Sch}_S^{\mathrm{op}} \to \mathrm{Ab}$ that Zariski locally on an affine open $U = \operatorname{Spec} R \subset S$ is isomorphic to $\operatorname{Spf} R[\![t]\!]$. The R–modules $R[\![t]\!]$ glue to the *structure sheaf* $\mathcal{O}_F$ on S and the R–modules $(R[\![t]\!]/t)\cdot dt$ glue to the line bundle $\omega_{F/S}$.[4] An *invariant differential* of a formal group F is a trivialization of $\omega_{F/S}$. A *coordinate* is a section s of $\mathcal{O}_F$ that is of the form $a_0 t + a_1 t^2 + \cdots$ with $a_0 \in R^\times$ for every local trivialization $F|_{\operatorname{Spec} R} \cong \operatorname{Spf} R[\![x]\!]$.

Remark 3.2 There are different ways to state the definition of a formal group, for example as an abelian group object in one-dimensional formal Lie varieties; see [12, Definitions 1.29 and 2.2]. To compare them, note that our formal groups are automatically fpqc sheaves since $\operatorname{Spf} R[\![t]\!]$ is an fpqc sheaf. On the other hand, a trivialization of the sheaf of differentials of a one-dimensional formal Lie variety over $\operatorname{Spec} R$ determines an equivalence to $\operatorname{Spf} R[\![t]\!]$, and such trivializations exist Zariski locally.

We note that the differential ds of a coordinate s of a formal group F is an invariant differential of F, sending $a_0 t + a_1 t^2 + \cdots$ to $a_0\, dt$ locally. If $S = \operatorname{Spec} R$, a coordinate of F is equivalent datum to an isomorphism $F \cong \operatorname{Spf} R[\![s]\!]$.

Recall that given an arbitrary even ring spectrum E, a complex orientation is an element in $\widetilde{E}^2(\mathbb{CP}^\infty)$ restricting to $1 \in \widetilde{E}^2(\mathbb{CP}^1)$ after a homeomorphism $\mathbb{CP}^1 \cong S^2$ is chosen. The formal spectrum $\operatorname{Spf} E^{2*}(\mathbb{CP}^\infty)$ is a formal group over $\operatorname{Spec} E^{2*}(\mathrm{pt})$ and the line bundle ω corresponds to $\widetilde{E}^*(\mathbb{CP}^1)$; it thus comes with a canonical invariant differential corresponding to $1 \in \widetilde{E}^2(\mathbb{CP}^1)$. A complex orientation is thus a coordinate of $\operatorname{Spf} E^{2*}(\mathbb{CP}^\infty)$ in degree $* = 1$ whose differential is the canonical invariant differential.

[4] If $p\colon C \to S$ is a (generalized) elliptic curve and F is the formal completion of $\mathcal{E}$, this agrees with $\omega_{C/S} = p_*\Omega^1_{C/S}$.

We want to apply this to $E = \mathrm{tmf}_1(n)$ for $n \geq 2$. Essentially by construction, the maps

$$\pi_{2*}\mathrm{tmf}_1(n) \to \pi_{2*}\mathrm{Tmf}_1(n) \to H^0(\overline{\mathcal{M}}_1(n); \omega^{\otimes *}_{\bar{\mathcal{C}}/\overline{\mathcal{M}}_1(n)})$$

are isomorphisms, where $\bar{\mathcal{C}}$ is the universal generalized elliptic curve over $\overline{\mathcal{M}}_1(n)$. For convenience, let $\overline{\mathcal{M}}^1_1(n)$ be the relative spectrum

$$\underline{\mathrm{Spec}}_{\overline{\mathcal{M}}_1(n)}\Big(\bigoplus \omega^{\otimes *}_{\bar{\mathcal{C}}/\overline{\mathcal{M}}_1(n)}\Big),$$

which is the total space of the $\mathbb{G}_m$–torsor associated with $\omega_{\bar{\mathcal{C}}/\overline{\mathcal{M}}_1(n)}$, ie classifies generalized elliptic curves with a point of exact order n and an invariant differential. The resulting morphism

$$\overline{\mathcal{M}}^1_1(n) \to \mathrm{Spec}\, H^0(\overline{\mathcal{M}}^1_1; \mathcal{O}_{\overline{\mathcal{M}}^1_1(n)}) \cong \mathrm{Spec}\, H^0(\overline{\mathcal{M}}_1; \omega^{\otimes *}_{\bar{\mathcal{C}}/\overline{\mathcal{M}}_1(n)}) \cong \mathrm{Spec}\, \pi_{2*}\mathrm{tmf}_1(n)$$

is an open immersion, whose image is covered by the nonvanishing loci of c_4 and Δ; see [38, Proposition 3.5]. We denote by $\mathcal{C}$ the pullback of $\bar{\mathcal{C}}$ to $\overline{\mathcal{M}}^1_1(n)$. Since

$$\mathrm{tmf}_1(n)[c_4]^{-1} \simeq \mathrm{Tmf}_1(n)[c_4^{-1}] \quad \text{and} \quad \mathrm{tmf}_1(n)[\Delta^{-1}] \simeq \mathrm{Tmf}_1(n)[\Delta^{-1}]$$

are elliptic cohomology theories, their formal groups are identified with the restrictions of $\widehat{\mathcal{C}}$ to the nonvanishing loci of c_4 and Δ, respectively, and as a result $\widehat{\mathcal{C}}$ becomes identified with the restriction of $\mathrm{Spf}\,\mathrm{tmf}_1(n)^{2*}(\mathbb{CP}^\infty)$ to $\overline{\mathcal{M}}^1_1(n)$. As $\overline{\mathcal{M}}^1_1(n) \subset \mathrm{Spec}\,\pi_{2*}\mathrm{tmf}_1(n)$ induces an isomorphism on global sections of the structure sheaf, coordinates on $\mathrm{Spf}\,\mathrm{tmf}_1(n)^{2*}(\mathbb{CP}^\infty)$ are in bijection with those on $\widehat{\mathcal{C}}$ and one checks that the canonical invariant differential on the former corresponds to the canonical invariant differential on the latter. Summarizing we obtain:

Lemma 3.3 *Complex orientations* $\mathrm{MU} \to \mathrm{tmf}_1(n)$ *are in bijection with coordinates of* $\widehat{\mathcal{C}}$, *which are homogeneous of degree one and have the canonical invariant differential as differential.*

The Hirzebruch genus relies on a specific such coordinate, which we will construct momentarily. Basically we will follow [20, Chapter 7], but present a more algebro-geometric approach and give an independent treatment. The key point is the existence of a certain meromorphic function on a cover of a given generalized elliptic curve. To the purpose of constructing this function, recall that every section P into the smooth part of a generalized elliptic curve $C \to S$ is an effective Cartier divisor [26, Lemma 1.2.2], ie the kernel $\mathcal{O}_C(-(P))$ of $\mathcal{O}_C \to P_*\mathcal{O}_S$ is a line bundle. Given any linear combination of sections P_i, we denote by $\mathcal{O}_C\big(\sum_i n_i(P_i)\big)$ the corresponding tensor product of line bundles.

Lemma 3.4 *Let $n \geq 2$ and S be a $\mathbb{Z}[1/n]$–scheme. Furthermore let C/S be a generalized elliptic curve with zero-section $e\colon S \to C$ and a chosen point $P\colon S \to C$ of exact order n in the smooth locus.*

(a) *The pullback of $e^*\mathcal{O}_C((P)-(e))$ to S is canonically isomorphic to $\omega_{C/S} = e^*\Omega^1_{C/S}$.*
(b) *Let λ be an invariant differential on C. Then there exists a unique meromorphic function h on C with an n–fold zero at e and an n–fold pole at P as the only pole whose restriction along e coincides with λ^n under the identification of the previous part.*
(c) *There exists a degree-n étale cover $q\colon C' \to C$ by a generalized elliptic curve and a meromorphic function f on C' with $f^n = q^*h$.*

Proof (a) Note that $\mathcal{O}_C(-(e))$ is the ideal sheaf associated to the closed immersion e and the pullback $e^*\mathcal{O}_C((P)-(e))$ coincides with $\mathcal{O}_C(-(e))/\mathcal{O}_C(-(e))^2$ viewed as an $\mathcal{O}_S$–module. Indeed, we can cover S by opens of the form $U \cap S$, where $U \cong \operatorname{Spec} R$ is an affine open in C not intersecting the image of P. The section e corresponds to an element $s \in R$ and $U \cap S \cong \operatorname{Spec} S/s$. Then $e^*\mathcal{O}_C((P)-(e))(U\cap S)$ is the S/s–module $sS \otimes_S S/s$, which is canonically isomorphic to the S/s–module sS/s^2S. For example, by [14, Proposition II.8.12], we obtain a canonical surjective map

$$\mathcal{O}_C(-(e))/\mathcal{O}_C(-(e))^2 \to e^*\Omega^1_{C/S} = \omega_{C/S}$$

between line bundles, which is hence an isomorphism.

(b) Consider the line bundle $\mathcal{O}_C(n(P)-n(e))$. Note that $n \cdot P - n \cdot e = e$ as points on C. By [26, Theorem 2.1.2] in the case that C is an elliptic curve, and by [6, Proposition II.2.7] for generalized elliptic curves, we deduce that $\mathcal{O}_C(n(P)-n(e))$ is the pullback of a line bundle $\mathcal{L}$ on S. By part (a), $\mathcal{L} = e^*p^*\mathcal{L} = \omega_{C/S}^{\otimes n}$. By [6, Proposition II.1.6], we see that the canonical map

$$\omega_{C/S}^{\otimes n} \to p_*p^*\omega_{C/S}^{\otimes n} \cong p_*\mathcal{O}_C(n(P)-n(e))$$

is an isomorphism. Thus

$$\Gamma(\mathcal{O}_C(n(P)-n(e))) \cong \Gamma(\omega_{C/S}^{\otimes n}),$$

where the isomorphism can be identified with the pullback along e. Thus, there is a unique section h of $\mathcal{O}_C(n(P)-n(e))$ whose image is λ^n.

(c) Consider the μ_n–torsor $q\colon C' \to C$ associated with the problem of extracting an n^{th} root out of q^*h as a section of $q^*\mathcal{O}_C((P)-(e))$, in other words the μ_n–torsor associated with the pair $(h, \mathcal{O}_C((P)-(e)))$ in the sense of [39, page 125]. By construction, the required root f exists on C'. By [6, Proposition II.1.17], C' has the structure of a generalized elliptic curve provided that we can lift e to C' and $C' \to S$ has geometrically connected fibers. For the first point, it suffices to provide a section of $C' \times_C S \to S$, ie to provide an n^{th} root of e^*h. Under the identification of part (a), this is provided by λ. For the second point, we assume that $S = \operatorname{Spec} K$ with K algebraically closed of characteristic not dividing n and that C' is not connected. The stabilizer of a component C_0' must be of the form μ_m with $m < n$, and thus $C' \cong C_0' \times_{\mu_m} \mu_n$. The μ_m–torsor C_0' is hence associated with a pair $(g, \mathcal{O}_C((P)-(e)))$ such that $g^{n/m} = h$. The section g provides a trivialization of $\mathcal{O}_C(m(P)-m(e))$. This implies $m \cdot P = e$ on C' [6, Corollaire II.2.4], in contradiction with P being of exact order n. □

Construction 3.5 Let $\mathcal{C}$ be the universal generalized elliptic curve with a point of exact order n over $\overline{\mathcal{M}}_1^1(n)$. It comes, by definition, with a canonical invariant differential λ. From the preceding lemma, we obtain an n–fold étale cover $q\colon \mathcal{C}' \to \mathcal{C}$ together with a meromorphic function f on $\mathcal{C}'$ whose pullback along a lift of e agrees with λ. This function f provides a coordinate for $\widehat{\mathcal{C}}' \cong \widehat{\mathcal{C}}$. Moreover, note that f is uniquely determined by the requirements in the lemma because $\mathcal{C}'$ is irreducible (since $\overline{\mathcal{M}}_1^1(n)$ is irreducible and the locus of smoothness of $\mathcal{C}'$ in it is dense) and thus every other n^{th} root of h would have to differ by a root of unity, resulting in a different pullback to $\overline{\mathcal{M}}_1^1(n)$.

Pulling the orientation induced from f back along a map $\operatorname{Spec} \mathbb{C} \to \overline{\mathcal{M}}_1(n)$ classifying $(\mathbb{C}/\Lambda, 1/n, dz)$ results exactly in the coordinate and orientation chosen in [20].

Theorem 3.6 *For every $n \geq 2$, there is a unique complex orientation of $\mathrm{MU} \to \mathrm{tmf}_1(n)$ realizing on homotopy groups the Hirzebruch genus. Moreover, this can be uniquely refined to a morphism $\mathrm{MU}_{\mathbb{R}} \to \mathrm{tmf}_1(n)$ of C_2–ring spectra.*

Proof The first part follows from Lemma 3.3 as the Hirzebruch genus is given by a coordinate on the formal group associated with the universal generalized elliptic curve on $\overline{\mathcal{M}}_1^1(n)$. For the second point, we recall from [24, Theorem 2.25] that C_2–ring morphisms $\mathrm{MU}_{\mathbb{R}} \to \mathrm{tmf}_1(n)$ are in bijection with Real orientations of $\mathrm{tmf}_1(n)$, ie a lift of a complex orientation to a class $\mathrm{tmf}_1(n)^{\rho}_{C_2}(\mathbb{CP}^\infty)$. As $\mathbb{CP}^\infty$ can be built by cells in dimensions $k\rho$, the strong-evenness of $\mathrm{tmf}_1(n)$ from Theorem 2.22 implies that the forgetful map

$$\mathrm{tmf}_1(n)^{\rho}_{C_2}(\mathbb{CP}^\infty) \to \mathrm{tmf}_1(n)^2(\mathbb{CP}^\infty)$$

is an isomorphism; thus every complex orientation of $\mathrm{tmf}_1(n)$ refines to a unique Real orientation. □

Remark 3.7 In [11], Franke already gave a related but different algebrogeometric treatment of the Hirzebruch genus.

Remark 3.8 After the first version of this article became available, Senger has shown in [43] that the map $\mathrm{MU} \to \mathrm{tmf}_1(n)$ actually refines to one of E_∞–ring spectra. He also gives a reformulation of our treatment above in terms of Θ^1–structures.

4 Compactness, formality and faithfulness of tmf(Γ)

Given a (tame) congruence subgroup of level n, we will show that $\mathrm{tmf}(\Gamma)$ is a faithful and perfect $\mathrm{tmf}[1/n]$–module. In contrast, for example, $\mathrm{tmf}_1(3)$ will not be a perfect $\mathrm{tmf}_0(3)$–module, not even rationally. The latter result relies on $\mathrm{tmf}_0(3)_{\mathbb{Q}}$ being formal (ie multiplicatively a graded Eilenberg–Mac Lane spectrum), a result we prove in greater generality in a subsection on its own.

4.1 All tmf(Γ) are perfect

Recall that for an A_∞–ring spectrum R, a perfect R–module is a compact object in the ∞–category of left R–modules. Equivalently, the ∞–category of perfect R–modules is the smallest stable ∞–subcategory of all left R–modules that contains R and is closed under retracts. The goal of this section is to show that the spectra $\mathrm{tmf}(\Gamma)$, in the cases we defined them, are perfect $\mathrm{tmf}[1/n]$–modules. The key technical tool is the following proposition.

Proposition 4.1 *Let R be an A_∞–ring spectrum such that*

(1) *$\pi_0 R$ is regular noetherian,*

(2) *all $\pi_n R$ are finitely generated $\pi_0 R$–modules, and*

(3) *$H\pi_0 R$ is perfect as a $\tau_{\geq 0} R$–module.*

Let furthermore M be a perfect R–module. Then $\tau_{\geq k} M$ is a perfect $\tau_{\geq 0} R$–module for every $k \in \mathbb{Z}$.

Lemma 4.2 *With notation as in the statement of the proposition, let X be a $\tau_{\geq 0} R$–module with only finitely many nontrivial homotopy groups, all finitely generated over $\pi_0 R$. Then X is a perfect $\tau_{\geq 0} R$–module.*

Proof By induction, we can reduce to the case that $\pi_* X$ is concentrated in a single degree n. Then $X = H\pi_n X$ acquires the structure of a $H\pi_0 R$–module and it is perfect as such because $\pi_0 R$ is regular noetherian and $\pi_n X$ is finitely generated. As $H\pi_0 R$ is perfect over $\tau_{\geq 0} R$, the same is thus true for X. □

Proof of Proposition 4.1 Let M be a perfect R–module. As the truth of the conclusion of the proposition is clearly preserved under retracts in M and also clear for $M = 0$, we can assume by induction that we have a cofiber sequence

$$\Sigma^l R \to N \to M \to \Sigma^{l+1} R$$

where $\tau_{\geq k} N$ is a perfect $\tau_{\geq 0} R$–module for all $k \in \mathbb{Z}$. Taking $\tau_{\geq l}$ on the first two objects gives a diagram

$$\begin{tikzcd}
\Sigma^l \tau_{\geq 0} R \arrow[r] \arrow[d] & \tau_{\geq l} N \arrow[r] \arrow[d] & M' \arrow[r] \arrow[d] & \Sigma^{l+1} \tau_{\geq 0} R \arrow[d] \\
\Sigma^l R \arrow[r] & N \arrow[r] & M \arrow[r] & \Sigma^{l+1} R
\end{tikzcd}$$

of cofiber sequences. As $\tau_{\geq l} N$ is a perfect $\tau_{\geq 0} R$–module, so is M'. Clearly, we have $\tau_{\geq l+1} M' \simeq \tau_{\geq l+1} M$. As the fiber of $\tau_{\geq l+1} M' \to M'$ fulfills the conditions of the previous lemma, $\tau_{\geq l+1} M$ is perfect as a $\tau_{\geq 0} R$–module.

For a general $k \in \mathbb{Z}$, we make a case distinction: assume first that $k \geq l+1$. Then the fiber of $\tau_{\geq k} M \to \tau_{\geq l+1} M$ is perfect by the previous lemma; hence $\tau_{\geq k} M$ is perfect as well. If $k \leq l+1$, consider the fiber of $\tau_{\geq l+1} M \to \tau_{\geq k} M$ instead. □

To apply Proposition 4.1 to topological modular forms, we need the following lemma.

Lemma 4.3 *For every $n \geq 1$, the* tmf$[1/n]$*–module* $H\pi_0$tmf$[1/n] = H\mathbb{Z}[1/n]$ *is perfect.*

Proof If $2|n$, there is a 3–cell complex X such that tmf$[1/n] \otimes X \simeq$ tmf$_1(2)[1/n]$; see [34, Theorem 4.13]. We have π_*tmf$_1(2)[1/n] = \mathbb{Z}[1/n][b_2, b_4]$. Killing b_2 and b_4 gives $H\mathbb{Z}[1/n]$. Thus, $H\mathbb{Z}[1/n]$ is a perfect tmf$_1(2)[1/n]$–module and hence also a perfect tmf$[1/n]$–module.

If $3|n$, there is an 8–cell complex X such that tmf$[1/n] \otimes X \simeq$ tmf$_1(3)[1/n]$; see [34, Theorem 4.10]. We have π_*tmf$_1(3)[1/n] = \mathbb{Z}[1/n][a_1, a_3]$. Killing a_1 and a_3 gives $H\mathbb{Z}[1/n]$ and thus $H\mathbb{Z}[1/n]$ is also a perfect tmf$[1/n]$–module in this case.

For the general case, let X_i be a collection of tmf$[1/n]$–modules. Consider

$$\Phi_k\colon \bigoplus_i \mathrm{Hom}_{\mathrm{tmf}[1/n]}\Bigl(H\mathbb{Z}\Bigl[\frac{1}{n}\Bigr], X_i\Bigl[\frac{1}{k}\Bigr]\Bigr) \to \mathrm{Hom}_{\mathrm{tmf}[1/n]}\Bigl(H\mathbb{Z}\Bigl[\frac{1}{n}\Bigr], \bigoplus_i X_i\Bigl[\frac{1}{k}\Bigr]\Bigr).$$

If $k = 2, 3$ or 6, then Φ_k is an equivalence by the previous results. As for every spectrum X, there is a cofiber sequence

$$\Sigma^{-1} X\bigl[\tfrac{1}{6}\bigr] \to X \to X\bigl[\tfrac{1}{2}\bigr] \oplus X\bigl[\tfrac{1}{3}\bigr] \to X\bigl[\tfrac{1}{6}\bigr]$$

and there is a cofiber sequence of maps between mapping spectra

$$\Sigma^{-1}\,\mathrm{fib}(\Phi_6) \to \mathrm{fib}(\Phi_1) \to \mathrm{fib}(\Phi_2 \oplus \Phi_3) \to \mathrm{fib}(\Phi_6).$$

It follows that Φ_1 is an equivalence as well and that $H\mathbb{Z}[1/n]$ is a perfect tmf$[1/n]$–module. □

Theorem 4.4 *Let Γ be a congruence subgroup of level n, which is tame or has a subgroup $\Gamma' \subset \Gamma$ of index 2 with Γ' tame. Then* tmf(Γ) *is a perfect* tmf$[1/n]$*–module.*

The same conclusion holds without the tameness hypothesis for any tmf$[1/n]$*–module R with a map $R \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ whose fiber has finitely generated homotopy groups over $\mathbb{Z}[1/n]$, concentrated in finitely many degrees.*

Proof According to [37, Proposition 2.12] the Tmf$[1/n]$–module Tmf(Γ) is perfect. All $\pi_k\mathrm{Tmf}[1/n]$ are finitely generated $\mathbb{Z}[1/n]$–modules. Furthermore, $H\pi_0\mathrm{tmf}[1/n] = H\mathbb{Z}[1/n]$ is a perfect tmf$[1/n]$–module by the previous lemma. This implies that $\tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is a perfect tmf$[1/n]$–module by Proposition 4.1.

For any R as in the statement of the theorem, R is thus perfect as well, by Lemma 4.2. To see that tmf(Γ) satisfies the hypotheses on R, note first that every $H^s(\overline{\mathcal{M}}(\Gamma); \omega^{\otimes t})$ is a finitely generated $\mathbb{Z}[1/n]$–module for every s and t since $\overline{\mathcal{M}}(\Gamma)$ is proper over $\mathbb{Z}[1/n]$. If Γ is tame, the cofiber of $\mathrm{tmf}(\Gamma) \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ is by construction $H\pi_1\mathrm{Tmf}(\Gamma)$ and $\pi_1\mathrm{Tmf}(\Gamma) \cong H^1(\overline{\mathcal{M}}(\Gamma); \omega)$. If there is a tame subgroup $\Gamma' \subset \Gamma$ of index 2, the cofiber $\mathrm{tmf}(\Gamma) \to \tau_{\geq 0}\mathrm{Tmf}(\Gamma)$ agrees with $\Sigma^\sigma HM$ for M the constant Mackey functor on $H^1(\overline{\mathcal{M}}(\Gamma'); \omega)$ by Remark 2.25. The exact sequence given in the same remark implies that the homotopy groups of $\Sigma^\sigma HM$ are concentrated in degrees 0 and 1 and are finitely generated $\mathbb{Z}[1/n]$–modules. □

We recall from [33] that a connective p–complete spectrum X is called an *fp-spectrum* if $H_*(X; \mathbb{F}_p)$ is finitely presented as a comodule over the dual Steenrod algebra. They show in [33, Proposition 3.2] that, equivalently, there is a finite spectrum F with nontrivial $\mathbb{F}_p$–homology such that the total group $\pi_*(X \otimes F)$ is finite. The following proposition can be deduced from the known $\mathbb{F}_p$–(co)homology of tmf (see for example [41, Section 21]) and was already noted in [33] for $p = 2$. We prefer to give a less computational proof though.

Proposition 4.5 *The p–completion of* tmf *is an fp-spectrum for all primes p.*

Proof We implicitly p–localize. For $p \neq 3$, [34, Theorem 4.10] implies the existence of a finite spectrum W with nontrivial $\mathbb{F}_p$–homology such that $\mathrm{tmf} \otimes W \simeq \mathrm{tmf}_1(3)$. Choose a complex V such that $BP_*V \cong BP_*/(p^{k_0}, v_1^{k_1}, v_2^{k_2})$ with k_0, k_1 and k_2 positive integers. As $\mathrm{TMF}_1(3)$ is Landweber exact, the sequence p, v_1, v_2 and hence the sequence $p^{k_0}, v_1^{k_1}, v_2^{k_2}$ is regular on $\pi_*\mathrm{TMF}_1(3)$. Since $\pi_*\mathrm{tmf}_1(3) = \mathbb{Z}_{(p)}[a_1, a_3]$ is an integral domain, the sequence is also regular on $\pi_*\mathrm{tmf}_1(3)$. Thus,

$$\pi_*\mathrm{tmf} \otimes W \otimes V \cong \pi_*\mathrm{tmf}_1(3) \otimes V \cong \pi_*\mathrm{tmf}_1(3)/(p^{k_0}, a_1^{k_1}, a_3^{k_2})$$

is a finitely generated $\mathbb{Z}/p^{k_0}$–algebra and of Krull dimension 0. Hence it is of finite length as a $\mathbb{Z}/p^{k_0}$–module, and thus finite.

Essentially the same argument works for $p = 3$ if we choose instead a complex W' with $\mathrm{tmf} \otimes W' \simeq \mathrm{tmf}_1(2)$ as in [34, Theorem 4.13]. □

Corollary 4.6 *The p–completion of* $\mathrm{tmf}(\Gamma)$ *for a congruence subgroup Γ of level n and p not dividing n is an fp-spectrum.*

For implications involving duality we refer to [33] and for an implication for the Hurewicz image in $H_*(\Omega^\infty \mathrm{tmf}(\Gamma); \mathbb{F}_p)$ to [28, Theorem 1.7].

4.2 All $\mathrm{tmf}(\Gamma)_{\mathbb{Q}}$ are formal

The goal of this section is to show that the E_∞–rings $\mathrm{tmf}(\Gamma)_{\mathbb{Q}}$ are formal. While this statement is interesting in its own right, we also need it for further pursuing compactness questions in the following subsection. We begin with the following consequence of Goerss–Hopkins obstruction theory.

Proposition 4.7 *Let A and B be E_∞–$H\mathbb{Q}$–algebras such that π_*A is smooth as a $\mathbb{Q}$–algebra. Then*

$$\pi_i \operatorname{Map}_{\mathrm{CAlg}}(A, B) \cong \begin{cases} \operatorname{Hom}_{\mathrm{grCRings}}(\pi_*A, \pi_*B) & \text{if } i = 0, \\ \operatorname{Hom}_{\pi_*A}(\Omega^1_{\pi_*A/\mathbb{Q}}, \pi_{*+i}B) & \text{if } i > 0, \end{cases}$$

where for π_i with $i > 0$ a basepoint is chosen if a map $A \to B$ exists.

Proof According to [13, Section 4] or [40, Section 6] with $E = H\mathbb{Q}$, there is an obstruction theory for lifting a morphism $\pi_* A \to \pi_* B$ to a morphism $A \to B$, where the obstructions lie in $\mathrm{Ext}^{n+1,n}_{\pi_* A}(\mathbb{L}^{E_\infty}_{\pi_* A/\mathbb{Q}}, \pi_* B)$, where $\mathbb{L}^{E_\infty}$ denotes the E_∞–cotangent complex. As we are working rationally, this coincides with other forms of the cotangent complexes. In particular, we obtain from the smoothness of $\pi_* A$ that $\mathbb{L}^{E_\infty}_{\pi_* A/\mathbb{Q}}$ is isomorphic to $\Omega^1_{\pi_* A/\mathbb{Q}}$ concentrated in degree 0, which again by smoothness is a projective $\pi_* A$–module. Thus the Ext-groups vanish and there is no obstruction to lifting a morphism $\pi_* A \to \pi_* B$ to a morphism $A \to B$. The same sources provide a spectral sequence computing $\pi_* \mathrm{Map}_{\mathrm{CAlg}}(A, B)$, which collapses by a similar Ext-calculation and gives the result. $\square$

Proposition 4.8 *Let $\mathcal{X}$ be a smooth Deligne–Mumford stacks over $\mathbb{Q}$ and $\mathcal{O}$ an even-periodic sheaf of E_∞–ring spectra on $\mathcal{X}$ such that $\pi_0\mathcal{O} \cong \mathcal{O}_{\mathcal{X}}$ and the $\pi_i\mathcal{O}_{\mathcal{X}}$ are quasicoherent. Assume further that $H^{i+1}(\mathcal{X}; \pi_i\mathcal{O}) = 0$ for all even $i \geq 1$. Then $\mathcal{O}$ is formal, ie equivalent to the (sheafification of the pre)sheaf $H\pi_*\mathcal{O}$ of graded Eilenberg–Mac Lane spectra.*

Proof Note first that $(\mathcal{X}, \mathcal{O})$ actually defines a nonconnective spectral Deligne–Mumford stack and in particular $\mathcal{O}$ is hypercomplete; see eg [37, Lemma B.2]. Set $\mathcal{O}' = H\pi_*\mathcal{O}$. Choosing an étale hypercover $U_\bullet \to \mathcal{X}$ by affines, we can compute $\mathrm{Map}_{\mathrm{CAlg}_{\mathcal{X}}}(\mathcal{O}, \mathcal{O}')$ as the totalization of the cosimplicial diagram

$$M^\bullet = \mathrm{Map}_{\mathrm{CAlg}}(\mathcal{O}(U_\bullet), \mathcal{O}'(U_\bullet)).$$

We observe using Proposition 4.7 that $\pi^0\pi_0 M^\bullet$ agrees with the set of ring morphisms $\pi_*\mathcal{O} \to \pi_*\mathcal{O}'$, in which we can pick an isomorphism f_0. By [4, Sections 5.2 and 2.4], the vanishing of $\pi^{i+1}\pi_i M^\bullet \cong H^{i+1}(\mathcal{X}, \pi_i\mathcal{O})$ for $i \geq 1$ suffices to lift f_0 to a multiplicative map $\mathcal{O} \to \mathcal{O}'$, which is automatically an equivalence. $\square$

Corollary 4.9 *For all $\overline{\mathcal{M}}(\Gamma)$ the rationalized Goerss–Hopkins–Miller–Hill–Lawson sheaf $\mathcal{O}^{\mathrm{top}}$ is formal.*

Proof We can apply the previous proposition, as $\overline{\mathcal{M}}(\Gamma)_{\mathbb{Q}}$ has cohomological dimension one. (See eg [36, Proposition 2.4(4)].) $\square$

Remark 4.10 In the original account of the construction of $\mathcal{O}^{\mathrm{top}}$ on $\overline{\mathcal{M}}_{\mathrm{ell}}$ in [9], $\mathcal{O}^{\mathrm{top}}_{\mathbb{Q}}$ is actually formal *by construction*. Our argument shows that this choice was necessary, not only for $\overline{\mathcal{M}}_{\mathrm{ell}}$, but also for $\overline{\mathcal{M}}(\Gamma)$. (The former was shown in a different manner already in [17, Proposition 4.47].)

Proposition 4.11 *Let Γ be a congruence group. Then the E_∞–rings* $\mathrm{tmf}(\Gamma)_{\mathbb{Q}}$ *are formal.*

Proof Set $R = H(H^0(\overline{\mathcal{M}}(\Gamma), \pi_* \mathcal{O}^{\mathrm{top}}_{\mathbb{Q}}))$. We want to construct an equivalence between R and $\mathrm{tmf}(\Gamma)_{\mathbb{Q}}$. By the preceding corollary, we know that $\mathcal{O}^{\mathrm{top}}_{\mathbb{Q}}$ on $\overline{\mathcal{M}}(\Gamma)$ is formal. In particular this provides us with compatible maps $R \to \mathcal{O}^{\mathrm{top}}(U)_{\mathbb{Q}}$ for all affines U étale over $\overline{\mathcal{M}}(\Gamma)$. Taking the homotopy limit, we obtain a map $R \to \mathrm{Tmf}(\Gamma)_{\mathbb{Q}}$. The uniqueness part of Theorem 2.12 identifies R with $\mathrm{tmf}(\Gamma)_{\mathbb{Q}}$. □

4.3 Not all tmf(Γ) are perfect

While we have seen above that $\mathrm{tmf}(\Gamma)$ for a congruence group of level n is always perfect as a $\mathrm{tmf}[1/n]$–module, we will see in this subsection that it is not necessarily compact as a $\mathrm{tmf}(\Gamma')[1/n]$–module for $\Gamma \subset \Gamma'$. The author learned this argument from Tyler Lawson.

Lemma 4.12 *For $R = \mathrm{tmf}(\Gamma)_{\mathbb{Q}}$, the R–module $H\pi_0 R$ can only be perfect if $\pi_* R$ is regular.*

Proof By [10, Theorem 19.1, Corollary 19.5 and Theorem 19.12], $\pi_* R$ is regular if and only if the graded $\mathbb{Q}$–vector space $\mathrm{Tor}^{\pi_* R}_*(\pi_0 R, \pi_0 R)$ is concentrated in finitely many dimensions. Because R is formal by Proposition 4.11, this Tor agrees with $\pi_*(H\pi_0 R \otimes_R H\pi_0 R)$. Clearly, $H\pi_0 R$ being a perfect R–module would imply the finite-dimensionality of this quantity. □

It is actually very rare that $\pi_* \mathrm{tmf}(\Gamma)_{\mathbb{Q}} \cong M_*(\Gamma; \mathbb{Q})$ is regular. One of the few exceptions is $\Gamma = \Gamma_1(3)$, where we obtain the ring $\mathbb{Q}[a_1, a_3]$. In contrast for $\Gamma = \Gamma_0(3)$, we obtain its C_2–fixed points, ie $\mathbb{Q}[a_1^2, a_3^2, a_1 a_3] \cong \mathbb{Q}[x, y, z]/xz - y^2$, which is not regular. Thus, $H\mathbb{Q}$ is a perfect $\mathrm{tmf}_1(3)$–module, but is by the previous lemma not a perfect $\mathrm{tmf}_0(3)_{\mathbb{Q}}$–module. We obtain:

Proposition 4.13 *The* $\mathrm{tmf}_0(3)$*–module* $\mathrm{tmf}_1(3)$ *is not perfect, not even rationally.*

4.4 All tmf(Γ) are faithful

The goal of this section is to show that if Γ is a congruence subgroup of level n, then $\mathrm{tmf}(\Gamma)$ is (if defined) a faithful $\mathrm{tmf}[1/n]$–module, ie tensoring with it is conservative.

Lemma 4.14 *For every congruence subgroup Γ of level n, the* $\mathrm{Tmf}[1/n]$*–module* $\mathrm{Tmf}(\Gamma)$ *is faithful.*

Proof By [35], the derived stack $(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$ is 0–affine, ie the global sections functor

$$\Gamma\colon \mathrm{QCoh}(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}}) \to \mathrm{Mod}_{\mathrm{Tmf}}$$

is a symmetric monoidal equivalence and the same holds after inverting n. Thus our claim is equivalent to showing that tensoring with $f_*\mathcal{O}^{\mathrm{top}}_{\overline{\mathcal{M}}(\Gamma)}$ for $f\colon \overline{\mathcal{M}}(\Gamma) \to \overline{\mathcal{M}}_{\mathrm{ell},\mathbb{Z}[1/n]}$ is conservative on $\mathrm{QCoh}(\overline{\mathcal{M}}_{\mathrm{ell}}, \mathcal{O}^{\mathrm{top}})$. This can be checked étale locally, where $f_*\mathcal{O}^{\mathrm{top}}_{\overline{\mathcal{M}}(\Gamma)}$ is free of positive rank as f is finite and flat (see eg [36, Proposition 2.4]) and of positive rank everywhere (as $\overline{\mathcal{M}}_{\mathrm{ell},\mathbb{Z}[1/n]}$ is irreducible and $\overline{\mathcal{M}}(\Gamma)$ not empty). $\square$

In the following we fix a congruence subgroup Γ and a multiplicatively closed subset S of $\mathbb{Z}$ such that $\mathrm{tmf}(\Gamma)_S$ is defined (ie Γ is tame or of index 2 in a tame Γ).

Proposition 4.15 *The* tmf_S*–module* $\mathrm{tmf}(\Gamma)_S$ *is faithful for every congruence subgroup* Γ.

Proof Let $M \in \mathrm{Mod}_{\mathrm{tmf}_S}$ with $M \otimes_{\mathrm{tmf}_S} \mathrm{tmf}(\Gamma)_S = 0$. It suffices to show that $M_{(p)} = 0$ for all p not in S. Consider the case $p = 2$ and localize everything implicitly at 2. As $\mathrm{tmf}_1(3)$ is faithful over tmf (see [34, Theorem 4.10]), it suffices to show that $M' = M \otimes_{\mathrm{tmf}} \mathrm{tmf}_1(3)$ vanishes. Our assumption implies

$$(M \otimes_{\mathrm{tmf}} \mathrm{Tmf}) \otimes_{\mathrm{Tmf}} \mathrm{Tmf}(\Gamma) = 0,$$

so by the faithfulness of $\mathrm{Tmf}(\Gamma)$ also $M \otimes_{\mathrm{tmf}} \mathrm{Tmf} = 0$. Thus, $M' \otimes_{\mathrm{tmf}_1(3)} \mathrm{Tmf}_1(3) = 0$. Moreover, $\mathrm{tmf}(\Gamma) \otimes_{\mathrm{tmf}} H\mathbb{Z}$ is a faithful $H\mathbb{Z}$–module as its π_0 is a faithful $\mathbb{Z}$–module. Thus $M' \otimes_{\mathrm{tmf}_1(3)} H\mathbb{Z} \simeq M \otimes_{\mathrm{tmf}} H\mathbb{Z} = 0$.

Recall now that $\pi_*\mathrm{tmf}_1(3) \cong \mathbb{Z}[a_1, a_3]$. The map $\mathrm{tmf}_1(3)[a_i^{-1}] \to \mathrm{Tmf}_1(3)[a_i^{-1}]$ is an equivalence for $i = 1, 3$ since the cofiber of $\mathrm{tmf}_1(3) \to \mathrm{Tmf}_1(3)$ is coconnective. Thus the considerations above imply that $M'[a_1^{-1}]$, $M'[a_3^{-1}]$ and $M'/(a_1, a_3)$ all vanish, which implies the vanishing of M'.

The argument for $p = 3$ is similar with $\mathrm{tmf}_1(2)$ in place of $\mathrm{tmf}_1(3)$ and for $p > 3$ we can use tmf itself as $\pi_*\mathrm{tmf}\left[\frac{1}{6}\right] \cong \mathbb{Z}\left[\frac{1}{6}\right][c_4, c_6]$ is a polynomial ring. $\square$

5 Splittings

Our goal in this setting is to show that $\mathrm{tmf}_1(n)$ often splits p–locally into small pieces.

Fixing a natural number $n \geq 2$ and a prime p not dividing n, we will work throughout this section implicitly p–locally. We demand that $M(\Gamma_1(n), \mathbb{Z}_{(p)}) \to M(\Gamma_1(n); \mathbb{F}_p)$ is surjective. In general, this is a subtle condition, but it is for example always fulfilled if $n \leq 28$; see [36, Remark 3.14]. Equivalently, we can ask that $H^1(\overline{\mathcal{M}}_1(n); \omega) \cong \pi_1 \mathrm{Tmf}_1(n)$ does not have p–torsion. We note that this leaves plenty of cases where $\pi_1 \mathrm{Tmf}_1(n) \neq 0$ and hence $\mathrm{tmf}_1(n)$ is not the naive connective cover of $\mathrm{Tmf}_1(n)$, of which the smallest is $n = 23$.

By Theorem 1.3 of [37], we have a splitting

$$\mathrm{Tmf}_1(n) \simeq \bigoplus_i \Sigma^{2n_i} R \tag{5.1}$$

of Tmf–modules, where R is $\mathrm{Tmf}_1(3)$, $\mathrm{Tmf}_1(2)$ or Tmf, depending on whether the prime p is 2, 3 or bigger than 3. In this splitting all n_i are nonnegative.

Theorem 5.2 *Under the conditions as above, we have a splitting*

$$\mathrm{tmf}_1(n) \simeq \bigoplus_i \Sigma^{2n_i} r,$$

where $r = \tau_{\geq 0} R$.

Proof Consider the composition

$$f \colon \bigoplus_i \Sigma^{2n_i} r \to \bigoplus_i \tau_{\geq 0} \Sigma^{2n_i} R \to \tau_{\geq 0} \mathrm{Tmf}_1(n).$$

Here, the second map is just the connective cover of (5.1) (using that $\tau_{\geq 0}$ commutes with direct sums) and the first map is the direct sum of the maps

$$\Sigma^{2n_i} r \simeq \tau_{\geq 2n_i} \Sigma^{2n_i} R \to \tau_{\geq 0} \Sigma^{2n_i} R.$$

Since all negative homotopy of R is in odd degrees, we see that f is an isomorphism on even homotopy groups. Moreover, the source has only homotopy groups in even degrees.

Recall that we defined $\mathrm{tmf}_1(n)$ as a pullback

$$\begin{tikzcd} \mathrm{tmf}_1(n) \arrow[r] \arrow[d] & H\mathbb{Z} \arrow[d] \\ \tau_{\geq 0}\mathrm{Tmf}_1(n) \arrow[r] & \tau_{[0,1]}\mathrm{Tmf}_1(n) \end{tikzcd}$$

where we still localize implicitly everywhere at p. This implies a fiber sequence

$$\mathrm{tmf}_1(n) \to \tau_{\geq 0}\mathrm{Tmf}_1(n) \to \Sigma H\pi_1\mathrm{Tmf}_1(n).$$

To factor f over $\mathrm{tmf}_1(n)$, it is enough to show that $H^1(\Sigma^{2n_i}r; A) = 0$ with any coefficients A. This is clear anyhow for $n_i \geq 1$, so assume $n_i = 0$. We know that $\tau_{[0,1]}r \simeq H\mathbb{Z}$ and we have $H^1(H\mathbb{Z}; A) \cong H^1(\mathbb{S}; A) = 0$ (as the cofiber of $\mathbb{S} \to H\mathbb{Z}$ is 1–connected).

Now $\pi_*\mathrm{tmf}_1(n)$ is concentrated in even degrees and $\mathrm{tmf}_1(n) \to \tau_{\geq 0}\mathrm{Tmf}_1(n)$ induces a π_*–isomorphism in even degrees. In total, we see that f induces an isomorphism on π_*. □

Remark 5.3 The condition that $\pi_1\mathrm{Tmf}_1(n) \cong H^1(\overline{\mathcal{M}}_1(n); \omega)$ does not have p–torsion is actually necessary in the preceding theorem. One can indeed show that $\mathrm{Tmf}_1(n)$ can be recovered as $\mathrm{tmf}_1(n) \otimes_{\mathrm{tmf}} \mathrm{Tmf}$. Thus a p–local tmf–linear splitting of $\mathrm{tmf}_1(n)$ into shifted copies of r implies a p–local splitting of $\mathrm{Tmf}_1(n)$ into copies of R. As the latter has torsion-free homotopy groups, such a splitting can indeed only occur if the homotopy groups of $\mathrm{Tmf}_1(n)$ are p–torsion-free as well.

We now fix $p = 2$ and are thus assuming that $\pi_1\mathrm{Tmf}_1(n) \cong H^1(\overline{\mathcal{M}}_1(n); \omega)$ does not have 2–torsion — this is true for all odd $2 \leq n < 65$ by [36, Remark 3.14], for example. In this setting we also want to prove connective versions of the C_2–equivariant refinement

$$\mathrm{Tmf}_1(n)_{(2)} \simeq_{C_2} \bigoplus_i \Sigma^{n_i\rho}\mathrm{Tmf}_1(3)_{(2)} \tag{5.4}$$

of (5.1) given in [37, Theorem 6.19], where ρ is the regular representation of C_2. We need the following lemma:

Lemma 5.5 *Let A be an abelian group without 2–torsion, and denote by $\underline{A}$ the corresponding constant C_2–Mackey functor. Then $\pi^{C_2}_{-\sigma}H\underline{A} \cong A \otimes \mathbb{Z}/2$, and the map*

$$[H\underline{\mathbb{Z}}, \Sigma^{\sigma}H\underline{A}]^{C_2} \xrightarrow{\pi_0^{C_2}} A \otimes \mathbb{Z}/2$$

is an isomorphism.

Proof Smashing the fundamental cofiber sequence

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

with $S^{-\sigma}$ and mapping out of it yields an exact sequence

$$\pi^e_{-1} H\underline{A} \leftarrow \pi^{C_2}_{-\sigma} H\underline{A} \leftarrow \pi^{C_2}_0 H\underline{A} \leftarrow \pi^e_0 H\underline{A}.$$

The rightmost arrow can be identified with the transfer $\mathrm{tr} = 2\colon A \to A$ of the constant Mackey functor, while $\pi^e_{-1} H\underline{A} = 0$. We obtain $\pi^{C_2}_{-\sigma} H\underline{A} \cong A \otimes \mathbb{Z}/2$ as claimed.

To finish the proof, we recall from Section 2.2 that $\tau_{\leq 1} C\overline{\eta} \simeq H\underline{\mathbb{Z}}$. As $\Sigma^\sigma H\underline{A} \leq 1$ in the slice filtration, this implies that $[H\underline{\mathbb{Z}}, \Sigma^\sigma H\underline{A}]^{C_2} \cong [C\overline{\eta}, \Sigma^\sigma H\underline{A}]^{C_2}$. This sits in a long exact sequence

$$0 = \pi^{C_2}_1 H\underline{A} \to [C\overline{\eta}, \Sigma^\sigma H\underline{A}] \to \pi^{C_2}_{-\sigma} H\underline{A} \to \pi^{C_2}_0 H\underline{A} = A.$$

As A does not have 2–torsion and we have shown above that $\pi^{C_2}_{-\sigma} H\underline{A} \cong A \otimes \mathbb{Z}/2$, the result follows. □

Theorem 5.6 *Assuming that $n \geq 3$ is odd and $H^1(\overline{\mathcal{M}}_1(n); \omega)$ does not have 2–torsion, we have 2–locally a C_2–equivariant splitting*

$$\mathrm{tmf}_1(n) \simeq \bigoplus_i \Sigma^{n_i \rho} \mathrm{tmf}_1(3).$$

Proof We localize everywhere implicitly at 2 and consider the map

$$\bigoplus_i \Sigma^{n_i \rho} \mathrm{tmf}_1(3) \to \bigoplus_i \tau_{\geq 0} \Sigma^{n_i \rho} \mathrm{Tmf}_1(3) \xrightarrow{\tau_{\geq 0}\Phi} \tau_{\geq 0} \mathrm{Tmf}_1(n),$$

for a chosen C_2–equivalence Φ between $\bigoplus_i \Sigma^{n_i \rho} \mathrm{Tmf}_1(3)$ and $\mathrm{Tmf}_1(n)$. We have a fiber sequence

$$\mathrm{tmf}_1(n) \to \tau_{\geq 0} \mathrm{Tmf}_1(n) \to \Sigma^\sigma H\underline{A},$$

where $A = H^1(\overline{\mathcal{M}}_1(n); \omega)$ since by [37, Theorem 6.16], $\Sigma^\sigma H\underline{A}$ is the 1–slice of $\mathrm{Tmf}_1(n)$. On $\pi^{C_2}_0$ this induces (using Lemma 5.5) a short exact sequence

$$0 \to \mathbb{Z} \to \pi^{C_2}_0 \mathrm{Tmf}_1(n) \xrightarrow{r} A \otimes \mathbb{Z}/2 \to 0. \tag{5.7}$$

The composite $\bigoplus \Sigma^{n_i \rho} \mathrm{tmf}_1(3) \to \Sigma^\sigma H\underline{A}$ factors over the 1–slice coconnective cover of the source, which agrees with $H\underline{\mathbb{Z}}$ since there is precisely one n_i equaling 0 (by considering nonequivariant homotopy groups). Using Lemma 5.5 again, the resulting map $H\underline{\mathbb{Z}} \to \Sigma^\sigma H\underline{A}$ is null if and only if the image $r(\Phi(1))$ of $\Phi(1)$ in $A \otimes \mathbb{Z}/2$ is 0.

We want to show that we can change Φ so that this is true. Using Φ, the C_2–spectrum $\mathrm{Tmf}_1(n)$ gets the structure of a $\mathrm{Tmf}_1(3)$–module. Thus, $\mathrm{Tmf}_1(3)$–module maps

$\bigoplus_{i=0}^N \Sigma^{n_i\rho}\mathrm{Tmf}_1(3) \to \mathrm{Tmf}_1(n)$ correspond to a sequence of classes $x_i \in \pi^{C_2}_{n_i\rho}\mathrm{Tmf}_1(n)$ by considering the images of $1 \in \pi^{C_2}_{n_i\rho}\Sigma^{n_i\rho}\mathrm{Tmf}_1(3)$. Denote the sequence corresponding to Φ by $e_0, \dots, e_N$. By possibly reordering, we can assume $n_0 = 0$. We construct a new map $\Phi' \colon \bigoplus_{i=0}^N \Sigma^{n_i\rho}\mathrm{Tmf}_1(3) \to \mathrm{Tmf}_1(n)$ corresponding to $x_0, x_1, \dots, x_N$ with $x_i = e_i$ for $i > 0$, and x_0 corresponding to the image of $u \in \mathbb{Z}$ in (5.7), where u maps to $\mathrm{res}^{C_2}_e(e_0)$ along the isomorphism $\mathbb{Z} \cong \pi^e_0\mathrm{tmf}_1(n) \to \pi^e_0\mathrm{Tmf}_1(n)$. As Φ' and Φ induce the same map on underlying homotopy groups, the map Φ' is an equivalence. By construction, $r(x_0) = 0$.

Thus the map

$$\bigoplus_i \Sigma^{n_i\rho}\mathrm{tmf}_1(3) \to \bigoplus_i \tau_{\geq 0}\Sigma^{n_i\rho}\mathrm{tmf}_1(3) \xrightarrow{\tau_{\geq 0}\Phi'} \tau_{\geq 0}\mathrm{Tmf}_1(n)$$

factors indeed over $\mathrm{tmf}_1(n)$. As before, the map $\Sigma^{n_i\rho}\mathrm{tmf}_1(3) \to \mathrm{tmf}_1(n)$ induces an isomorphism on underlying homotopy groups. Both source and target are strongly even and thus the map is a C_2–equivariant equivalence by [18, Lemma 3.4]. □

References

[1] **M Ando**, **M J Hopkins**, **C Rezk**, *Multiplicative orientations of KO–theory and of the spectrum of topological modular forms*, preprint (2010) Available at https://rezk.web.illinois.edu/koandtmf.pdf

[2] **S Araki**, **K Iriye**, *Equivariant stable homotopy groups of spheres with involutions, I*, Osaka Math. J. 19 (1982) 1–55 MR Zbl

[3] **M Behrens**, **M Hill**, **M J Hopkins**, **M Mahowald**, *Detecting exotic spheres in low dimensions using* coker J, J. Lond. Math. Soc. 101 (2020) 1173–1218 MR Zbl

[4] **A K Bousfield**, *Homotopy spectral sequences and obstructions*, Israel J. Math. 66 (1989) 54–104 MR Zbl

[5] **B Conrad**, *Arithmetic moduli of generalized elliptic curves*, J. Inst. Math. Jussieu 6 (2007) 209–278 MR Zbl

[6] **P Deligne**, **M Rapoport**, *Les schémas de modules de courbes elliptiques*, from "Modular functions of one variable, II" (P Deligne, W Kuyk, editors), Lecture Notes in Math. 349, Springer (1973) 143–316 MR Zbl

[7] **F Diamond**, **J Im**, *Modular forms and modular curves*, from "Seminar on Fermat's last theorem" (V K Murty, editor), CMS Conf. Proc. 17, Amer. Math. Soc., Providence, RI (1995) 39–133 MR Zbl

[8] **F Diamond**, **J Shurman**, *A first course in modular forms*, Graduate Texts in Math. 228, Springer (2005) MR Zbl

[9] **C L Douglas**, **J Francis**, **A G Henriques**, **M A Hill** (editors), *Topological modular forms*, Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) MR Zbl

[10] **D Eisenbud**, *Commutative algebra*, Graduate Texts in Math. 150, Springer (1995) MR Zbl

[11] **J Franke**, *On the construction of elliptic cohomology*, Math. Nachr. 158 (1992) 43–65 MR Zbl

[12] **P G Goerss**, *Quasi-coherent sheaves on the moduli stack of formal groups*, preprint (2008) arXiv 0802.0996

[13] **P G Goerss**, **M J Hopkins**, *Moduli spaces of commutative ring spectra*, from "Structured ring spectra" (A Baker, B Richter, editors), London Math. Soc. Lecture Note Ser. 315, Cambridge Univ. Press (2004) 151–200 MR Zbl

[14] **R Hartshorne**, *Algebraic geometry*, Graduate Texts in Math. 52, Springer (1977) MR Zbl

[15] **R Herrera**, *Elliptic genera of level N on complex π_2–finite manifolds*, C. R. Math. Acad. Sci. Paris 344 (2007) 317–320 MR Zbl

[16] **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

[17] **M Hill**, **T Lawson**, *Topological modular forms with level structure*, Invent. Math. 203 (2016) 359–416 MR Zbl

[18] **M A Hill**, **L Meier**, *The C_2–spectrum $\mathrm{Tmf}_1(3)$ and its invertible modules*, Algebr. Geom. Topol. 17 (2017) 1953–2011 MR Zbl

[19] **F Hirzebruch**, *Elliptic genera of level N for complex manifolds*, from "Differential geometrical methods in theoretical physics" (K Bleuler, M Werner, editors), NATO Adv. Sci. Inst. Ser. C: Math. Phys. Sci. 250, Kluwer, Dordrecht (1988) 37–63 MR Zbl

[20] **F Hirzebruch**, **T Berger**, **R Jung**, *Manifolds and modular forms*, Aspects of Mathematics E20, Friedr. Vieweg & Sohn, Braunschweig (1992) MR Zbl

[21] **E Höning**, **B Richter**, *Detecting and describing ramification for structured ring spectra*, preprint (2021) arXiv 2101.12655

[22] **M J Hopkins**, **T Lawson**, *Strictly commutative complex orientation theory*, Math. Z. 290 (2018) 83–101 MR Zbl

[23] **M J Hopkins**, **M Mahowald**, *From elliptic curves to homotopy theory*, preprint (2000) Available at `https://tinyurl.com/HopkinsMahowald2000`

[24] **P Hu**, **I Kriz**, *Real-oriented homotopy theory and an analogue of the Adams–Novikov spectral sequence*, Topology 40 (2001) 317–399 MR Zbl

[25] **D C Isaksen**, **G Wang**, **Z Xu**, *Stable homotopy groups of spheres: from dimension* 0 *to* 90, preprint (2020) arXiv 2001.04511

[26] **N M Katz**, **B Mazur**, *Arithmetic moduli of elliptic curves*, Annals of Mathematics Studies 108, Princeton Univ. Press (1985) MR Zbl

[27] **I M Krichever**, *Generalized elliptic genera and Baker–Akhiezer functions*, Mat. Zametki 47 (1990) 34–45 MR Zbl In Russian; translated in Math. Notes 47 (1990) 132–142

[28] **N J Kuhn**, *Adams filtration and generalized Hurewicz maps for infinite loopspaces*, Invent. Math. 214 (2018) 957–998 MR Zbl

[29] **T Lawson**, *The Shimura curve of discriminant 15 and topological automorphic forms*, Forum Math. Sigma 3 (2015) art. id. e3 MR Zbl

[30] **L G Lewis, Jr**, *The equivariant Hurewicz map*, Trans. Amer. Math. Soc. 329 (1992) 433–472 MR Zbl

[31] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl

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

[33] **M Mahowald**, **C Rezk**, *Brown–Comenetz duality and the Adams spectral sequence*, Amer. J. Math. 121 (1999) 1153–1177 MR Zbl

[34] **A Mathew**, *The homology of tmf*, Homology Homotopy Appl. 18 (2016) 1–29 MR Zbl

[35] **A Mathew**, **L Meier**, *Affineness and chromatic homotopy theory*, J. Topol. 8 (2015) 476–528 MR Zbl

[36] **L Meier**, *Additive decompositions for rings of modular forms*, Doc. Math. 27 (2022) 427–488 MR Zbl

[37] **L Meier**, *Topological modular forms with level structure: decompositions and duality*, Trans. Amer. Math. Soc. 375 (2022) 1305–1355 MR Zbl

[38] **L Meier**, **V Ozornova**, *Rings of modular forms and a splitting of* $\mathrm{TMF}_0(7)$, Selecta Math. 26 (2020) art. id. 7 MR Zbl

[39] **J S Milne**, *Étale cohomology*, Princeton Mathematical Series No. 33, Princeton Univ. Press (1980) MR Zbl

[40] **P Pstrągowski**, **P VanKoughnett**, *Abstract Goerss–Hopkins theory*, Adv. Math. 395 (2022) art. id. 108098 MR Zbl

[41] **C Rezk**, *Supplementary notes for Math* 512 (*version* 0.18), lecture notes (2002) Available at `https://rezk.web.illinois.edu/512-spr2001-notes.pdf`

[42] **E Riehl**, **D Verity**, *Homotopy coherent adjunctions and the formal theory of monads*, Adv. Math. 286 (2016) 802–888 MR Zbl

[43] **A Senger**, *Obstruction theory and the level n elliptic genus*, Compos. Math. 159 (2023) 2000–2021 MR Zbl

[44] **V Stojanoska**, *Duality for topological modular forms*, Doc. Math. 17 (2012) 271–311 MR Zbl

[45] **K Česnavičius**, *A modular description of $\mathcal{X}_0(n)$*, Algebra Number Theory 11 (2017) 2001–2089 MR Zbl

[46] **G Wang**, **Z Xu**, *The triviality of the 61–stem in the stable homotopy groups of spheres*, Ann. of Math. 186 (2017) 501–580 MR Zbl

[47] **J Wang**, **Y Wang**, **Z Yu**, *Elliptic genera of level N for complete intersections*, Front. Math. China 16 (2021) 1043–1062 MR Zbl

[48] **E Witten**, *The index of the Dirac operator in loop space*, from "Elliptic curves and modular forms in algebraic topology" (P S Landweber, editor), Lecture Notes in Math. 1326, Springer (1988) 161–181 MR Zbl

Mathematical Institute, Utrecht University
Utrecht, Netherlands

f.l.m.meier@uu.nl

Received: 13 May 2021 Revised: 4 July 2022

Algebraic & Geometric Topology **23**:8 (2023) 3587–3613
DOI: 10.2140/agt.2023.23.3587
Published: 5 November 2023

Asymptotic dimension of graphs of groups and one-relator groups

PANAGIOTIS TSELEKIDIS

We prove a new inequality for the asymptotic dimension of HNN-extensions. We deduce that the asymptotic dimension of every finitely generated one-relator group is at most two, confirming a conjecture of A Dranishnikov. As corollaries we calculate the exact asymptotic dimension of right-angled Artin groups and we give a new upper bound for the asymptotic dimension of fundamental groups of graphs of groups.

20E08, 20F65

1 Introduction

In 1993, M Gromov introduced the notion of the asymptotic dimension of metric spaces (see [12]) as an invariant of finitely generated groups. It can be shown that if two metric spaces are quasi-isometric then they have the same asymptotic dimension. The asymptotic dimension asdim X of a metric space X is defined by: asdim $X \le n$ if and only if, for every $R > 0$, there exists a uniformly bounded covering $\mathcal{U}$ of X such that the R–multiplicity of $\mathcal{U}$ is smaller than or equal to $n+1$ (ie every R–ball in X intersects at most $n+1$ elements of $\mathcal{U}$). There are many equivalent ways to define the asymptotic dimension of a metric space. It turns out that the asymptotic dimension of an infinite tree is 1 and the asymptotic dimension of $\mathbb{E}^n$ is n.

In 1998, the asymptotic dimension achieved particular prominence in geometric group theory after the publication of a paper of Guoliang Yu (see [24]) which proved the Novikov higher signature conjecture for manifolds whose fundamental group has finite asymptotic dimension. Unfortunately, not all finitely presented groups have finite asymptotic dimension. For example, Thompson's group F has infinite asymptotic dimension since it contains $\mathbb{Z}^n$ for all n. However, we know for many classes of groups that they have finite asymptotic dimension. For instance, hyperbolic, relative hyperbolic, mapping class groups of surfaces and one-relator groups have finite asymptotic dimension (see G Bell and A Dranishnikov [3], Bestvina, Bromberg and Fujiwara [6], Osin [18] and D Matsnev [17]). The exact computation of the asymptotic dimension of groups or finding the optimal upper bound is more delicate. Another remarkable result is that of Buyalo and Lebedeva (see [7]), where in 2006 they established the equality, for hyperbolic groups,

$$\operatorname{asdim} G = \dim \partial_\infty G + 1.$$

The inequalities of Bell and Dranishnikov (see [2; 9]) play a key role in finding an upper bound for the asymptotic dimension of groups. However, in some cases the upper bounds that the inequalities of Bell and Dranishnikov provide us are quite far from being optimal. An example is the asymptotic dimension of one-relator groups.

We prove some new inequalities that can be a useful tool for the computation of the asymptotic dimension of groups. As an application we give the optimal upper bound for the asymptotic dimension of one-relator groups which was conjectured by Dranishnikov. As a further corollary we calculate the exact asymptotic dimension of any right-angled Artin group (Theorem 1.2) — this has been proven earlier by N Wright [23] by different methods.

The first inequality and one of the main results we prove is the following:

Theorem 1.1 *Let $G*_N$ be an HNN-extension of the finitely generated group G over N. Then*

$$\operatorname{asdim} G*_N \leq \max\{\operatorname{asdim} G, \operatorname{asdim} N + 1\}.$$

Next, we calculate the asymptotic dimension of the right-angled Artin groups. To be more precise, let Γ be a finite simplicial graph. We denote by $A(\Gamma)$ the *right-angled Artin group* (RAAG) associated to the graph Γ. We set

$$\operatorname{Sim}(\Gamma) = \max\{n : \Gamma \text{ contains the 1–skeleton of the standard } (n-1)\text{–simplex } \Delta^{n-1}\}.$$

Then by applying Theorem 1.1 we obtain the following:

Theorem 1.2 *Let Γ be a finite simplicial graph. Then*

$$\operatorname{asdim} A(\Gamma) = \operatorname{Sim}(\Gamma).$$

In 2005, Bell and Dranishnikov (see [4]) gave a proof that the asymptotic dimension of one-relator groups is finite and also gave an upper bound, namely the length of the relator plus one. Let $G = \langle S \mid r \rangle$ be a finitely generated one-relator group such that $|r| = n$. Then

$$\operatorname{asdim} G \leq n + 1.$$

To prove this upper bound, Bell and Dranishnikov used an inequality for the asymptotic dimension of HNN-extensions; see [2]. In particular, let G be a finitely generated group and let N be a subgroup of G. Then

$$\operatorname{asdim} G *_N \leq \operatorname{asdim} G + 1.$$

In 2006, Matsnev (see [17]) proved a sharper upper bound for the asymptotic dimension of one-relator groups: if $G = \langle S \mid r \rangle$ is a one-relator group, then

$$\operatorname{asdim} G \leq \left\lceil \tfrac{1}{2} \operatorname{length}(r) \right\rceil.$$

Here by $\lceil a \rceil$ ($a \in \mathbb{R}$) we denote the minimal integer greater than or equal to a.

Applying Theorem 1.1, we answer a conjecture of Dranishnikov (see [8]) giving the optimal upper bound for the asymptotic dimension of one-relator groups.

Theorem 1.3 *Let G be a finitely generated one-relator group. Then*

$$\operatorname{asdim} G \leq 2.$$

We note that R C Lyndon (see [14]) has shown that the *cohomological dimension* of a torsion-free one-relator group is smaller than or equal to 2. Our result can be seen as a large-scale analog of this. We note that the large-scale geometry of one-relator groups can be quite complicated; for example, one-relator groups can have very large isoperimetric functions; see eg Platonov [19].

It is worth noting that L Sledd showed that the Assouad–Nagata dimension of any finitely generated $C'\left(\frac{1}{6}\right)$ group is at most two; see [20].

Theorem 1.3 combined with the results of M Kapovich and B Kleiner (see [13]) leads us to a description of the boundary of hyperbolic one-relator groups.

We determine also the one-relator groups that have asymptotic dimension exactly two. We prove that every infinite finitely generated one-relator group G that is not a free group or a free product of a free group and a finite cyclic group has asymptotic dimension equal to 2 (Proposition 3.5). We obtain the following:

Corollary *Let G be finitely generated freely indecomposable one-relator group which is not cyclic. Then*

$$\operatorname{asdim} G = 2.$$

Moreover, we describe the finitely generated one-relator groups:

Corollary *Let G be a finitely generated one-relator group. Then one of the following is true*:

(i) *G is finite cyclic*, *and* $\operatorname{asdim} G = 0$.

(ii) *G is a nontrivial free group or a free product of a nontrivial free group and a finite cyclic group*, *and* $\operatorname{asdim} G = 1$.

(iii) *G is an infinite freely indecomposable not cyclic group or a free product of a nontrivial free group and an infinite freely indecomposable not cyclic group*, *and* $\operatorname{asdim} G = 2$.

Using Theorem 1.1 and an inequality of Dranishnikov about the asymptotic dimension of amalgamated products (see [9]) we obtain a more general theorem for the asymptotic dimension of fundamental groups of graphs of groups:

Theorem 1.4 *Let $(\mathbb{G}, Y)$ be a finite graph of groups with vertex groups $\{G_v : v \in Y^0\}$ and edge groups $\{G_e : e \in Y^1_+\}$. Then*

$$\operatorname{asdim} \pi_1(\mathbb{G}, Y, \mathbb{T}) \leq \max_{v \in Y^0, e \in Y^1_+} \{\operatorname{asdim} G_v, \operatorname{asdim} G_e + 1\}.$$

Using the previous theorem, we can obtain, for example, that the asymptotic dimension of a graph of surface groups (with genus ≥ 2) with free edge groups is two. Theorem 1.4 says that the asymptotic dimension doesn't *jump* as long as there exists a vertex group with asymptotic dimension greater than the asymptotic dimension of any edge group.

The paper is organized as follows. In Section 2 we prove the inequality for the asymptotic dimension of HNN-extensions. In Section 2.1 we compute the asymptotic dimension of RAAGs. Next, in Section 3 we give the optimal upper bound for the asymptotic dimension of one-relator groups. In Section 3.1 we describe the one-relator groups with asymptotic dimension 0, 1 and 2. In Section 4 a new upper bound for the asymptotic dimension of graphs of groups is obtained.

Acknowledgments I would like to thank Panos Papasoglu for his valuable advice during the development of this research work. I owe my deepest gratitude to anonymous referee(s) for their time spent reviewing this article. I would also like to offer my special thanks to Mark Hagen and Richard Wade for their very useful comments.

2 Asymptotic dimension of HNN-extensions

Let X be a metric space and $\mathcal{U}$ a covering of X. We say that the covering $\mathcal{U}$ is d–*bounded* or d–*uniformly bounded* if $\sup_{U\in\mathcal{U}}\{\operatorname{diam} U\} \le d$. The *Lebesgue number* $L(\mathcal{U})$ of the covering $\mathcal{U}$ is

$$L(\mathcal{U}) = \sup\{\lambda : \text{if } A \subseteq X \text{ with } \operatorname{diam} A \le \lambda \text{ then there exists } U \in \mathcal{U} \text{ such that } A \subseteq U\}.$$

We recall that the order $\operatorname{ord}(\mathcal{U})$ of the cover $\mathcal{U}$ is the smallest number n (if it exists) such that each point of the space belongs to at most n sets in the cover. For a metric space X, we say that (r,d)–$\dim X \le n$ if, for $r > 0$, there exists a d–bounded cover $\mathcal{U}$ of X with $\operatorname{ord}(\mathcal{U}) \le n+1$ and with Lebesgue number $L(\mathcal{U}) > r$. We refer to such a cover as an (r,d)–*cover* of X. The following proposition is due to Bell and Dranishnikov (see [2]):

Proposition 2.1 *For a metric space X, $\operatorname{asdim} X \le n$ if and only if there exists a function $d(r)$ such that $(r, d(r))$–$\dim X \le n$ for all $r > 0$.*

We recall that the family X_i of subsets of X satisfies the inequality $\operatorname{asdim} X_i \le n$ *uniformly* if, for every $R > 0$, there exists a D–bounded covering $\mathcal{U}_i$ of X_i with R–$\operatorname{mult}(\mathcal{U}_i) \le n+1$ for every i. For the proofs of Theorems 2.2 and 2.3 see [1].

Theorem 2.2 (infinite union theorem) *Let $X = \bigcup_a X_a$ be a metric space where the family $\{X_a\}$ satisfies the inequality $\operatorname{asdim} X_a \le n$ uniformly. Suppose further that, for every $r > 0$, there is a subset $Y_r \subseteq X$ with $\operatorname{asdim} Y_r \le n$, so that $d(X_a \setminus Y_r, X_b \setminus Y_r) \ge r$ whenever $X_a \ne X_b$. Then $\operatorname{asdim} X \le n$.*

Theorem 2.3 (finite union theorem) *For every metric space presented as a finite union $X = \bigcup_i X_i$,*

$$\operatorname{asdim} X = \max\{\operatorname{asdim} X_i\}.$$

A partition of a metric space X is a presentation as a union $X = \bigcup_i W_i$ such that $\operatorname{Int}(W_i) \cap \operatorname{Int}(W_j) = \varnothing$ whenever $i \ne j$. We denote by ∂W_i the topological boundary of W_i and by $\operatorname{Int}(W_i)$ the topological interior. We have that $\partial W \cap \operatorname{Int}(W) = \varnothing$. The boundary can be written as

$$\partial W_i = \{x \in X : d(x, W_i) = d(x, X \setminus W_i) = 0\}.$$

For the proof of the following theorem see [9]:

Theorem 2.4 (partition theorem) *Let X be a geodesic metric space. Suppose that for every $R > 0$ there is $d > 0$ and a partition $X = \bigcup_i W_i$ with $\operatorname{asdim} W_i \le n$ uniformly in i and such that (R, d)–$\dim\bigl(\bigcup_i \partial W_i\bigr) \le n - 1$, where ∂W_i is taken with the metric restricted from X. Then $\operatorname{asdim} X \le n$.*

Let G be a finitely generated group, N a subgroup of G and $\phi : N \to G$ a monomorphism. We set $\overline{G} = G*_N$, the HNN-extension of G over the subgroup N with respect to the monomorphism ϕ. We fix a finite generating set S for the group G. Then the set $\overline{S} = S \cup \{t, t^{-1}\}$ is a finite generating set for the group $\overline{G}$ and we set $C(\overline{G}) = \operatorname{Cay}(\overline{G}, \overline{S})$, its Cayley graph.

Normal forms for HNN-extensions There are two types of *normal forms* for HNN-extensions: the right normal form and the left normal form. We use both.

Right normal form Let S_N and $S_{\phi(N)}$ be sets of representatives of *right cosets* of G/N and of $G/\phi(N)$, respectively. Then every $w \in \overline{G}$ has a unique normal form $w = g t^{\epsilon_1} s_1 t^{\epsilon_2} s_2 \cdots t^{\epsilon_k} s_k$ where $g \in G$, $\epsilon_i \in \{-1, 1\}$, if $\epsilon_i = 1$ then $s_i \in S_N$, if $\epsilon_i = -1$ then $s_i \in S_{\phi(N)}$, and if $s_i = 1$ then $\epsilon_i \epsilon_{i+1} > 0$. We say that the *length* of the right normal form of w is k.

Left normal form Let ${}_N S$ and ${}_{\phi(N)} S$ be sets of representatives of *left cosets* of G/N and of $G/\phi(N)$, respectively. Then every $w \in \overline{G}$ has a unique normal form $w = s_1 t^{\epsilon_1} s_2 t^{\epsilon_2} \cdots s_k t^{\epsilon_k} g$ where $g \in G$, $\epsilon_i \in \{-1, 1\}$, if $\epsilon_i = 1$ then $s_i \in_{\phi(N)} S$, if $\epsilon_i = -1$ then $s_i \in_N S$, and if $s_i = 1$ then $\epsilon_{i-1}\epsilon_i > 0$. We say that the *length* of the left normal form of w is k.

We observe that the *lengths* of the right and the left normal form of an element coincide, and denote this length by $l(w)$.

Convention When we write a normal form we mean the right normal form, unless otherwise stated.

The group $\overline{G} = G*_N$ acts on its Bass–Serre tree T. There is a natural projection $\pi : G*_N \to T$ defined by the action: $\pi(g) = gG$.

Lemma 2.5 *The map $\pi : \overline{G} \to T$ extends to a simplicial map from the Cayley graph, $\pi : C(\overline{G}, S) \to T$, which is 1–Lipschitz.*

Proof Let $g \in \overline{G}$ and $s \in \overline{S}$. Then the vertex g is mapped to the vertex $\pi(g) = \pi(gs) = gG$. If $s \in S$, then the edge $[g, gs]$ is mapped to the vertex $\pi(g) = \pi(gs) = gG$.

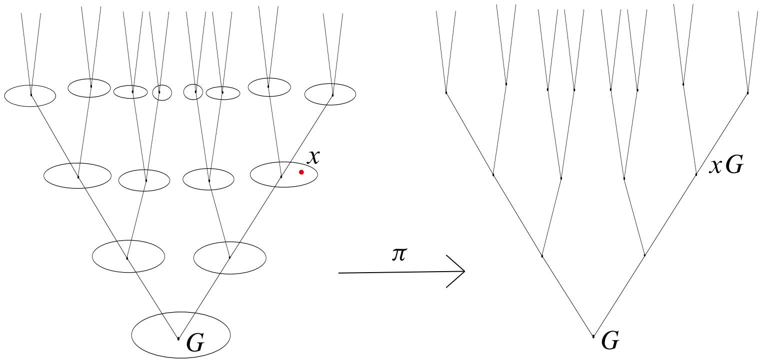


Figure 1: An illustration of the projection $\pi : C(\overline{G}, S) \to T$.

If $s \in \{t, t^{-1}\}$, without loss of generality we may assume that $s = t$, and so the edge $[g, gs]$ is mapped to the edge $[\pi(g), \pi(gs)] = [gG, gtG]$ of T.

We observe that the simplicial map $\pi : C(\overline{G}) \to T$ is 1–Lipschitz. $\square$

The base vertex G separates T into two parts, $T_- \setminus G$ and $T_+ \setminus G$, where

$$\pi^{-1}(T_+) = \{w \in \overline{G} : \text{if } w = gt^{\epsilon_1} s_1 t^{\epsilon_2} s_2 \cdots t^{\epsilon_k} s_k \text{ is the normal form of } w \text{ then } \epsilon_1 = 1\}$$

and similarly

$$\pi^{-1}(T_-) = \{w \in \overline{G} : \text{if } w = gt^{\epsilon_1} s_1 t^{\epsilon_2} s_2 \cdots t^{\epsilon_k} s_k \text{ is the normal form of } w \text{ then } \epsilon_1 = -1\}.$$

We note that both $T_+ \setminus G$ and $T_- \setminus G$ are unions of connected components of T and $\pi^{-1}(T_+)$ and $\pi^{-1}(T_-)$ are unions of connected components of $C(\overline{G})$. See Figure 2 for an illustration of T_- and T_+.

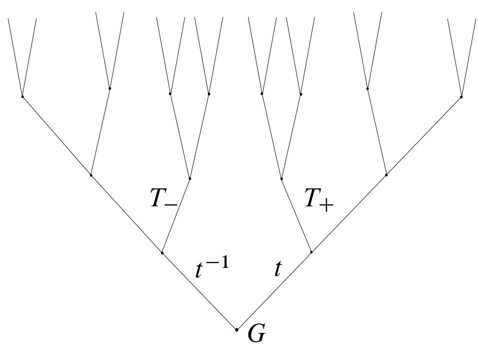


Figure 2: An illustration of T_+ and T_-.

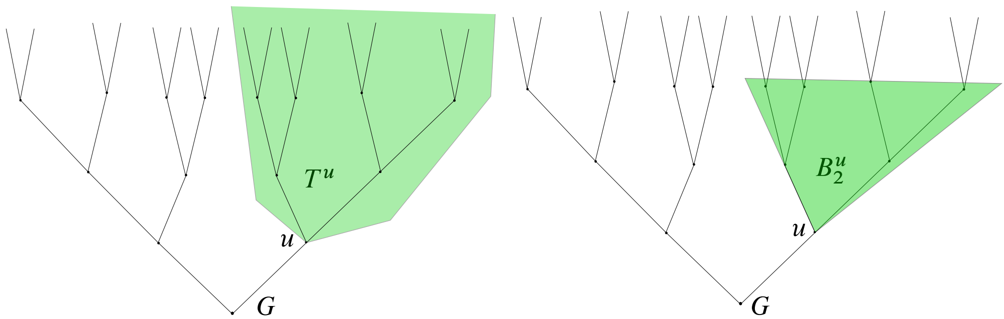


Figure 3: Left: an illustration of T^u. Right: an illustration of B_r^u, where $r = 2$.

We consider the Bass–Serre tree T as a metric space with the simplicial metric $\bar{d}$. If Y is a graph, we denote by Y^0 or $V(Y)$ the vertices of Y. For $u \in T^0$ we denote by $|u|$ the distance to the vertex with label G. We note that the distance of the vertex wG from G in the Bass–Serre tree T equals the length $l(w)$ of the normal form of w, $|wG| = l(w)$.

We recall that a *full subgraph* of a graph Γ is a subgraph formed from a subset of vertices V and from all of the edges that have both endpoints in the subset V. If A is a subgraph of Γ we define the *edge closure* $E(A)$ of A to be the full subgraph of Γ formed from $V(A)$. Obviously, $V(E(A)) = V(A)$.

We fix some notation on the Bass–Serre tree T and on the Cayley graph.

In the tree T We denote by B_r^T the r–ball in T centered at G ($r \in \mathbb{N}$). There is a partial order on vertices of T defined by setting $v \leq u$ if and only if v lies in the geodesic segment $[G, u]$ joining the base vertex G with u. For $u \in T^0$ of nonzero level (ie $u \neq G$) and $r > 0$, we set

$$T^u = E(\{v \in T^0 : u \leq v\}), \quad B_r^u = E(\{v \in T^u : |v| \leq |u| + r\}).$$

For every vertex $u \in T^0$ represented by a coset $g_u G$, we have $B_r^u = g_u B_r^T \cap T^u$. We also observe that $B_r^u = E(\{v \in T^u : \bar{d}(v, u) \leq r\})$. See Figure 3 for an illustration of the sets T^u and B_r^u

In the Cayley graph For $R \in \mathbb{N}$, let

$$M_R = \{g \in \overline{G} : \operatorname{dist}(g, N \cup \phi(N)) = R\}.$$

Letting $u = g_u G$, we set $M_R^u = g_u M_R \cap \pi^{-1}(T^u)$. We observe that $\pi(M_R^u) \subseteq B_R^u$ since π is 1–Lipschitz.

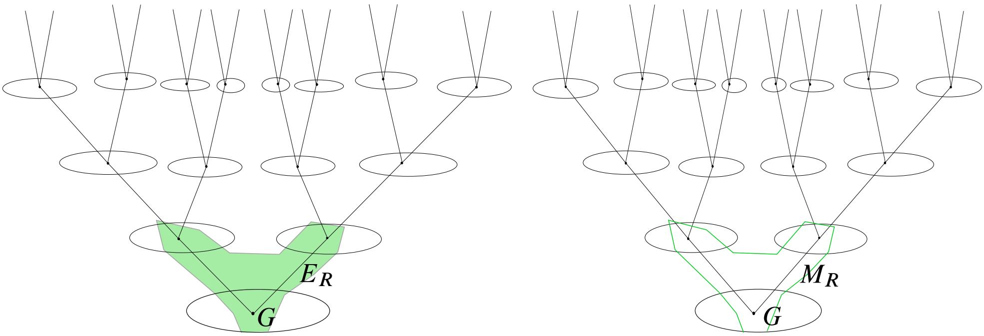


Figure 4: Left: an illustration of E_R. Right: an illustration of M_R.

Letting $u = g_u G$, we set $E_R = E\big(N_R(N \cup \phi(N))\big)$ and

$$E_R^u = g_u E_R \cap \pi^{-1}(T^u).$$

Obviously, $M_R^u \subseteq E_R^u \subseteq \pi^{-1}(B_R^u)$.

Convention We associate every $u \in T^0$ to an element $g_u \in \overline{G}$ such that

(i) $u = g_u G$, and

(ii) if the left normal form of g_u is $s_1 t^{\epsilon_1} s_2 t^{\epsilon_2} \cdots s_k t^{\epsilon_k} g$ then $g = 1_{\overline{G}}$.

We see that in this way we may define a bijective map from T^0 to the set $\mathcal{G}_T$ which consists of the elements of $\overline{G}$ such that conditions (i) and (ii) hold.

Proposition 2.6 *If $4 < 4R \le r$ and the distinct vertices $u, u' \in T^0$, satisfy $|u|, |u'| \in \{nr : n \in \mathbb{N}\}$, then*

$$d(M_R^u, M_R^{u'}) \ge 2R.$$

Proof We distinguish two cases. See the left and right parts of Figure 5 for cases 1 and 2, respectively.

Case 1 ($|u| \ne |u'|$) Recall that every path γ in $C(\overline{G})$ projects to a path $\pi(\gamma)$ in the tree T. Then, since

$$M_R^u = g_u M_R \cap \pi^{-1}(T^u) \subseteq \pi^{-1}(B_R^u), \quad M_R^{u'} = g_{u'} M_R \cap \pi^{-1}(T^{u'}) \subseteq \pi^{-1}(B_R^{u'})$$

and π is 1–Lipschitz,

$$d(M_R^u, M_R^{u'}) \ge \bar{d}(B_R^u, B_R^{u'}) \ge r - R \ge 3R.$$

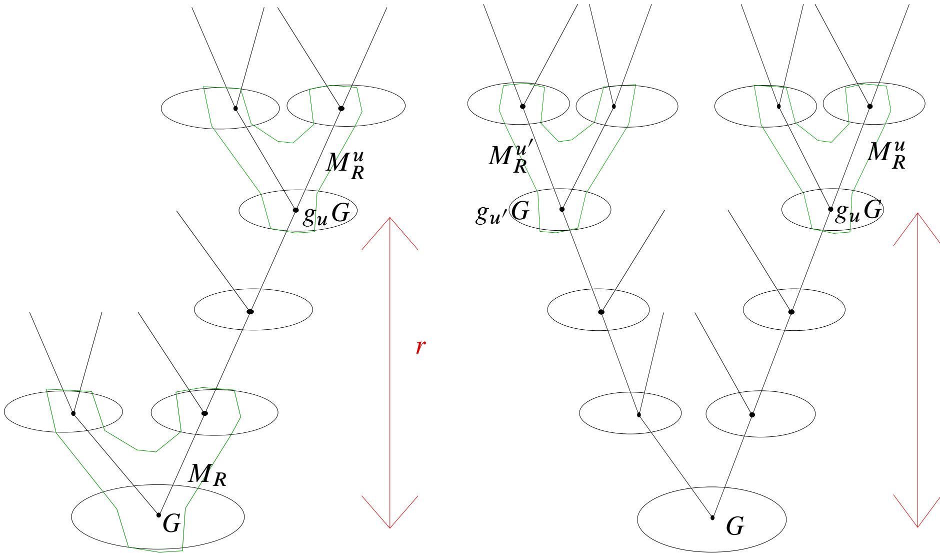


Figure 5: Left: an illustration of Case 1 of Proposition 2.6, where $u' = G$. Right: an illustration of Case 2 of Proposition 2.6.

Case 2 $(|u| = |u'|$ with $u \neq u')$ Denote by ζ_0 the last vertex of the common geodesic segment $[G, \zeta_0]$ of the geodesics $[G, u]$ and $[G, u']$. Observe that $\bar{d}(u, \zeta_0), \bar{d}(u', \zeta_0) \geq 1$. Let $x \in M_R^u$, $y \in M_R^{u'}$ and let γ be a geodesic from x to y. Then the path $\pi(\gamma)$ passes

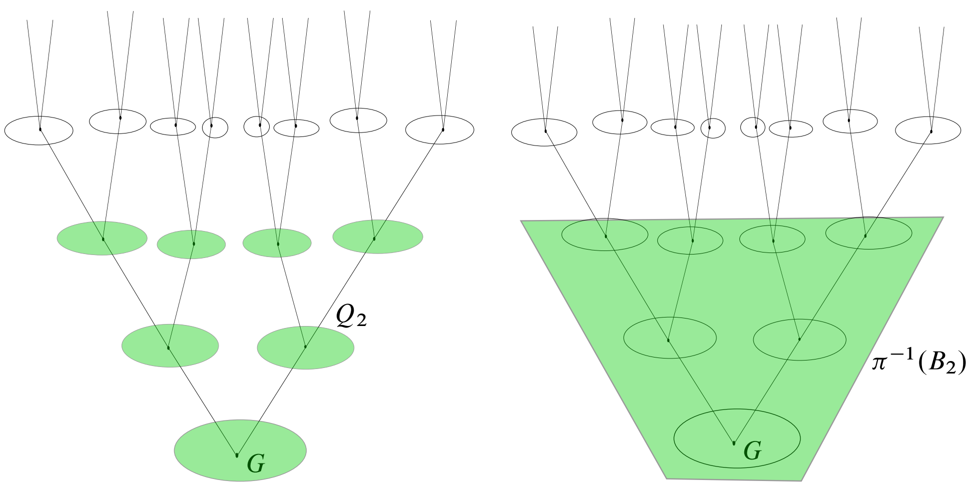


Figure 6: Left: an illustration of Q_m, for $m = 2$ (Proposition 2.8). We note that $Q_m = V(\pi^{-1}(B_r))$. Right: an illustration of $\pi^{-1}(B_r)$, where $r = 2$.

through the vertices u, u' and ζ_0, so the geodesic γ intersects both $g_u(N \cup \phi(N))$ and $g_{u'}(N \cup \phi(N))$. Hence

$$\begin{aligned} d(x,y) &\geq \operatorname{dist}\bigl(x, g_u(N \cup \phi(N))\bigr) + \operatorname{dist}\bigl(y, g_{u'}(N \cup \phi(N))\bigr) + \operatorname{length}([\zeta_0, u']) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad + \operatorname{length}([\zeta_0, u]) \\ &\geq R + R + 2 = 2(R+1). \end{aligned}$$

□

For $w \in G*_N$, we denote by $\|w\|$ the distance from w to $1_{\overline{G}}$ in the Cayley graph $\operatorname{Cay}(\overline{G}, \overline{S})$.

Lemma 2.7 *Let $w = gt^{\epsilon_1}s_1t^{\epsilon_2}s_2\cdots t^{\epsilon_k}s_k$ be the normal form of w. Then*

$$\|w\| \geq d(s_k, N) \quad \text{if } \epsilon_k = 1 \qquad \text{and} \qquad \|w\| \geq d(s_k, \phi(N)) \quad \text{if } \epsilon_k = -1.$$

Proof Without loss of generality we assume that $\epsilon_k = 1$. Let

$$w = \Bigl(\prod_{i_0=1}^{m_0} s_{i_0}\Bigr)t^{\epsilon_1}\Bigl(\prod_{i_1=1}^{m_1} s_{i_1}\Bigr)t^{\epsilon_2}\cdots t\Bigl(\prod_{i_k=1}^{m_k} s_{i_k}\Bigr)$$

be a shortest presentation of w in the alphabet $\overline{S}$ (we note that $s_{i_j} \notin \{t, t^{-1}\}$). We set $\prod_{i_j=1}^{m_j} s_{i_j} = g_j$ for every $j \in \{1, \dots, k\}$. Then $w = gt^{\epsilon_1}g_1t^{\epsilon_2}s_2\cdots tg_k = w_0tg_k$. The first step when we rewrite w in normal form starting from the previous presentation is to write $g_k = ns_k$ (where $n \in N$). Then

$$\|w\| \geq \|g_k\| = \|ns_k\| = d(ns_k, 1) = d(s_k, n^{-1}) \geq d(s_k, N).$$

□

We note that there exists an amalgamated product analog of the following proposition, proved by Dranishnikov in [9]:

Proposition 2.8 *Suppose that* asdim $G \leq n$. *Let*

$$Q_m = \{w \in \overline{G} : l(w) \leq m\}.$$

Then asdim $Q_m \leq n$, *for every $m \in \mathbb{N}$.*

Proof We set $P_\lambda = \{w \in \overline{G} : l(w) = \lambda\}$. It is enough to show that asdim $P_\lambda \leq n$, for every $\lambda \in \mathbb{N}$. Indeed, since

$$Q_m = \bigcup_{i=0}^{m} P_i,$$

by the finite union theorem we obtain that asdim $Q_m \leq n$.

Claim *For $\lambda \in \mathbb{N}$ we have* asdim $P_\lambda \leq n$.

Proof We use induction on λ. We have $P_0 = G$, so asdim $P_0 \le n$. We observe that $P_\lambda \subseteq P_{\lambda-1}tG \cup P_{\lambda-1}t^{-1}G$. Using the finite union theorem it suffices to show that $\operatorname{asdim}(P_\lambda \cap P_{\lambda-1}tG) \le n$ and $\operatorname{asdim}(P_\lambda \cap P_{\lambda-1}t^{-1}G) \le n$; we show the first.

To show that asdim $P_\lambda \cap P_{\lambda-1}tG \le n$, we use the infinite union theorem. For $r > 0$ we set $Y_r = P_{\lambda-1}tN_r(N)$. We claim that

$$Y_r \subseteq N_{r+1}(P_{\lambda-1}).$$

Indeed, if $z \in Y_r$ then $z = z_0tz_1$, where $z_0 \in P_{\lambda-1}$ and $z_1 \in N_r(N)$. Since $z_1 \in N_r(N)$, there exists $n \in N$ with $d(n, z_1) \le$ r. So $z = z_0tnn^{-1}z_1 = z_0\phi(n)tn^{-1}z_1$, and

$$d(z, P_{\lambda-1}) \le d(z, z_0\phi(n)) = \|tn^{-1}z_1\| \le \|t\| + \|t^{-1}z_1\| \le 1 + r.$$

Hence Y_r and $P_{\lambda-1}$ are quasi-isomorphic, so asdim $Y_r \le n$. □

We consider the family xtG where $x \in P_{\lambda-1}$. For $xtG \ne ytG$, we have

$$d(xtG \setminus Y_r, ytG \setminus Y_r) = d(xtg, yth) = \|g^{-1}t^{-1}x^{-1}yth\|,$$

where $g, h \in G \setminus N_r(N)$. The first step when we rewrite $g^{-1}t^{-1}x^{-1}yth$ in normal form is to make the substitution $h = ns_k$, where $n \in N$ and $s_k \in S_N$, so $g^{-1}t^{-1}x^{-1}yth = g^{-1}t^{-1}x^{-1}y\phi(n)ts_k$. Since $h \in G \setminus N_r(N)$, we have $\|s_k\| = \|n^{-1}h\| \ge d(h, N) \ge r$.

By Lemma 2.7 we obtain that $\|g^{-1}t^{-1}x^{-1}y\phi(n)ts_k\| \ge \|s_k\| \ge r$.

Finally, by observing that xtG and G are isometric, we deduce that $\operatorname{asdim}(xtG) \le n$ uniformly. Since all the conditions of the infinite union theorem hold,

$$\operatorname{asdim}(P_\lambda \cap P_{\lambda-1}tG) \le n$$

for every $\lambda \in \mathbb{N}$. □

We observe that $E(Q_m) = \pi^{-1}(B_m^T)$ and $Q_m = \overline{G} \cap \pi^{-1}(B_m^T)$.

For $w \in \overline{G}$, we set $T^w = T^{\pi(w)}$, where $\pi(w) = wG$.

We note that there was an attempt to prove the following theorem in [16], however, there is a gap in that proof. We give a few details about this gap right after the proof of Theorem 2.9.

Theorem 2.9 *Let $G*_N$ be an HNN-extension of the finitely generated group G over N. Then*

$$\operatorname{asdim} G*_N \le \max\{\operatorname{asdim} G, \operatorname{asdim} N + 1\}.$$

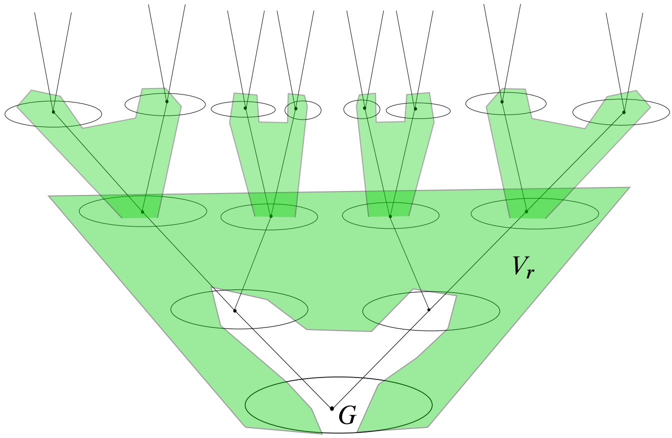


Figure 7: An illustration of V_r.

Proof Let $n = \max\{\operatorname{asdim} G, \operatorname{asdim} N + 1\}$. We denote by $\pi : C(\overline{G}, S) \to T$ the map of Lemma 2.5. We recall that we denote by $l(g)$ the length of the normal form of g.

We use the partition theorem (Theorem 2.4). Let $R, r \in \mathbb{N}$ be such that $R > 1$ and $r > 4R$.

We set

$$U_r = E\Bigg[\Big(\pi^{-1}(B^T_{r-1}) \cap E\big(\{g \in \overline{G} : d(g, N \cup \phi(N)) \geq R\}\big)\Big) \cup \Bigg(\bigcup_{u \in \partial B^T_r} E^u_R\Bigg)\Bigg],$$

where $E^u_R = g_u E\big(N_R(N \cup \phi(N))\big) \cap \pi^{-1}(T^u)$. We recall that

$$M_R = \{g \in \overline{G} : d(g, N \cup \phi(N)) = R\}.$$

Let A_R be the collection of the edges between the elements of $M_R \subseteq U_r$. We have that $A_R \subseteq U_r$. We define V_r to be the set obtained by removing the interior of the edges of A_R from U_r. Formally,

$$V_r = U_r \setminus \{\text{interior}(e) : e \in A_R\}.$$

See Figure 7 for an illustration of V_r. We observe that U_r and V_r are subgraphs of $C(\overline{G})$, $\partial U_r = \partial V_r$ and $V_r \cap \overline{G} = U_r \cap \overline{G}$. Obviously, $\bigcup_{u \in \partial B^T_r} E^u_R \subseteq V_r$. We also have

$$V_r \cap \overline{G} = \big(\overline{G} \cap \pi^{-1}(B^T_{r-1}) \cap E\big(\{g \in \overline{G} : d(g, N \cup \phi(N)) \geq R\}\big)\big) \cup \Bigg(\overline{G} \cap \bigcup_{u \in \partial B^T_r} E^u_R\Bigg).$$

To be more precise, $V_r \cap \overline{G}$ consists of those $wx \in \overline{G}$ such that $d(w, N \cup \phi(N)) \geq R$ and, if $w = g_0 t^{\epsilon_1} g_1 \cdots t^{\epsilon_k} g_k$ is the normal form of w, then either $k \leq r-1$, or $g_k = 1$ and $k = r$, while, if $x \neq 1$, then $k = r$, $g_k = 1$ and $d(x, N \cup \phi(N)) \leq R$.

For every vertex $u \in T^0$ satisfying $|u| \in \{nr : n \in \mathbb{N}\}$, we define

$$V_r^u = g_u V_r \cap \pi^{-1}(T^u).$$

Obviously, the sets V_r^u are subgraphs of $C(\overline{G})$ and $V_r^u \nsubseteq \overline{G}$. We observe that $V_r \subseteq \pi^{-1}(B_{r+R}^T)$, so $V_r^u \subseteq \pi^{-1}(B_{r+R}^u)$. Obviously, for every h such that $h = g_1 t^{\epsilon_1} g_2 \cdots t^{\epsilon_r}$ is the *left normal form* of h, $u \leq g_u h G$ and $|u| + r = |g_u h G|$, we have that

$$(g_u M_R \cap \pi^{-1}(T^{g_u G})) \cup (g_u h M_R \cap \pi^{-1}(T^{g_u h G})) \subseteq V_r^u.$$

We also observe that

$$g_u M_R \cap \pi^{-1}(T^{g_u G}) \subseteq \partial V_r^u,$$

which can also be written as

$$M_R^u = M_R^{g_u G} \subseteq \partial V_r^u.$$

We set $V_r^G = V_r$.

Consider the partition

$$C(\overline{G}) = \pi^{-1}(T) = \Bigg(\bigcup_{|u| \in \{nr : n \in \mathbb{N}_+ \cup \{0\}\}} V_r^u \Bigg) \cup E\big(N_R(N \cup \phi(N))\big). \tag{1}$$

We set

$$Z = \Bigg(\bigcup_{|u| \in \{nr : n \in \mathbb{N}_+\} \cup \{0\}} \partial V_r^u \Bigg) \cup \partial E\big(N_R(N \cup \phi(N))\big).$$

Observe that if $V_r^u \cap V_r^v \neq \varnothing$, then either $u \leq v$ and $|u| + r = |v|$ or $u \geq v$ and $|v| + r = |u|$. If $V_r^u \cap V_r^v \neq \varnothing$ is such that $u \leq v$ and $|u| + r = |v|$, then

$$V_r^u \cap V_r^v = M_R^v.$$

We deduce that

$$Z = \Bigg(\bigcup_{|u| \in \{nr : n \in \mathbb{N}_+\}} M_R^u \Bigg) \cup M_R.$$

We will show that there exists $d > 0$ such that (R,d)–dim $Z \leq n - 1$. Since M_R is quasi-isometric to $N_R(N \cup \phi(N))$, which is quasi-isometric to $N \cup \phi(N)$, we have that asdim $M_R \leq n - 1$. Then for $R > 0$ there exists an (R,d)–covering $\mathcal{U}$ of M_R with $\mathrm{ord}(\mathcal{U}) \leq n$. In view of Proposition 2.6, the covering

$$\mathcal{V} = \mathcal{U} \cup \bigcup_{|u| \in \{nr : n \in \mathbb{N}_+\}} (g_u \mathcal{U} \cap M_R^u)$$

is an (R,d)–covering of Z with $\mathrm{ord}(\mathcal{V}) \leq n$. We conclude that (R,d)–dim $Z \leq n - 1$.

Next, we will show that $\operatorname{asdim} V_r^u \le n$ and $\operatorname{asdim} N_R(N \cup \phi(N)) \le n$ uniformly. This will complete our proof, since all the conditions of the partition theorem are satisfied. It suffices to show that $\operatorname{asdim} V_r^u \le n$ uniformly and

$$\operatorname{asdim} N_R(N \cup \phi(N)) \le n.$$

We observe that $V_r \subseteq \pi^{-1}(B^T_{r+R}) \subseteq N_1(Q_{r+R})$, so by Proposition 2.8 we have that $\operatorname{asdim} V_r^u \le n$. Since there are at most two isometric classes for the sets V_r^u (for $u \ne G$) of our partition, we conclude that $\operatorname{asdim} V_r^u \le n$ uniformly. Finally, $\operatorname{asdim} N_R(N \cup \phi(N)) \le n-1$ since $N_R(N \cup \phi(N))$ is quasi-isometric to $N \cup \phi(N)$.

By the partition theorem (Theorem 2.4), $\operatorname{asdim} C(\overline{G}) = \operatorname{asdim} \pi^{-1}(T) \le n$. □

We now give a few details about the gap we found in [16]. We use the same notation as on pages 2276–2279 of [16]. We are not going to redefine all the symbols we use. We consider the HNN-extension $A{*}_C = \langle S_A \mid R_A, ct = tf(c)\rangle$ with respect to a monomorphism f. The idea was to construct a partition for $\pi^{-1}(K_1)$. The building block of this partition is the set $V_r = X_+ \cap \bigl(\bigcap_{|u|=r} X_-^u\bigr)$, where $u = g_u C$ are vertices of K_1.

The problem is that the set V_r is empty when the index $[A:C]$ is at least 2. One can find two vertices u and v of K_1 (where $|u| = |v| = r$) such that $X_-^v \cap X_-^u = \varnothing$. To see that, one must investigate how the dual graph K *behaves* under translations. For example, $K^u = g_u K_1$ when g_u (where u is a vertex of K_1) has a normal form ending with t, while $K^u = g_u K_0$ when g_u (where u is a vertex of K_1) has a normal form ending with t^{-1}.

2.1 Right-angled Artin groups

We use the following theorem of Bell, Dranishnikov and J Keesling; see [5].

Theorem 2.10 *If A and B are finitely generated groups then*

$$\operatorname{asdim} A * B = \max\{\operatorname{asdim} A, \operatorname{asdim} B\}.$$

Let Γ be a finite simplicial graph with n vertices. The *right-angled Artin group* (RAAG) $A(\Gamma)$ associated to the graph Γ has the presentation

$$A(\Gamma) = \langle \{s_u : u \in V(\Gamma)\} \mid \{[s_u, s_v] : [u,v] \in E(\Gamma)\}\rangle.$$

By $[s_u, s_v] = s_u s_v s_u^{-1} s_v^{-1}$ we mean the commutator. We set

$$\operatorname{Val}(\Gamma) = \max\{\operatorname{valency}(u) : u \in V(\Gamma)\}.$$

By valency(u) of a vertex u we denote the number of edges incident to the vertex u. Clearly $\mathrm{Val}(\Gamma) \leq \mathrm{rank}(A(\Gamma)) - 1$.

If Γ is a simplicial graph, we denote by 1–skel(Γ) the 1–skeleton of Γ. Recall that a *full subgraph* of a graph Γ is a subgraph formed from a subset of vertices V and from all of the edges that have both endpoints in the subset V.

Conventions Let Γ be simplicial graph, $u \in V(\Gamma)$ and $e \in E(\Gamma)$. We denote by

(i) $\Gamma \setminus \{u\}$ the full subgraph of Γ formed from $V(\Gamma) \setminus \{u\}$,

(ii) $\Gamma \setminus e$ the subgraph of Γ such that $V(\Gamma \setminus e) = V(\Gamma)$ and $E(\Gamma \setminus e) = E(\Gamma) \setminus \{e\}$.

Lemma 2.11 *Let Γ be a finite simplicial graph. Then*

$$\mathrm{asdim}\, A(\Gamma) \leq \mathrm{Val}(\Gamma) + 1.$$

Proof By Theorem 2.10, it suffices to prove Lemma 2.11 for connected simplicial graphs, so assume Γ is a connected simplicial graph. We use induction on $\mathrm{rank}(A(\Gamma))$. For $\mathrm{rank}(A(\Gamma)) = 1$ we have that $A(\Gamma)$ is $\mathbb{Z}$, so the statement holds. We assume that the statement holds for every $k \leq n$ and we show that it holds for $n+1$ (for $n+1 \geq 2$).

Let Γ be a simplicial graph with $n+1$ vertices. We remove a vertex u from the graph Γ such that valency$(u) = \mathrm{Val}(\Gamma) = m \geq 1$. Let's denote by v_i (for $i \in \{1, \dots, m\}$) the vertices of Γ which are adjacent to u. We set $\Gamma' = \Gamma \setminus \{u\}$. Obviously, $\mathrm{Val}(\Gamma') \leq \mathrm{Val}(\Gamma)$. We denote by Y the full subgraph of Γ formed from $\{v_1, \dots, v_m\}$.

We observe that the RAAG $A(\Gamma)$ is an HNN-extension of the RAAG $A(\Gamma')$. To be more precise, we have that

$$A(\Gamma) = A(\Gamma') *_{A(Y)}.$$

By Theorem 2.9 we obtain that

$$\mathrm{asdim}\, A(\Gamma) \leq \max\{\mathrm{asdim}\, A(\Gamma'), \mathrm{asdim}\, A(Y) + 1\}.$$

We observe that $\mathrm{Val}(Y) \leq \mathrm{Val}(\Gamma) - 1$, so by the inductive hypothesis $(\mathrm{rank}(A(Y)) \leq n)$,

$$\mathrm{asdim}\, A(Y) \leq \mathrm{Val}(Y) + 1 \leq \mathrm{Val}(\Gamma).$$

Since $\mathrm{rank}\, A(\Gamma') = n$, again by the inductive hypothesis, we deduce that

$$\mathrm{asdim}\, A(\Gamma') \leq \mathrm{Val}(\Gamma') + 1 \leq \mathrm{Val}(\Gamma) + 1.$$

Combining the three previous inequalities, we obtain

$$\mathrm{asdim}\, A(\Gamma) \leq \max\{\mathrm{Val}(\Gamma) + 1, \mathrm{Val}(\Gamma) + 1\} = \mathrm{Val}(\Gamma) + 1. \qquad \square$$

Using the previous lemma we can compute the exact asymptotic dimension of $A(\Gamma)$. We note that this has already been computed by Wright [23] using different methods.

We set

$$\mathrm{Sim}(\Gamma) = \max\{n : \Gamma \text{ contains the 1–skeleton of the standard } (n{-}1)\text{–simplex } \Delta^{n-1}\}.$$

Obviously if $\Gamma' \subseteq \Gamma$, then $\mathrm{Sim}(\Gamma') \leq \mathrm{Sim}(\Gamma)$.

Theorem 2.12 *Let Γ be a finite simplicial graph. Then*

$$\operatorname{asdim} A(\Gamma) = \mathrm{Sim}(\Gamma).$$

Proof By Theorem 2.10, it suffices to prove Theorem 2.12 for connected simplicial graphs, so assume Γ is a connected simplicial graph.

Claim 1
$$\mathrm{Sim}(\Gamma) \leq \operatorname{asdim} A(\Gamma).$$

Proof Let $\mathrm{Sim}(\Gamma) = n$. We observe that $\mathbb{Z}^n = A(S_{n-1}) \leq A(\Gamma)$. It follows that

$$n = \operatorname{asdim} \mathbb{Z}^n \leq \operatorname{asdim} A(\Gamma). \qquad \square$$

Claim 2
$$\operatorname{asdim} A(\Gamma) \leq \mathrm{Sim}(\Gamma).$$

Proof We use induction on $\operatorname{rank}(A(\Gamma))$. For $\operatorname{rank}(A(\Gamma)) = 1$ we have that $A(\Gamma)$ is $\mathbb{Z}$, so the statement holds. We assume that the statement holds for every $r \leq m$, and we show that holds for $m+1$ as well. Let Γ be a connected simplicial graph with $m+1$ vertices.

Let $\mathrm{Sim}(\Gamma) = n$. Then Γ contains the 1–skeleton of the standard $(n{-}1)$–simplex S_{n-1} (where $S_{n-1} = 1\text{–skel}(\Delta^{n-1})$).

Case 1 $(\Gamma = S_{n-1})$ Then $m+1 = n$, so by Lemma 2.11 we have $\operatorname{asdim} A(S_{n-1}) \leq \mathrm{Val}(S_{n-1}) + 1$. By observing that $\mathrm{Val}(S_{n-1}) = n-1$, we obtain that

$$\operatorname{asdim} A(S_{n-1}) \leq n = \mathrm{Sim}(\Gamma).$$

Case 2 $(S_{n-1} \subsetneqq \Gamma)$ We will remove a vertex $u \in V(S_{n-1})$. Let's denote by v_i (for $i \in \{1, \dots, k\}$) the vertices of Γ which are adjacent to u. We set $\Gamma' = \Gamma \setminus \{u\}$. Obviously $\mathrm{Sim}(\Gamma') \leq n$. We denote by Y the full subgraph of Γ formed from $\{v_1, \dots, v_k\}$.

We observe that the RAAG $A(\Gamma)$ is an HNN-extension of the RAAG $A(\Gamma')$. To be more precise,

$$A(\Gamma) = A(\Gamma') *_{A(Y)} .$$

By Theorem 2.9 we obtain that

$$\operatorname{asdim} A(\Gamma) \leq \max\{\operatorname{asdim} A(\Gamma'), \operatorname{asdim} A(Y) + 1\}. \tag{2}$$

Since $\mathrm{Sim}(\Gamma') \leq n$ and $\mathrm{rank}(\Gamma') \leq m$, by the inductive assumption

$$\operatorname{asdim} A(\Gamma') \leq \mathrm{Sim}(\Gamma') \leq n. \tag{3}$$

We observe that $\mathrm{Sim}(Y) \leq n-1$ and $\mathrm{rank}(Y) \leq m$. Then by the inductive hypothesis we obtain

$$\operatorname{asdim} A(Y) + 1 \leq \mathrm{Sim}(Y) + 1 \leq n. \tag{4}$$

by (2), (3) and (4) we conclude that

$$\operatorname{asdim} A(\Gamma) \leq n = \mathrm{Sim}(\Gamma). \qquad \square$$

3 Asymptotic dimension of one-relator groups

Theorem 3.1 *Let G be a finitely generated one-relator group. Then*

$$\operatorname{asdim} G \leq 2.$$

Proof Let $G = \langle S \mid r \rangle$ be a presentation of G where S is finite and r is a cyclically reduced word in $S \cup S^{-1}$. To omit trivial cases, we assume that S contains at least two elements and $|r| > 0$ (we denote by $|r|$ the length of the relator r in the free group $F(S)$).

We may assume that every letter of S appears in r. Otherwise our group G is isomorphic to a free product $H * F$ of a finitely generated one-relator group H with relator r and generating set $S_H \subseteq S$ consisting of all letters which appear in r and a free group F with generating set the remaining letters of S. We recall that the asymptotic dimension of any finitely generated nonabelian free group is equal to one. Then $\operatorname{asdim} G = \max\{\operatorname{asdim} H, \operatorname{asdim} F\} = \max\{\operatorname{asdim} H, 1\}$; see [2].

We denote by $\epsilon_r(s)$ the exponent sum of a letter $s \in S$ in a word r, and by $\mathrm{oc}_r(s)$ the minimum number of the positions of appearance of the elements of the set

$$\{s^k \text{ for some } 0 \neq k \in \mathbb{Z}\}$$

in a cyclically reduced word r. For example, if $r = abcab^{10}a^{-2}c^{-1}$, then $\mathrm{oc}_r(a) = 3$, $\mathrm{oc}_r(b) = 2$, $\mathrm{oc}_r(c) = 2$ and $\epsilon_r(c) = 0$.

We observe that, if there exists $b \in S$ such that $\mathrm{oc}_r(b) = 1$, then the group G is free (see [15, Theorem 5.1, page 198]), so $\operatorname{asdim} G = 1$. From now on we assume that, for every $s \in S$, we have that $\mathrm{oc}_r(s) \geq 2$ (so $|r| \geq 4$).

The proof is by induction on the length of r. We observe that if $|r| = 4$ then the statement of the theorem holds, since by the result of Matsnev [17] we have that

$\operatorname{asdim} G \le \frac{1}{2}\lfloor |r| \rfloor = \frac{4}{2} = 2$ (where $\lfloor * \rfloor$ is the floor function). We assume that the statement of the theorem holds for all one-relator groups with relator length smaller than or equal to $|r|-1$.

We follow the arguments from the book of Lyndon and Schupp (see [15, Theorem 5.1, page 198]) and the book of Wise (see [22, Construction 18.5]). We distinguish two cases.

Case 1 (there exists a letter $a \in S$ such that $\epsilon_r(a) = 0$) We shall exhibit G as an HNN-extension of a one-relator group G_1 whose defining relator has shorter length than r, over a finitely generated free subgroup F. Let $S = \{a = s_1, s_2, s_3, s_4, \dots, s_k\}$. Set $s_i^{(j)} = a^j s_i a^{-j}$ for $j \in \mathbb{Z}$ and for $k \ge i \ge 2$. Rewrite r, scanning it from left to right and changing any occurrence of $a^j s_i$ to $s_i^{(j)} a^j$, collecting the powers of adjacent a–letters together and continuing with the leftmost occurrence of a or its inverse in the modified word. We denote by r' the modified word in terms of $s_i^{(j)}$. We note that by doing this we make at least one cancellation of a and its inverse. The resulting word r', which represents r in terms of $s_i^{(j)}$ and their inverses, has length at most $|r|-2$. For example, if $r = as_2s_3as_2^4a^{-2}s_3$ then $r' = s_2^{(1)}s_3^{(1)}(s_2^{(2)})^4 s_3^{(0)}$.

Let m and M be the minimal and the maximal superscripts, respectively, of all $s_i^{(j)}$ (for $i \ge 2$) occurring in r'. To be more precise,

$$m = \min\{j : s_i^{(j)} \text{ occurs in } r'\} \quad \text{and} \quad M = \max\{j : s_i^{(j)} \text{ occurs in } r'\}.$$

Continuing our example, $m = 0$ and $M = 2$.

Claim 1.1 *In Case 1 we have $M - m > 0$ and $m \le 0 \le M$.*

We may assume, replacing r with a suitable permutation if necessary, that r begins with a^k for some $k \ne 0$. Then we can write $r = a^k swa^ntz$, where $k, n \ne 0$, $a \notin \{s,t\} \subseteq S$ and both a and a^{-1} do not appear in the word z ($\mathrm{oc}_z(a) = 0$). Then we observe that the letter s has as superscript k in the word r' while t has as superscript 0 in the word r'. Since $k \ne 0$, we have that $M - m > 0$. This completes the proof of Claim 1.1.

Claim 1.2 *The group G has a presentation*

$$\langle \{a, s_i^{(j)} : i = 2, \dots, k,\ j = m, \dots, M\} \mid \{r', as_i^{(j')}a^{-1}(s_i^{(j'+1)})^{-1} : j' = m, \dots, M-1\}\rangle.$$

To verify the claim, let H be the group defined by the presentation given above. The map $\phi\colon G \to H$ defined by

$$a \mapsto a, \quad s_i \mapsto s_i^{(0)}$$

is a homomorphism since $\phi(r) = r'$. On the other hand, the map $\psi : H \to G$ defined by

$$a \mapsto a, \quad s_i^{(j)} \mapsto a^j s_i a^{-j}$$

is also a homomorphism since all relators of H are sent to 1_G.

It is easy to verify that $\psi \circ \phi$ is the identity map of G. The homomorphism $\phi \circ \psi : H \to H$ maps $a \mapsto a$, $s_i^{(0)} \mapsto s_i \mapsto s_i^{(0)}$ and $s_i^{(j)} \mapsto a^j s_i a^{-j} \mapsto a^j s_i^{(0)} a^{-j}$. Now we show that $s_i^{(j)} = a^j s_i^{(0)} a^{-j}$. We have

$$a^1 s_i^{(0)} a^{-1} = s_i^{(1)}, \quad a^1 s_i^{(1)} a^{-1} = s_i^{(2)}, \ldots \quad a^1 s_i^{(j-1)} a^{-1} = s_i^{(j)}.$$

Combining these equations $s_i^{(j)} = a^1 s_i^{(j-1)} a^{-1} = a^2 s_i^{(j-2)} a^{-2} = \cdots = a^j s_i^{(0)} a^{-j}$, so $\phi \circ \psi = \mathrm{id}_H$.

Since $\phi \circ \psi$ and $\psi \circ \phi$ are the identity maps on H and G, respectively, we deduce that ϕ is an isomorphism. This completes the proof of Claim 1.2.

We set

$$G_1 = \langle \{ s_i^{(j)} : i = 2, \ldots, k, \ j = m, \ldots, M \} \mid r' \rangle,$$

and note that there exists a letter $s_{i_m} \in S$ such that $s_{i_m}^{(m)}$ appears in r' and a letter $s_{i_M} \in S$ such that $s_{i_M}^{(M)}$ appears in r'.

Let F and Λ be the subgroups of G_1 generated by

$$X = \{ s_i^{(j)} : i = 2, \ldots, k, \ j = m, \ldots, M-1 \}$$

and

$$Y = \{ s_i^{(j)} : i = 2, \ldots, k, \ j = m+1, \ldots, M \},$$

respectively.

Claim 1.3 *The groups F and Λ are free subgroups of G_1.*

This claim follows by the Freiheitssatz (see [15, Theorem 5.1, page 198]); since X omits a generator of G_1 occurring in r' (this is the letter $s_{i_M}^{(M)}$) the subgroup F is free. The same holds for Λ, since Y omits the letter $s_{i_m}^{(m)}$.

Claim 1.4 *We have that $G \simeq G_1 *_F$.*

In particular, the map $s_i^{(j)} \mapsto s_i^{(j+1)}$ from X to Y extends to an isomorphism from F to Λ. Thus H is exhibited as the HNN-extension of G_1 over the finitely generated free group F using a as a stable letter. Since $G \simeq H$ (Claim 1.2),

$$G \simeq G_1 *_F.$$

By the fact that $|r'| < |r|$ and the inductive assumption we have that $\operatorname{asdim} G_1 \leq 2$. To conclude, we apply the inequality for HNN-extensions (Theorem 2.9): $\operatorname{asdim} G \leq \max\{\operatorname{asdim} G_1, \operatorname{asdim} F + 1\} = \max\{\operatorname{asdim} G_1, 2\} = 2$.

Case 2 ($|\epsilon_r(s)| \geq 1$ for every letter $s \in S$) Let $S = \{a = s_1, b = s_2, s_3, s_4, \dots, s_k\}$ and $S_1 = \{t, x, s_i : 3 \leq i \leq k\}$. We consider the homomorphism between the free group $F(S)$ and the free group $F(S_1)$

$$\phi\colon a \mapsto t^{-\epsilon_r(b)}x, \quad b \mapsto t^{\epsilon_r(a)}, \quad s_i \mapsto s_i \quad \text{for } 3 \leq i \leq k. \tag{5}$$

We set

$$\Gamma = \langle S_1 \mid r(t, x, s_3, \dots, s_k) \rangle,$$

where we denote by $r(t, x, s_3, \dots, s_k)$ the modified word in terms of t, x and s_i for $3 \leq i \leq k$ which is obtained from r when we replace a generator s with $\phi(s)$. Then ϕ induces a homomorphism

$$\phi\colon G \to \Gamma.$$

The following claim shows that the homomorphism ϕ is actually a monomorphism into Γ, so we have an embedding of G into Γ via ϕ:

Claim 2.1 *The homomorphism $\phi\colon G \to \Gamma$ is a monomorphism.*

Proof We set $S_2 = \{a, t, s_i : 3 \leq i \leq k\}$ and $S_1 = \{x, t, s_i : 3 \leq i \leq k\}$. We define $g\colon F(S) \to F(S_2)$ and $f\colon F(S_2) \to F(S_1)$ by

$$g\colon a \mapsto a, \quad b \mapsto t^{\epsilon_r(a)}, \quad s_i \mapsto s_i \quad \text{for } 3 \leq i \leq k,$$

and

$$f\colon a \mapsto t^{-\epsilon_r(b)}x, \quad t \mapsto t, \quad s_i \mapsto s_i \quad \text{for } 3 \leq i \leq k.$$

We set $r_2 = g(r)$, $G_2 = \langle S_2 \mid r_2 \rangle$ and $r_1 = f \circ g(r) = r(t, x, s_3, \dots, s_k)$, and we observe that $\Gamma = \langle S_1 \mid r_1 \rangle$. Then g induces a homomorphism $\bar{g}\colon G \mapsto G_2$ and f induces a homomorphism $\bar{f}\colon G_2 \mapsto \Gamma$. Obviously, $\phi = \bar{f} \circ \bar{g}$.

We can easily see that $\bar{f}$ is an isomorphism. Indeed, the homomorphism $\psi\colon \Gamma \to G_2$ given by

$$x \mapsto t^{\epsilon_r(b)}a, \quad t \mapsto t, \quad s_i \mapsto s_i \quad \text{for } 3 \leq i \leq k$$

is the inverse homomorphism of $\bar{f}$.

It is enough to prove that $\bar{g}$ is a monomorphism. This follows by the fact that the group G_2 is the amalgamated product $G *_{\mathbb{Z}} \langle t \rangle$, where $\mathbb{Z} = \langle \lambda \rangle$, and $\psi_1(\lambda) = b$ and $\psi_2(\lambda) = t^{\epsilon_r(a)}$ are the corresponding monomorphisms. We can see that $\bar{g}$ is the inclusion of G into the amalgamated product, so $\bar{g}$ is injective. □

We denote by $r(t,x,s_3,\dots,s_k)$ the modified word in terms of t, x and s_i for $3\le i\le k$ which can be obtained from r when we replace a generator s with $\phi(s)$ and p with the cyclically reduced $r(t,x,s_3,\dots,s_k)$. We observe that $\epsilon_p(t)=0$ and that x occurs in p.

If the letter t occurs in the word p, from Case 1 we have that Γ is an HNN-extension of some group H over a free subgroup F, namely $\Gamma = H*_F$. As in Case 1, by assuming that p starts with t or t^{-1} we introduce new variables $s_i^{(j)}=t^j s_i t^{-j}$. Using these variables, we rewrite p as a word w, eliminating all occurrences of t and its inverse. Then we observe that $|w|\le |r|-1$. By using the inductive assumption for w we obtain

$$\operatorname{asdim} G \le \operatorname{asdim}\Gamma \le 2.$$

If the letter t does not occur in the word p, we observe that

$$|p|\le |r|-1.$$

Then

$$\Gamma = \langle t\rangle * \Gamma',$$

where

$$\Gamma' = \langle \{x, s_i : i=3,\dots,k\} \mid p\rangle.$$

Since $\operatorname{asdim}(G_1*G_2)=\max\{\operatorname{asdim} G_1, \operatorname{asdim} G_2\}$ holds (see [2]) we have that

$$\operatorname{asdim}\Gamma = \max\{1, \operatorname{asdim}\Gamma'\}.$$

Then, by the inductive assumption for p, $\operatorname{asdim}\Gamma'\le 2$. Finally, we conclude that

$$\operatorname{asdim} G \le \operatorname{asdim}\Gamma \le 2. \qquad \square$$

3.1 One-relator groups with asymptotic dimension two

We recall that a nontrivial group H is *freely indecomposable* if H cannot be expressed as a free product of two nontrivial groups.

A natural question derived from Theorem 3.1 is which one-relator groups have asymptotic dimension two. In this subsection, we will show that the asymptotic dimension of every finitely generated one-relator group that is not a free group or a free product of a free group and a finite cyclic group is exactly two.

We will use Propositions 3.2 and 3.3 from [10] and [21], respectively.

Proposition 3.2 *Let G be an infinite finitely generated one-relator group with torsion. If G has more than one end, then G is a free product of a nontrivial free group and a freely indecomposable one-relator group.*

Proposition 3.3 *Let G be a torsion-free infinite finitely generated group. If G is virtually free, then it is free.*

Lemma 3.4 *Let G be an infinite finitely generated one-relator group that is not a free group or a free product of a nontrivial free group and a freely indecomposable one-relator group. Then G is not virtually free.*

Proof If G has torsion, by Proposition 3.2 G has exactly one end, so G cannot be virtually free. If G is torsion free, by Proposition 3.3 we obtain that G is free and this is a contradiction by the assumption of the lemma. □

We note that every finite one-relator group is cyclic. To see this, it is enough to observe that every one-relator group with at least two generators has infinite abelianization.

The following proposition is the main result of this subsection:

Proposition 3.5 *Let G be a finitely generated one-relator group that is not a free group or a free product of a free group and a finite cyclic group. Then*

$$\operatorname{asdim} G = 2.$$

Proof By Theorem 3.1, $\operatorname{asdim} G \leq 2$. If G is finite then it is cyclic. If G is infinite, $1 \leq \operatorname{asdim} G$. By a theorem of T Gentimis [11], $\operatorname{asdim} G = 1$ if and only if G is virtually free. We assume that G is an infinite virtually free group. So, if G is torsion free, then by Proposition 3.3 we obtain that G is free. If G has torsion, then, by Lemma 3.4, G is a free product of a nontrivial free group and a freely indecomposable one-relator group G_1. Observe that, if G_1 is an infinite noncyclic group, then by the same lemma G_1 is not virtually free, so G is not virtually free either, which is a contradiction. We conclude that $\operatorname{asdim} G = 2$. □

Corollary *Let G be a finitely generated freely indecomposable one-relator group which is not cyclic. Then*

$$\operatorname{asdim} G = 2.$$

Proposition 3.6 [15, Proposition 5.13, page 107] *Let $G = \langle x_1, \dots, x_n \mid r \rangle$ be a finitely generated one-relator group, where r is of minimal length under $\operatorname{Aut}(F(\{x_1, \dots, x_n\}))$ and contains exactly the generators $x_1, \dots, x_k$ for some k with $0 \leq k \leq n$. Then G is isomorphic to the free product $G_1 * G_2$, where $G_1 = \langle x_1, \dots, x_k \mid r \rangle$ is freely indecomposable and G_2 is free with basis $\{x_{k+1}, \dots, x_n\}$.*

The above results combine to give the following corollary, which describes the finitely generated one-relator groups:

Corollary *Let G be a finitely generated one-relator group. Then one of the following is true*:

(i) *G is finite cyclic*, *and* $\operatorname{asdim} G = 0$.

(ii) *G is a nontrivial free group or a free product of a nontrivial free group and a finite cyclic group*, *and* $\operatorname{asdim} G = 1$.

(iii) *G is an infinite freely indecomposable not cyclic group or a free product of a nontrivial free group and an infinite freely indecomposable noncyclic group*, *and* $\operatorname{asdim} G = 2$.

We can further describe the boundaries of hyperbolic one-relator groups. We recall the following result of Buyalo and Lebedeva (see [7]) for hyperbolic groups:

$$\operatorname{asdim} G = \dim \partial_\infty G + 1.$$

Let G be an infinite finitely generated hyperbolic one-relator group that is not virtually free. By Gentimis [11] we obtain that $\operatorname{asdim} G \neq 1$, so $\operatorname{asdim} G = 2$. Using the previous equality we obtain that G has one-dimensional boundary. Applying a theorem of Kapovich and Kleiner (see [13]) we can describe the boundaries of hyperbolic one-relator groups:

Proposition 3.7 *Let G be a hyperbolic one-relator group. Then* $\operatorname{asdim} G = 0, 1$ *or* 2.

(i) *If* $\operatorname{asdim} G = 0$, *then G is finite.*

(ii) *If* $\operatorname{asdim} G = 1$, *then G is virtually free and the boundary is a Cantor set.*

(iii) *If* $\operatorname{asdim} G = 2$, *providing that G does not split over a virtually cyclic subgroup*, *then one of the following holds*:

  (a) *$\partial_\infty G$ is a Menger curve.*

  (b) *$\partial_\infty G$ is a Sierpinski carpet.*

  (c) *$\partial_\infty G$ is homeomorphic to S^1.*

4 Graphs of groups

We will prove a general theorem for the asymptotic dimension of fundamental groups of finite graphs of groups.

Theorem 4.1 *Let $(\mathbb{G}, Y)$ be a finite graph of groups with vertex groups $\{G_v : v \in Y^0\}$ and edge groups $\{G_e : e \in Y^1_+\}$. Then*

$$\operatorname{asdim} \pi_1(\mathbb{G}, Y, \mathbb{T}) \le \max_{v \in Y^0, e \in Y^1_+} \{\operatorname{asdim} G_v, \operatorname{asdim} G_e + 1\}.$$

Proof We use induction on the number $\#E(Y)$ of edges of the graph Y. For $\#E(Y) = 1$ we distinguish two cases. The first case is when the fundamental group $\pi_1(\mathbb{G}, Y, \mathbb{T})$ is an amalgamated product. Here the theorem follows by the inequality of Dranishnikov (see [9])

$$\operatorname{asdim} A *_C B \le \max\{\operatorname{asdim} A, \operatorname{asdim} B, \operatorname{asdim} C + 1\}.$$

The second case is when the fundamental group $\pi_1(\mathbb{G}, Y, \mathbb{T})$ is an HNN-extension. Here the theorem follows by Theorem 2.9.

We assume that the theorem holds for $E(Y) \le m$. Let $(\mathbb{G}, Y)$ be a finite graph of groups with $\#E(Y) = m + 1$. We denote by $\mathbb{T}$ a maximal tree of Y.

We distinguish two cases:

Case 1 $(Y = \mathbb{T})$ We remove a terminal edge $e' = [v, u]$ from the graph Y so that the full subgraph of Y, denoted by Γ and formed from the vertices $V(Y) \setminus \{u\}$, is connected. We observe that Γ is also a tree, which we denote by $\mathbb{T}'$.

Then $\pi_1(\mathbb{G}, Y, \mathbb{T}) = \pi_1(\mathbb{G}, \Gamma, \mathbb{T}') *_{G_{e'}} G_u$, so by the inequality for amalgamated products of Dranishnikov (see [9]),

$$\operatorname{asdim} \pi_1(\mathbb{G}, Y, \mathbb{T}) \le \max\{\operatorname{asdim} \pi_1(\mathbb{G}, \Gamma, \mathbb{T}'), \operatorname{asdim} G_u, \operatorname{asdim} G_{e'} + 1\}.$$

Since $\#E(\Gamma) = m$, by the inductive assumption we obtain that

$$\operatorname{asdim} \pi_1(\mathbb{G}, \Gamma, \mathbb{T}') \le \max_{v \in Y^0 \setminus \{u\}, e \in Y^1_+ \setminus \{e'\}} \{\operatorname{asdim} G_v, \operatorname{asdim} G_e + 1\},$$

so

$$\operatorname{asdim} \pi_1(\mathbb{G}, Y, \mathbb{T}) \le \max_{v \in Y^0, e \in Y^1_+} \{\operatorname{asdim} G_v, \operatorname{asdim} G_e + 1\}.$$

Case 2 $(\mathbb{T} \subsetneq Y)$ We remove from Y an edge $e' = [v, u]$ which doesn't belong to $\mathbb{T}$. Since the tree $\mathbb{T}$ is a maximal tree of Y and $e' \notin E(\mathbb{T})$, we have that the graph $\Gamma = Y \setminus e'$ is connected and $\mathbb{T} \subseteq \Gamma$. Then $\pi_1(\mathbb{G}, Y, \mathbb{T}) = \pi_1(\mathbb{G}, \Gamma, \mathbb{T}) *_{G_{e'}}$, so by the inequality for HNN-extensions (Theorem 2.9) we have

$$\operatorname{asdim} \pi_1(\mathbb{G}, Y, \mathbb{T}) \le \max\{\operatorname{asdim} \pi_1(\mathbb{G}, \Gamma, \mathbb{T}), \operatorname{asdim} G_{e'} + 1\}.$$

Since $\#E(\Gamma)=m$, by the inductive assumption

$$\operatorname{asdim}\pi_1(\mathbb{G},\Gamma,\mathbb{T})\le \max_{v\in Y^0, e\in Y^1_+\setminus\{e'\}}\{\operatorname{asdim}G_v, \operatorname{asdim}G_e+1\},$$

so

$$\operatorname{asdim}\pi_1(\mathbb{G},Y,\mathbb{T})\le \max_{v\in Y^0, e\in Y^1_+}\{\operatorname{asdim}G_v, \operatorname{asdim}G_e+1\}. \qquad \square$$

We obtain as a corollary the following:

Proposition 4.2 *Let $(\mathbb{G},Y)$ be a finite graph of groups with vertex groups $\{G_v : v\in Y^0\}$ and edge groups $\{G_e : e\in Y^1_+\}$. We assume that*

$$\max_{e\in Y^1_+}\{\operatorname{asdim}G_e\} < \max_{v\in Y^0}\{\operatorname{asdim}G_v\}=n.$$

Then $\operatorname{asdim}\pi_1(\mathbb{G},Y,\mathbb{T})=n$.

As a corollary of Proposition 4.2, the asymptotic dimension of a graph of one-ended hyperbolic groups with n–dimensional boundary with free edge groups is $n+1$.

References

[1] **G Bell**, **A Dranishnikov**, *On asymptotic dimension of groups*, Algebr. Geom. Topol. 1 (2001) 57–71 MR Zbl

[2] **G Bell**, **A Dranishnikov**, *On asymptotic dimension of groups acting on trees*, Geom. Dedicata 103 (2004) 89–101 MR Zbl

[3] **G Bell**, **A Dranishnikov**, *Asymptotic dimension*, Topology Appl. 155 (2008) 1265–1296 MR Zbl

[4] **G Bell**, **A Dranishnikov**, *Asymptotic dimension in Będlewo*, Topology Proc. 38 (2011) 209–236 MR Zbl

[5] **G C Bell**, **A N Dranishnikov**, **J E Keesling**, *On a formula for the asymptotic dimension of free products*, Fund. Math. 183 (2004) 39–45 MR Zbl

[6] **M Bestvina**, **K Bromberg**, **K Fujiwara**, *Constructing group actions on quasi-trees and applications to mapping class groups*, Publ. Math. Inst. Hautes Études Sci. 122 (2015) 1–64 MR Zbl

[7] **S V Buyalo**, **N D Lebedeva**, *Dimensions of locally and asymptotically self-similar spaces*, Algebra i Analiz 19 (2007) 60–92 MR Zbl In Russian; translated in St. Petersburg Math. J. 19 (2008) 45–65

[8] **A Dranishnikov**, *Asymptotic dimension*, web resource, in "Problems in geometric group theory" Available at `https://sites.google.com/a/scu.edu/rscott/pggt`

[9] **A Dranishnikov**, *On asymptotic dimension of amalgamated products and right-angled Coxeter groups*, Algebr. Geom. Topol. 8 (2008) 1281–1293 MR Zbl

[10] **J Fischer**, **A Karrass**, **D Solitar**, *On one-relator groups having elements of finite order*, Proc. Amer. Math. Soc. 33 (1972) 297–301 MR Zbl

[11] **T Gentimis**, *Asymptotic dimension of finitely presented groups*, Proc. Amer. Math. Soc. 136 (2008) 4103–4110 MR Zbl

[12] **M Gromov**, *Asymptotic invariants of infinite groups*, from "Geometric group theory, II" (G A Niblo, M A Roller, editors), London Math. Soc. Lecture Note Ser. 182, Cambridge Univ. Press (1993) 1–295 MR Zbl

[13] **M Kapovich**, **B Kleiner**, *Hyperbolic groups with low-dimensional boundary*, Ann. Sci. École Norm. Sup. 33 (2000) 647–669 MR Zbl

[14] **R C Lyndon**, *Cohomology theory of groups with a single defining relation*, Ann. of Math. 52 (1950) 650–665 MR Zbl

[15] **R C Lyndon**, **P E Schupp**, *Combinatorial group theory*, Ergebnisse der Math. 89, Springer (1977) MR Zbl

[16] **J M Mackay**, **A Sisto**, *Embedding relatively hyperbolic groups in products of trees*, Algebr. Geom. Topol. 13 (2013) 2261–2282 MR Zbl

[17] **D Matsnev**, *Asymptotic dimension of one relator groups*, Colloq. Math. 111 (2008) 85–89 MR Zbl

[18] **D Osin**, *Asymptotic dimension of relatively hyperbolic groups*, Int. Math. Res. Not. 2005 (2005) 2143–2161 MR Zbl

[19] **A N Platonov**, *An isoparametric function of the Baumslag–Gersten group*, Vestnik Moskov. Univ. Ser. I Mat. Mekh. (2004) 12–17 MR Zbl In Russian; translated in Moscow Univ. Math. Bull. 59 (2004) 12–17

[20] **L Sledd**, *Assouad–Nagata dimension of finitely generated $C'(1/6)$ groups*, Topology Appl. 299 (2021) art. id. 107634 MR Zbl

[21] **J Stallings**, *Groups of dimension 1 are locally free*, Bull. Amer. Math. Soc. 74 (1968) 361–364 MR Zbl

[22] **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, Ann. of Math. Stud. 209, Princeton Univ. Press (2021) MR Zbl

[23] **N Wright**, *Finite asymptotic dimension for* $\mathrm{CAT}(0)$ *cube complexes*, Geom. Topol. 16 (2012) 527–554 MR Zbl

[24] **G Yu**, *The Novikov conjecture for groups with finite asymptotic dimension*, Ann. of Math. 147 (1998) 325–355 MR Zbl

Mathematical Institute, University of Oxford
Oxford, United Kingdom

panagiotis.tselekidis@queens.ox.ac.uk

Received: 3 June 2021 Revised: 11 April 2022

Algebraic & Geometric Topology **23**:8 (2023) 3615–3653
DOI: 10.2140/agt.2023.23.3615
Published: 5 November 2023

Pressure metrics for deformation spaces of quasifuchsian groups with parabolics

HARRISON BRAY
RICHARD CANARY
LIEN-YUNG KAO

We produce a mapping class group–invariant pressure metric on the space $\mathrm{QF}(S)$ of quasiconformal deformations of a cofinite-area fuchsian group uniformizing S. Our pressure metric arises from an analytic pressure form on $\mathrm{QF}(S)$ which is degenerate only on pure bending vectors on the fuchsian locus. Our techniques also show that the Hausdorff dimension of the limit set varies analytically.

57K32

1 Introduction

We construct a pressure metric on the quasifuchsian space $\mathrm{QF}(S)$ of quasiconformal deformations, within $\mathsf{PSL}(2,\mathbb{C})$, of a fuchsian group Γ in $\mathsf{PSL}(2,\mathbb{R})$ whose quotient $\mathbb{H}^2/\Gamma$ has finite area and is homeomorphic to the interior of a compact surface S. Our pressure metric is a mapping class group–invariant path metric which is a Riemannian metric on the complement of the submanifold of fuchsian representations. Our metric and its construction generalize work of Bridgeman [9], in which $\mathbb{H}^2/\Gamma$ is a closed surface.

McMullen [30] initiated the study of pressure metrics, by constructing a pressure metric on the Teichmüller space of a closed surface. His pressure metric is one way of formalizing Thurston's notion of constructing a metric on Teichmüller space as the "Hessian of the length of a random geodesic" (see also Wolpert [48], Bonahon [4] and Fathi and Flaminio [18]) and like Thurston's metric it agrees with the classical Weil–Petersson metric (up to scalar multiplication). Subsequently, Bridgeman [9] constructed a pressure metric on quasifuchsian space, Bridgeman, Canary, Labourie and Sambarino [10] constructed pressure metrics on deformation spaces of Anosov representations, and Pollicott and Sharp [33] constructed pressure metrics on spaces of metric graphs (see also Kao [21]). The main tool in the construction of these pressure metrics is the thermodynamic formalism for topologically transitive Anosov flows with compact support and their associated well-behaved finite Markov codings.

The major obstruction to extending the constructions of pressure metrics to deformation spaces of geometrically finite (rather than convex cocompact) Kleinian groups and related settings is that the support of the recurrent portion of the geodesic flow is not compact and hence there is not a well-behaved finite Markov coding. Mauldin and Urbański [29] and Sarig [39] extended the thermodynamical formalism to the setting of topologically mixing Markov shifts with countable alphabet and the BIP property. In the case of finite-area hyperbolic surfaces, Stadlbauer [42] and Ledrappier and Sarig [27] construct and study a topologically mixing countable Markov coding with the BIP property for the recurrent portion of the geodesic flow of the surface. In previous work, Kao [23] showed how to adapt the thermodynamic formalism in the setting of the Stadlbauer–Ledrappier–Sarig coding to construct pressure metrics on Teichmüller spaces of punctured surfaces.

We adapt the techniques developed by Bridgeman [9] and Kao [23] into our setting to construct a pressure metric which can again be naturally interpreted as the Hessian of the (renormalized) length of a random geodesic.

Theorem 9.1 *If S is a compact surface with nonempty boundary, the pressure form $\mathbb{P}$ on $\mathrm{QF}(S)$ induces a $\mathrm{Mod}(S)$–invariant path metric, which is an analytic Riemannian metric on the complement of the fuchsian locus.*

Moreover, if $v \in T_\rho(\mathrm{QF}(S))$, *then* $\mathbb{P}(v,v) = 0$ *if and only if* ρ *is fuchsian and* v *is a pure bending vector.*

The control obtained from the thermodynamic formalism allows us to see that the topological entropy of the geodesic flow of the quasifuchsian hyperbolic 3–manifold varies analytically over $\mathrm{QF}(S)$. We recall that the topological entropy $h(\rho)$ of ρ is the exponential growth rate of the number of closed orbits of the geodesic flow of $N_\rho = \mathbb{H}^3/\rho(\Gamma)$ of length at most T. More precisely, if

$$R_T(\rho) = \{[\gamma] \in [\Gamma] \mid 0 < \ell_\rho(\gamma) \le T\},$$

where $[\Gamma]$ is the collection of conjugacy classes in Γ and $\ell_\rho(\gamma)$ is the translation length of the action of $\rho(\gamma)$ on $\mathbb{H}^3$, then the topological entropy is given by

$$h(\rho) = \lim_{T\to\infty} \frac{\# R_T(\rho)}{T}.$$

Sullivan [45] showed that the topological entropy and the Hausdorff dimension of the limit set agree for quasifuchsian groups. So we see that the Hausdorff dimension of the limit set varies analytically over $\mathrm{QF}(S)$, generalizing a result of Ruelle [36] for quasifuchsian deformation spaces of closed surfaces. Schapira and Tapie [40, Theorem 6.2] previously established that the entropy is C^1 on $\mathrm{QF}(S)$ and computed its derivative (as a special case of a much more general result).

Corollary 5.3 *If* S *is a compact surface with nonempty boundary, then the Hausdorff dimension of the limit set varies analytically over* $\mathrm{QF}(S)$.

Concretely, the pressure form $\mathbb{P}$ at a representation ρ_0 is the Hessian of the renormalized pressure intersection $J(\rho_0,\cdot)$ at ρ_0. The pressure intersection of $\rho, \eta \in \mathrm{QF}(S)$ is given by

$$I(\rho,\eta) = \lim_{T\to\infty} \frac{1}{|R_T(\rho)|} \sum_{[\gamma]\in R_T(\rho)} \frac{\ell_\eta(\gamma)}{\ell_\rho(\gamma)}$$

and the renormalized pressure intersection is given by

$$J(\rho,\eta) = \frac{h(\eta)}{h(\rho)} \lim_{T\to\infty} \frac{1}{|R_T(\rho)|} \sum_{[\gamma]\in R_T(\rho)} \frac{\ell_\eta(\gamma)}{\ell_\rho(\gamma)}.$$

The pressure intersection was first defined by Burger [12] for pairs of convex cocompact fuchsian representations. Schapira and Tapie [40] defined an intersection function for negatively curved manifolds with an entropy gap at infinity, by generalizing the geodesic stretch considered by Knieper [26] in the compact setting. Their definition applies in a much more general framework, but agrees with our notion in this setting; see [40, Proposition 2.17].

Let (Σ^+, σ) be the Stadlbauer–Ledrappier–Sarig coding of a fuchsian group Γ giving a finite-area uniformization of S. If $\rho \in \mathrm{QF}(S)$, we construct a roof function $\tau_\rho \colon \Sigma^+ \to \mathbb{R}$ whose periods are translation lengths of elements of $\rho(\Gamma)$. The key technical work in the paper is a careful analysis of these roof functions. In particular, we show that they vary analytically over $\mathrm{QF}(S)$; see Proposition 3.1. If P is the Gurevich pressure function (on the space of all well-behaved roof functions), then the topological entropy $h(\rho)$ of ρ is the unique solution of $P(-t\tau_\rho) = 0$. Our actual working definition of the intersection function will be expressed in terms of equilibrium states on Σ^+ for the functions $-h(\rho)\tau_\rho$, but we will show in Theorem 10.3 that this thermodynamical definition agrees with the more geometric definition given above.

Following Burger [12], if $\rho, \eta \in \mathrm{QF}(S)$, we define the *Manhattan curve*

$$\mathcal{C}(\rho, \eta) = \{(a, b) \mid a, b \geq 0,\ a + b > 0 \text{ and } P(-a\tau_\rho - b\tau_\eta) = 0\}.$$

The following result generalizes work of Burger [12] and Kao [22]:

Theorem (Theorems 6.1 and 10.3) *If S is a compact surface with nonempty boundary and $\rho, \eta \in \mathrm{QF}(S)$, then $\mathcal{C}(\rho, \eta)$*

(1) *is a closed subsegment of an analytic curve*,

(2) *has endpoints $(h(\rho), 0)$ and $(0, h(\eta))$, and*

(3) *is strictly convex*, *unless ρ and η are conjugate in* $\mathrm{Isom}(\mathbb{H}^3)$.

Moreover, *the tangent line to $\mathcal{C}(\rho, \eta)$ at $(h(\rho), 0)$ has slope $-I(\rho, \eta)$.*

We use Theorem 6.1 in our proof of a rigidity result for the renormalized pressure intersection (see Corollary 7.2) and in our proof that pressure intersection is analytic on $\mathrm{QF}(S) \times \mathrm{QF}(S)$ (see Proposition 7.1). We also use it to obtain a rigidity theorem for weighted entropy in the spirit of the Bishop–Steger rigidity theorem for fuchsian groups; see [3]. If $a, b > 0$ and $\rho, \eta \in \mathrm{QF}(S)$, we define the weighted entropy

$$h^{a,b}(\rho, \eta) = \lim \frac{1}{T} \#\{[\gamma] \in [\Gamma] \mid 0 < a\ell(\rho(\gamma)) + b\ell(\eta(\gamma)) \leq T\}.$$

Corollary 6.3 *If S is a compact surface with nonempty boundary*, *$\rho, \eta \in \mathrm{QF}(S)$ and $a, b > 0$*, *then*

$$h^{a,b}(\rho, \eta) \leq \frac{h(\rho)h(\eta)}{bh(\rho) + ah(\eta)}$$

with equality if and only if $\rho = \eta$.

Other viewpoints If $\rho \in \mathrm{QF}(S)$, then $N_\rho = \mathbb{H}^3/\rho(\Gamma)$ is a geometrically finite hyperbolic 3–manifold. As such, its dynamics may be analyzed using techniques from dynamics which do not rely on symbolic dynamics. For example, it naturally fits into the frameworks for geometrically finite, negatively curved manifolds developed by Dal'bo, Otal and Peigné [14], negatively curved Riemannian manifolds with bounded geometry as studied by Paulin, Pollicott and Schapira [32], and negatively curved manifolds with an entropy gap at infinity as studied by Schapira and Tapie [40]. In particular, the existence of equilibrium states and their continuous variation in our setting also follows from the work of Schapira and Tapie [40].

Since all the geodesic flows of manifolds in $\mathrm{QF}(S)$ are Hölder orbit equivalent, one should be able to think of them all as arising from an analytically varying family of Hölder potential functions on the geodesic flow of a fixed hyperbolic 3–manifold. However, for the construction of the pressure metric it will be necessary to know that the pressure function is at least twice differentiable. Results of this form do not yet seem to be available without symbolic dynamics. We have therefore chosen to develop the theory entirely from the viewpoint of the coding throughout the paper.

Iommi, Riquelme and Velozo [20] have previously used the Dal'bo–Peigné coding [16] to study negatively curved manifolds of extended Schottky type. These manifolds include the hyperbolic 3–manifolds associated to all quasiconformal deformations of finitely generated fuchsian groups whose quotients have infinite area. In particular, they perform a phase transition analysis and show the existence and uniqueness of equilibrium states in their setting. The symbolic approach to phase transition analysis can be traced back to Iommi and Jordan [19]. Riquelme and Velozo [34] work in a more general setting which includes quasifuchsian groups with parabolics, but without a coding, and obtain a phase transition analysis for the pressure function as well as the existence of equilibrium measures.

Acknowledgments The authors would like to thank Francois Ledrappier, Mark Pollicott, Ralf Spatzier and Dan Thompson for helpful conversations during the course of their investigation. We also thank the referees, whose suggestions greatly improved the exposition.

Canary was partially supported by the grants DMS-1564362 and DMS-1906441 from the National Science Foundation (NSF). Kao was partially supported by the grant DMS-1703554 from NSF.

2 Background

2.1 Quasifuchsian space

Let S be a compact orientable surface with nonempty boundary and suppose that $\Gamma \subset \mathsf{PSL}(2,\mathbb{R})$ is a discrete torsion-free group such that $\mathbb{H}^2/\Gamma$ is a finite-area hyperbolic surface homeomorphic to the interior of S. We say that $\rho\colon \Gamma \to \mathsf{PSL}(2,\mathbb{C})$ is *quasifuchsian* if there exists a quasiconformal homeomorphism $\phi\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ such that $\rho(\gamma) = \phi\gamma\phi^{-1}$ for all $\gamma \in \Gamma$. Equivalently, ρ is quasifuchsian if and only if there is an orientation-preserving bilipschitz homeomorphism from $N_\rho = \mathbb{H}^3/\rho(\Gamma)$ to $N = \mathbb{H}^3/\Gamma$ in the homotopy class determined by ρ (see Douady and Earle [17]). Let $\mathrm{QC}(\Gamma) \subset \mathrm{Hom}(\Gamma, \mathsf{PSL}(2,\mathbb{C}))$ denote the space of all quasifuchsian representations. We recall — see Maskit [28, Theorem 2] — that $\rho\colon \Gamma \to \mathsf{PSL}(2,\mathbb{C})$ is quasifuchsian if and only if ρ is discrete and faithful, $\rho(\partial S)$ is parabolic and $\rho(\Gamma)$ preserves a Jordan curve in $\widehat{\mathbb{C}}$.

The *quasifuchsian space* is given by

$$\mathrm{QF}(S) = \mathrm{QC}(\Gamma)/\mathsf{PSL}(2,\mathbb{C}) \subset X(S) = \mathrm{Hom}_{\mathrm{tp}}(\Gamma, \mathsf{PSL}(2,\mathbb{C}))/\!\!/\mathsf{PSL}(2,\mathbb{C}),$$

where $\mathrm{Hom}_{\mathrm{tp}}(\Gamma, \mathsf{PSL}(2,\mathbb{C}))$ is the space of type-preserving representations of Γ into $\mathsf{PSL}(2,\mathbb{C})$ (ie representations taking parabolic elements of Γ to parabolic elements of $\mathsf{PSL}(2,\mathbb{C})$). We call $X(S)$ the relative character variety and it has the structure of a projective variety. The space $\mathrm{QF}(S)$ is a smooth open subset of $X(S)$, so is naturally a complex analytic manifold. (See Kapovich [24, Section 4.3] for details.) Bers [2] showed that $\mathrm{QF}(S)$ admits a natural identification with $\mathcal{T}(S) \times \mathcal{T}(S)$, where $\mathcal{T}(S)$ is the Teichmüller space of S.

If $\rho \in \mathrm{QC}(\Gamma)$ and ϕ is a quasiconformal map such that $\rho(\gamma) = \phi\gamma\phi^{-1}$ for all $\gamma \in \Gamma$, then ϕ restricts to a ρ–equivariant map $\xi_\rho\colon \Lambda(\Gamma) \to \Lambda(\rho(\Gamma))$, where $\Lambda(\rho(\Gamma))$ is the limit set of $\rho(\Gamma)$, ie the smallest closed $\rho(\Gamma)$–invariant subset of $\widehat{\mathbb{C}}$. Notice that, since ξ_ρ is ρ–equivariant, it must take the attracting fixed point γ^+ of a hyperbolic element $\gamma \in \Gamma$ to the attracting fixed point $\rho(\gamma)^+$ of $\rho(\gamma)$. Since attracting fixed points of hyperbolic elements are dense in $\Lambda(\Gamma)$, ξ_ρ depends only on ρ (and not on the choice of quasiconformally conjugating map ϕ). We now record well-known fundamental properties of this limit map.

Lemma 2.1 *If $\rho \in \mathrm{QC}(\Gamma)$, then there exists a ρ–equivariant bi-Hölder continuous map*

$$\xi_\rho\colon \Lambda(\Gamma) \to \Lambda(\rho(\Gamma)).$$

Moreover, if $x \in \Lambda(\Gamma)$, then $\xi_\rho(x)$ varies complex analytically over $\mathrm{QC}(\Gamma)$.

Proof Since each ξ_ρ is the restriction of a quasiconformal map $\phi\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ and quasiconformal maps are bi-Hölder (see [1, Theorem 10.3.2]), ξ_ρ is also bi-Hölder.

Suppose that $\{\rho_z\}_{z\in\Delta}$ is a complex analytic family of representations in $\mathrm{QC}(\Gamma)$ parametrized by the unit disk Δ. Sullivan [46, Theorem 1] showed that there is a continuous map $F\colon \Lambda(\Gamma)\times\Delta \to \widehat{\mathbb{C}}$ such that, if $z\in\Delta$, then $F(\cdot\,,z) = \xi_{\rho_z}$, and, if $x\in\Lambda(\Gamma)$, then $F(x,\cdot)$ varies holomorphically in z. Hartogs' theorem then implies that $\xi_\rho(x)$ varies complex analytically over all of $\mathrm{QC}(\Gamma)$. □

2.2 Countable Markov shifts

A two-sided *countable Markov shift* with countable alphabet $\mathcal{A}$ and transition matrix $\mathbb{T}\in\{0,1\}^{\mathcal{A}\times\mathcal{A}}$ is the set

$$\Sigma = \{x = (x_i)\in\mathcal{A}^{\mathbb{Z}} \mid t_{x_i x_{i+1}} = 1 \text{ for all } i\in\mathbb{Z}\}$$

equipped with a shift map $\sigma\colon\Sigma\to\Sigma$ which takes $(x_i)_{i\in\mathbb{Z}}$ to $(x_{i+1})_{i\in\mathbb{Z}}$. Notice that the shift simply moves the letter in place i into place $i-1$, ie it shifts every letter one place to the left.

Associated to any two-sided countable Markov shift Σ is the one-sided countable Markov shift

$$\Sigma^+ = \{x = (x_i)\in\mathcal{A}^{\mathbb{N}} \mid t_{x_i x_{i+1}} = 1 \text{ for all } i\in\mathbb{N}\}$$

equipped with a shift map $\sigma\colon\Sigma^+\to\Sigma^+$ which takes $(x_i)_{i\in\mathbb{N}}$ to $(x_{i+1})_{i\in\mathbb{N}}$. In this case, the shift deletes the letter x_1 and moves every other letter one place to the left. There is a natural projection map $p^+\colon\Sigma\to\Sigma^+$ given by $p^+(x) = x^+ = (x_i)_{i\in\mathbb{N}}$, which simply forgets all the terms to the left of x_1. Notice that $p^+\circ\sigma = \sigma\circ p^+$. We will work entirely with one-sided shifts, except in the final section.

One says that (Σ^+,σ) is *topologically mixing* if, for all $a,b\in\mathcal{A}$, there exists $N = N(a,b)$ such that, if $n\ge N$, then there exists $x\in\Sigma$ such that $x_1 = a$ and $x_n = b$. The shift (Σ^+,σ) has the big images and preimages property (BIP) if there exists a finite subset $\mathcal{B}\subset\mathcal{A}$ such that, if $a\in\mathcal{A}$, then there exists $b_0,b_1\in\mathcal{B}$ such that $t_{b_0,a} = 1 = t_{a,b_1}$.

Given a one-sided countable Markov shift (Σ^+,σ) and a function $g\colon\Sigma^+\to\mathbb{R}$, let

$$V_n(g) = \sup\{|g(x)-g(y)| : x,y\in\Sigma^+,\ x_i = y_i \text{ for all } 1\le i\le n\}$$

be the n^{th} *variation* of g. We say that g is *locally Hölder continuous* if there exists $C > 0$ and $\theta \in (0,1)$ such that

$$V_n(g) \le C\theta^n$$

for all $n \in \mathbb{N}$. We say that two locally Hölder continuous functions $f\colon \Sigma^+ \to \mathbb{R}$ and $g\colon \Sigma^+ \to \mathbb{R}$ are *cohomologous* if there exists a locally Hölder continuous function $h\colon \Sigma^+ \to \mathbb{R}$ such that

$$f - g = h - h \circ \sigma.$$

Sarig [37] considers the associated *Gurevich pressure* of a locally Hölder continuous function $g\colon \Sigma^+ \to \mathbb{R}$, given by

$$P(g) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\substack{x \in \mathrm{Fix}^n \\ x_1 = a}} e^{S_n g(x)}$$

for some (any) $a \in \mathcal{A}$ where

$$S_n(g)(x) = \sum_{i=1}^{n} g(\sigma^i(x))$$

is the *ergodic sum* and $\mathrm{Fix}^n = \{x \in \Sigma^+ \mid \sigma^n(x) = x\}$. The pressure of a locally Hölder continuous function f need not be finite, but Mauldin and Urbański [29] provide the following characterization of when $P(f)$ is finite:

Theorem 2.2 (Mauldin and Urbański [29, Theorem 2.1.9]) *Suppose that (Σ^+, σ) is a one-sided countable Markov shift which has BIP and is topologically mixing. If f is locally Hölder continuous, then $P(f)$ is finite if and only if*

$$Z_1(f) = \sum_{a \in \mathcal{A}} e^{\sup\{f(x) : x_1 = a\}} < +\infty.$$

A Borel probability measure m on Σ^+ is said to be a *Gibbs state* for a locally Hölder continuous function $g\colon \Sigma^+ \to \mathbb{R}$ if there exists a constant $B > 1$ and $C \in \mathbb{R}$ such that

$$\frac{1}{B} \le \frac{m([a_1, \dots, a_n])}{e^{S_n g(x) - nC}} \le B$$

for all $x \in [a_1, \dots, a_n]\}$, where $[a_1, \dots, a_n]$ is the *cylinder* consisting of all $x \in \Sigma^+$ such that $x_i = a_i$ for all $1 \le i \le n$. Sarig [39, Theorem 4.9] shows that a locally Hölder continuous function f on a topologically mixing one-sided countable Markov shift with BIP such that $P(f)$ is finite admits a Gibbs state μ_f. Mauldin and Urbański [29, Theorem 2.2.4] show that, if a locally Hölder continuous function f on a topologically mixing one-sided countable Markov shift with BIP admits a Gibbs state, then f admits a unique shift-invariant Gibbs state. We summarize their work in the statement below:

Theorem 2.3 [29, Theorem 2.2.4; 39, Theorem 4.9] *Suppose that (Σ^+,σ) is a one-sided countable Markov shift which has BIP and is topologically mixing. If f is locally Hölder continuous and $P(f)$ is finite, then f admits a unique shift-invariant Gibbs state μ_f.*

The *transfer operator* is a central tool in the thermodynamic formalism. Recall that the *transfer operator* $\mathcal{L}_f\colon C^b(\Sigma^+)\to C^b(\Sigma^+)$ of a locally Hölder continuous function f over Σ^+ is defined by

$$\mathcal{L}_f(g)(x)=\sum_{y\in\sigma^{-1}(x)} e^{f(y)}g(y) \quad \text{for all } x\in\Sigma^+.$$

If (Σ^+,σ) is topologically mixing and has the BIP property, ν is a Borel probability measure for Σ^+ and $(\mathcal{L}_f)^*(\nu)=e^{P(f)}\nu$ (where $(\mathcal{L}_f)^*$ is the dual of transfer operator), then ν is a Gibbs state for f; see Mauldin and Urbański [29, Theorem 2.3.3].

A σ–invariant Borel probability measure m on Σ^+ is said to be an *equilibrium measure* for a locally Hölder continuous function $g\colon\Sigma^+\to\mathbb{R}$ if

$$P(g)=h_\sigma(m)+\int_{\Sigma^+} g\,dm,$$

where $h_\sigma(m)$ is the measure-theoretic entropy of σ with respect to the measure m. Mauldin and Urbański [29] give a criterion guaranteeing the existence of a unique equilibrium state:

Theorem 2.4 [29, Theorem 2.2.9] *Suppose that (Σ^+,σ) is a one-sided countable Markov shift which has BIP and is topologically mixing. If f is locally Hölder continuous, ν_f is a shift-invariant Gibbs state for f and $-\int f\,d\nu_f<+\infty$, then ν_f is the unique equilibrium measure for f.*

We say that $\{g_u\colon\Sigma^+\to\mathbb{R}\}_{u\in M}$ is a *real analytic family* if M is a real analytic manifold and, for all $x\in\Sigma^+$, $u\to g_u(x)$ is a real analytic function on M. Mauldin and Urbański [29, Theorem 2.6.12 and Propositions 2.6.13 and 2.6.14] — see also Sarig [38, Corollary 4; 39, Theorems 5.10 and 5.13] — prove real analyticity properties of the pressure function and evaluate its derivatives. We summarize their results in Theorem 2.5. Here the *variance* of a locally Hölder continuous function $f\colon\Sigma^+\to\mathbb{R}$ with respect to a probability measure m on Σ^+ is given by

$$\operatorname{Var}(f,m)=\lim_{n\to\infty}\frac{1}{n}\int_{\Sigma^+} S_n\Big(\Big(f-\int_{\Sigma^+} f\,dm\Big)^2\Big)\,dm.$$

Theorem 2.5 (Mauldin–Urbański, Sarig) *Suppose that (Σ^+, σ) is a one-sided countable Markov shift which has BIP and is topologically mixing. If $\{g_u : \Sigma^+ \to \mathbb{R}\}_{u \in M}$ is a real analytic family of locally Hölder continuous functions such that $P(g_u) < \infty$ for all u, then $u \to P(g_u)$ is real analytic.*

Moreover, if $v \in T_{u_0}M$ and there exists a neighborhood U of u_0 in M such that $-\int_{\Sigma^+} g_u \, dm_{g_{u_0}} < \infty$ if $u \in U$, then

$$D_v P(g_u) = \int_{\Sigma^+} D_v(g_u(x)) \, dm_{g_{u_0}}$$

and

$$D_v^2 P(g_u) = \mathrm{Var}(D_v g_u, m_{g_{u_0}}) + \int_{\Sigma^+} D_v^2 g_u \, dm_{g_{u_0}},$$

where $m_{g_{u_0}}$ is the unique equilibrium state for g_{u_0}.

2.3 The Stadlbauer–Ledrappier–Sarig coding

Stadlbauer [42] and Ledrappier and Sarig [27] describe a one-sided countable Markov shift (Σ^+, σ) with alphabet $\mathcal{A}$ which encodes the recurrent portion of the geodesic flow on $T^1(\mathbb{H}^2/\Gamma)$. In this section, we will sketch the construction of this coding and recall its crucial properties.

They begin with the classical coding of a free group, as described by Bowen and Series [7]. One begins with a fundamental domain D_0 for a free convex cocompact fuchsian group Γ, containing the origin in the Poincaré disk model, all of whose vertices lie in $\partial\mathbb{H}^2$, such that the set $\mathcal{S}$ of face pairings of D_0 is a minimal symmetric generating set for Γ. One then labels any translate $\gamma(D_0)$ by the group element γ. Any geodesic ray r_z beginning at the origin and ending at $z \in \Lambda(\Gamma)$ passes through an infinite sequence of translates, so we get a sequence $c(z) = (\gamma_k)_{k \in \mathbb{N}}$. One may then turn this into an infinite sequence in $\mathcal{S}$ by considering $b(z) = (\gamma_k \gamma_{k-1}^{-1})_{k \in \mathbb{N}}$ (where we adopt the convention that $\gamma_0 = \mathrm{id}$.) If Γ is convex cocompact, this produces a well-behaved one-sided Markov shift $(\Sigma_{\mathrm{BS}}^+, \sigma)$ with finite alphabet $\mathcal{S}$. The obvious map $\omega \colon \Sigma_{\mathrm{BS}}^+ \to \Lambda(\Gamma)$ which takes $b(z)$ to z is Hölder and $(\Sigma_{\mathrm{BS}}^+, \sigma)$ encodes the recurrent portion of the geodesic flow of $\mathbb{H}^2/\Gamma$.

If one attempts to implement this procedure when Γ is not convex cocompact, then one must omit all geodesic rays which end at a parabolic fixed point and there is no natural way to do this from a coding perspective. Moreover, if one simply restricts ω to the allowable words then ω will not be Hölder in this case. (To see that ω will not be Hölder, choose $x, y \in \Sigma_{\mathrm{BS}}^+$ so that $x_i = y_i = \alpha$ for all $1 \le i \le n$, where α is a parabolic face pairing, and $x_{n+1} \neq y_{n+1}$; then $d_{\Sigma^+_{\mathrm{BS}}}(x, y) = e^{-n}$, while $d_{\partial\mathbb{H}^2}(\omega(x), \omega(y))$ is comparable to $1/n^2$.)

Roughly, the Stadlbauer–Ledrappier–Sarig coding begins with $c(z) = (\gamma_k)$ and clumps together all terms in $b(z) = (\gamma_k \gamma_{k-1}^{-1})$ which lie in a subword which is a high power of a parabolic element. One must then append to our alphabet all powers of minimal word length parabolic elements and disallow infinite words beginning or ending in infinitely repeating parabolic elements. When Γ is geometrically finite, but not of cofinite area, Dal'bo and Peigné [16] implemented this process to powerful effect for geometrically finite fuchsian groups with infinite-area quotients. However, when Γ has cofinite area, the actual description is more intricate. The states Stadlbauer, Ledrappier and Sarig use record a finite amount of information about both the past and the future of the trajectory.

Let $\mathcal{C}$ be the collection of all freely reduced words in $\mathcal{S}$ which have minimal word length in their conjugacy class and generate a maximal parabolic subgroup of Γ. Notice that the minimal word length representative of a conjugacy class of α is unique up to cyclic permutation. (One may in fact choose D_0 so that all but one pair of parabolic elements of $\mathcal{C}$ is conjugate to a face pairing.) Since there are only finitely many conjugacy classes of maximal parabolic subgroups of Γ, $\mathcal{C}$ is finite. They then choose a sufficiently large even number $2N$ so that the length of every element of $\mathcal{C}$ divides $2N$ and let $\mathcal{C}^*$ be the collection of powers of elements of $\mathcal{C}$ of length exactly $2N$. (One may assume that two elements of $\mathcal{C}^*$ share a subword of length at least 2 if and only if they are cyclic permutations of one another.)

Let $\mathcal{A}_1$ be the set of all strings $(b_0, b_1, \dots, b_{2N})$ in $\mathcal{S}$ such that $b_0 b_1 \cdots b_{2N}$ is freely reduced in $\mathcal{S}$ and such that neither $b_1 b_2 \cdots b_{2N}$ nor $b_0 b_1 \cdots b_{2N-1}$ lies in $\mathcal{C}^*$. Let $\mathcal{A}_2$ be the set of all freely reduced strings of the form $(b, w^s, w_1, \dots, w_{k-1}, c)$, where $w = w_1 \cdots w_{2N} \in \mathcal{C}^*$, $b \in \mathcal{S} - \{w_{2N}\}$, $1 \le k \le 2N$, $s \ge 1$ and $c \in \mathcal{S} - \{w_k\}$.

Let $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ and define functions

$$r : \mathcal{A} \to \mathbb{N} \quad \text{and} \quad G : \mathcal{A} \to \Gamma$$

by letting $r(a) = 1$ if $a \in \mathcal{A}_1$ and $r(b, w^s, w_1, \dots, w_{k-1}, c) = s + 1$ otherwise. If $a = (b_0, b_1, \dots, b_{2N}) \in \mathcal{A}_1$, then $G(a) = b_1$. If $a = (b, w^s, w_1 \cdots w_{k-1}, c)$, then let $G(a) = w^{s-1} w_1 \cdots w_{k+1}$. Notice that, by construction, if $n \in \mathbb{N}$, then

$$\#(r^{-1}(n)) \le \#(\mathcal{C}^*)(\#(\mathcal{S})^2)(2N).$$

So, $r^{-1}(n)$ is always nonempty and there exists D such that $r^{-1}(n)$ has size at most D for all $n \in \mathbb{N}$, ie there are at most D states associated to each positive integer.

Given a geodesic ray r_z beginning at the origin and ending at a point z in the set $\Lambda_c(\Gamma)$ of points in the limit set which are not parabolic fixed points, let $c(z) = (\gamma_k)_{k\in\mathbb{N}}$ be the sequence of elements of Γ which record the translates of D_0 through which r_z passes. Let $b(z) = (b_k(z)) = (\gamma_k \gamma_{k-1}^{-1}) \in \mathcal{S}^{\mathbb{N}}$. We then associate to r_z a finite collection of infinite words in $\mathcal{S}^{\mathbb{N}\cup\{0\}}$, by allowing b_0 to be any element of $\mathcal{S}$ such that $b_0 b_1 \cdots b_{2N}$ does not lie in $\mathcal{C}^*$.

Suppose we have a word $(b_k)_{k\in\mathbb{N}\cup\{0\}}$ arising from the previous construction. If $(b_0, b_1, \dots, b_{2N}) \in \mathcal{A}_1$, then let $x_1 = (b_0, b_1, \dots, b_{2N})$ and shift (b_i) rightward by 1 to compute x_2. If not, let x_1 be the unique substring of $b_0 b_1 \cdots b_k \cdots$ which begins at b_0 and is an element of $\mathcal{A}_2$. Then $x_1 = (b_0, w^s, w_1 \cdots w_{k-1}, b_v)$ for some $w \in \mathcal{C}^*$, $s \in \mathbb{N}$ and $v = 2Ns + k - 1$. In this case, we shift (b_i) rightward by $2N(s-1)+k+1$ to compute x_2. One then simply proceeds iteratively. By construction, if $x_i \in \mathcal{A}_2$, then x_{i+1} must lie in $\mathcal{A}_1$.

Examples If Γ uniformizes a once-punctured torus, then $\mathcal{S} = \{\alpha, \alpha^{-1}, \beta, \beta^{-1}\}$ is a minimal symmetric generating set for Γ and

$$\mathcal{C} = \big\{\alpha\beta\alpha^{-1}\beta^{-1}, \beta\alpha^{-1}\beta^{-1}\alpha, \alpha^{-1}\beta^{-1}\alpha\beta, \beta^{-1}\alpha\beta\alpha^{-1}, \beta\alpha\beta^{-1}\alpha^{-1}, \alpha\beta^{-1}\alpha^{-1}\beta, \beta^{-1}\alpha^{-1}\beta\alpha, \alpha^{-1}\beta\alpha\beta^{-1}\big\}.$$

If Γ uniformizes a four times–punctured sphere, then one may choose D_0 so that $\mathcal{S} = \{\alpha, \alpha^{-1}, \beta, \beta^{-1}, \gamma, \gamma^{-1}\}$ and

$$\mathcal{C} = \{\alpha, \alpha^{-1}, \beta, \beta^{-1}, \gamma, \gamma^{-1}, \alpha\beta\gamma, \beta\gamma\alpha, \gamma\alpha\beta, \gamma^{-1}\beta^{-1}\alpha^{-1}, \beta^{-1}\alpha^{-1}\gamma^{-1}, \alpha^{-1}\gamma^{-1}\beta^{-1}\}.$$

The following proposition encodes crucial properties of the coding:

Proposition 2.6 (Ledrappier and Sarig [27, Lemma 2.1] and Stadlbauer [42]) *Suppose that $\mathbb{H}^2/\Gamma$ is a finite-area hyperbolic surface, then (Σ^+, σ) is topologically mixing, has the big images and preimages property (BIP), and there exists a locally Hölder continuous finite-to-one map*

$$\omega\colon \Sigma^+ \to \Lambda(\Gamma)$$

such that $\omega(x) = \lim(G(x_1)\cdots G(x_n))(0)$ and $\omega(x) = G(x_1)\omega(\sigma(x))$. Moreover, if γ is a hyperbolic element of Γ, then there exists $x \in \mathrm{Fix}^n$ for some $n \in \mathbb{N}$, unique up to cyclic permutation, such that γ is conjugate to $G(x_1)\cdots G(x_n)$.

Notice that every element of $\mathcal{A}$ can be preceded and succeeded by some element of $\mathcal{A}_1$, so (Σ^+, σ) clearly has BIP. The topological mixing property is similarly easy to see directly from the definition, so the main claim of this proposition is that ω is locally Hölder continuous.

Another crucial property of the coding is that the translates of the origin associated to the Stadlbauer–Ledrappier–Sarig coding approach points in the limit set conically (see property (1) on page 15 in Ledrappier and Sarig [27]).

Lemma 2.7 (Ledrappier and Sarig [27, property (1) on page 15]) *Given $y \in \mathbb{H}^2$, there exists $L > 0$ such that, if $x \in \Sigma^+$ and $n \in \mathbb{N}$, then*

$$d(G(x_1)G(x_2)\cdots G(x_n)(0), \overrightarrow{y\omega(x)}) \le L.$$

Since the proof of Lemma 2.7 appears in the middle of a rather technical discussion in [27], we will sketch a proof in our language. Choose a compact subset $\widehat{K}$ of $\mathbb{H}^2/\Gamma$ so that its complement is a collection of cusp regions bounded by curves which are images of horocycles in $\mathbb{H}^2$. Without loss of generality we may assume that y is the origin in the Poincaré disk model for $\mathbb{H}^2$. Notice that, if the portion of $\overrightarrow{b\omega(x)}$ between $\gamma_s(D_0)$ and $\gamma_{s+t}(D_0)$ lies entirely in the complement of the preimage of $\widehat{K}$, and $t > s$, then $\gamma_{s+t}\gamma_s^{-1}$ is a subword of a power of an element in $\mathcal{C}$. Let K be the intersection of the preimage of $\widehat{K}$ with D_0. Notice that we may assume that $y \in K$ (by perhaps enlarging $\widehat{K}$). Suppose the last $2N+1$ letters of x_n are $b_r \cdots b_{r+2N}$, then $\overrightarrow{0\omega(x)}$ intersects one of $\gamma_r(K), \dots, \gamma_{r+2N}(K)$ (since otherwise $b_r \cdots b_{r+2N-1}$ or $b_{r+1} \cdots b_{r+2N+1}$ would lie in $\mathcal{C}^*$, which is disallowed). But then

$$d(G(x_1)\cdots G(x_n)(y), \overrightarrow{y\omega(x)}) \le R + \operatorname{diam}(K),$$

where

$$R = \max\bigl\{d(y, (s_1 \cdots s_p)(y)) \mid s_i \in \mathcal{S},\ p \in \{1, \dots, 2N\}\bigr\}.$$

3 Roof functions for quasifuchsian groups

If $\rho \in \mathrm{QC}(\Gamma)$, we define a *roof function* $\tau_\rho\colon \Sigma^+ \to \mathbb{R}$ by setting

$$\tau_\rho(x) = B_{\xi_\rho(\omega(x))}\bigl(b_0, \rho(G(x_1))(b_0)\bigr),$$

where $b_0 = (0,0,1)$ and $B_z(x,y)$ is the Busemann function based at $z \in \partial\mathbb{H}^3$ which measures the signed distance between the horoballs based at z through x and y. In the

Poincaré upper half-space model, we write the Busemann function explicitly as

$$\widehat{B}_z(p,q) = \log \frac{|p-z|^2 h(p)}{|q-z|^2 h(q)},$$

where $z \in \mathbb{C} \subset \partial\mathbb{H}^3$, $p,q \in \mathbb{H}^3$ and $h(p)$ is the Euclidean height of p above the complex plane, and $\widehat{B}_\infty(p,q) = h(p)/h(q)$.

It follows from the cocycle property of the Busemann function that

$$S_m\tau_\rho(x) = \sum_{i=0}^{m-1} \tau_\rho(\sigma^i(x)) = B_{\xi_\rho(\omega(x))}\big(b_0, \rho(G(x_1)\cdots G(x_m))(b_0)\big).$$

In particular, if $x = \overline{(x_1,\dots,x_m)} \in \Sigma^+$, then

$$S_m\tau_\rho(x) = \ell_\rho(G(x_1)\cdots G(x_m)).$$

We say that the roof function τ_ρ is *eventually positive* if there exists $C > 0$ and $N \in \mathbb{N}$ such that, if $n \geq N$ and $x \in \Sigma^+$, then $S_n\tau_\rho(x) \geq C$.

The following lemma records crucial properties of our roof functions. It generalizes similar results of Ledrappier and Sarig [27, Lemmas 2.2 and 3.1] in the fuchsian setting.

Proposition 3.1 *The family $\{\tau_\rho\}_{\rho\in \mathrm{QC}(\Gamma)}$ of roof functions is a real analytic family of locally Hölder continuous, eventually positive functions.*

Moreover, if $\rho \in \mathrm{QC}(\Gamma)$, then there exists $C_\rho > 0$ and $R_\rho > 0$ such that

$$2\log r(x_1) - C_\rho \leq \tau_\rho(x) \leq 2\log r(x_1) + C_\rho$$

and

$$|S_n\tau_\rho(x) - d(b_0, G(x_1)\cdots G(x_n)(b_0))| \leq R_\rho$$

for all $x \in \Sigma^+$ and $n \in \mathbb{N}$.

Proof Since $\xi_\rho(q)$ varies complex analytically in ρ for all $q \in \Lambda(\Gamma)$, by Lemma 2.1, and $B_z(b_0, y)$ is real analytic in $z \in \widehat{\mathbb{C}}$ and $y \in \mathbb{H}^3$, we see that $\tau_\rho(x)$ varies analytically over $\mathrm{QC}(\Gamma)$ for all $x \in \Sigma^+$.

Recall — see Douady and Earle [17] — that there exists $K = K(\rho) > 1$ and a ρ–equivariant K–bilipschitz map $\phi\colon \mathbb{H}^2 \to \mathbb{H}^3$ such that $\phi(y_0) = b_0$, where y_0 is the origin in the disk model for $\mathbb{H}^2$. Therefore, if L is the constant from Lemma 2.7 and $x \in \Sigma^+$, then $\rho(G(x_1)\cdots G(x_n))(b_0)$ lies within KL of the K–bilipschitz ray

$\phi(\overrightarrow{y_0\omega(x)})$. The fellow-traveler property for $\mathbb{H}^3$ implies that there exists $R=R(K)>0$ such that any K–bilipschitz geodesic ray lies a Hausdorff distance at most R from the geodesic ray with the same endpoints. Therefore, if $M=KL+R$, then, for all $n\in\mathbb{N}$,

$$d\big(\rho(G(x_1)\cdots G(x_n))(b_0),\overrightarrow{b_0\xi_\rho(\omega(x))}\big)\le M.$$

We next obtain our claimed bounds on the roof function. If $x\in\Sigma^+$, then

$$|\tau_\rho(x)|\le d\big(\rho(G(x_1))(b_0),b_0\big),$$

so, if $a\in\mathcal{A}$, there exists C_a such that, if $x_1=a$, then $|\tau_\rho(x)|\le C_a$. Since our alphabet is infinite, our work is not done.

If $w\in\mathcal{C}^*$, we may normalize so that $\rho(w)(z)=z+1$ and $b_0=(0,0,b_w)$ in the upper half-space model for $\mathbb{H}^3$. If $z\in\mathbb{C}\subset\partial\mathbb{H}^3$ and $r>0$, we let $B(z,r)$ denote the Euclidean ball of radius r about z in $\mathbb{C}$. Since g_a has length at most $2N+1$ in the alphabet $\mathcal{S}$, we may define

$$c_w=\max\{|\rho(g_a)(b_0)|:G(a)=w^sg_a\text{ for some }a\in\mathcal{A}_2\},$$

where $|\rho(g_a)(b_0)|$ is the Euclidean distance from $\rho(g_a)(b_0)$ to $0=(0,0,0)$. Suppose that $x\in\Sigma^+$, $r(x_1)\ge 2$ and $G(x_1)=w^sg_a$, where $s=r(a)-2$. By definition, $\rho(g_a)(b_0)\in B(0,c_w)$, so

$$\rho(w^sg_a)(b_0)=\rho(w^s)(\rho(g_a)(b_0))\in\rho(w^s)(B(0,c_w))=B(s,c_w).$$

Let $S=\max\{e^Mc_w:w\in\mathcal{C}^*\}$. If $s>S$, then b_0 does not lie in $B(s,e^Mc_w)$, but $\overrightarrow{b_0\xi_\rho(\omega(x))}$ passes through $B(s,e^Mc_w)$, which implies that $\xi_\rho(\omega(x))\in B(s,e^Mc_w)$. It then follows from our formula for the Busemann function that

$$\begin{aligned}\tau_\rho(x)&=\log\frac{|b_0-\xi_\rho(\omega(x))|^2h(\rho(w^sg_a)(b_0))}{|\rho(w^sg_a)(b_0)-\xi_\rho(\omega(x))|^2h(b_0)}\\&\le\log\frac{(b_w^2+(s+e^Mc_w)^2)h(\rho(g_a)(b_0))}{h(\rho(g_a)(b_0))^2b_w}=\log\frac{b_w^2+(s+e^Mc_w)^2}{h(\rho(g_a)(b_0))b_w}.\end{aligned}$$

Similarly,

$$\tau_\rho(x)\ge\log\frac{(b_w^2+(s-e^Lc_w)^2)h(\rho(g_a)(b_0))}{\big(h(\rho(g_a)(b_0))^2+e^{2M}c_w^2\big)b_w}.$$

Since there are only finitely many choices of g_a, it is easy to see that there exists C_w such that

$$2\log r(x_1)-C_w\le\tau_\rho(x)\le 2\log r(x_1)+C_w$$

whenever $x \in \Sigma^+$, $r(x_1) > S+2$ and $G(x_1) = w^s g_a$. Since there are only finitely many w in $\mathcal{C}^*$ and only finitely many words a with $r(a) \le S+2$, we see that there exists C_ρ such that

$$2\log r(x_1) - C_\rho \le \tau_\rho(x) \le 2\log r(x_1) + C_\rho$$

for all $x \in \Sigma^+$.

We next show that τ_ρ is locally Hölder continuous. Since ω is locally Hölder continuous, there exists A and $\alpha > 0$ such that, if $x, y \in \Sigma^+$ and $x_i = y_i$ for $1 \le i \le n$, then

$$d(\omega(x), \omega(y)) \le Ae^{-\alpha n}.$$

Since ξ_ρ is Hölder, there exists C and $\beta > 0$ such that $d(\xi_\rho(z), \xi_\rho(w)) \le Cd(z,w)^\beta$ for all $z, w \in \Lambda(\Gamma)$, so

$$d\big(\xi_\rho(\omega(x)), \xi_\rho(\omega(y))\big) \le CA^\beta e^{-\alpha\beta n}.$$

If $a \in \mathcal{A}$, then let

$$D_a = \sup\left\{\left|\frac{\partial}{\partial z}\Big|_{z=z_0} B_z\big(b_0, \rho(G(a))(b_0)\big)\right| : z_0 = \xi_\rho(\omega(x)) \text{ and } x_1 = a\right\},$$

so

$$\sup\{|\tau_\rho(x) - \tau_\rho(y)| : x, y \in [a, x_2, \dots, x_n]\} \le D_a CA^\beta e^{-\alpha\beta n}.$$

However, the best general estimate one can have on D_a is $O(r(a))$, so we will have to dig a little deeper.

We again work in the upper half-space model, and assume that $r(a) > S+2$ and $G(a) = w^s g_a$, where $s = r(a) - 2$ and normalize as before so that $\rho(w)(z) = z+1$. We then map the limit set into the boundary of the upper half-space model by setting $\hat{\xi}_\rho = T \circ \xi_\rho$, where T is a conformal automorphism which takes the Poincaré ball model to the upper half-space model and takes the fixed point of $\rho(w)$ to ∞. Notice that T is K_w–bilipschitz on $T^{-1}(B(0, e^M c_w))$. Therefore, if $x, y \in [a, x_2, \dots, x_n]$, then

$$|\hat{\xi}_\rho(x) - \hat{\xi}_\rho(y) = |\hat{\xi}_\rho(w^{-s}(x)) - \hat{\xi}_\rho(w^{-s}(x))| \le K_w CA^\beta e^{-\alpha\beta(n-1)}.$$

Moreover, there exists D_w such that

$$\left|\frac{\partial}{\partial z}\Big|_{z=z_0} \hat{B}_z\big(b_0, \rho(G(a))(b_0)\big)\right| \le D_w$$

if $z_0 \in \rho(w)^s(B(0, e^M c_w))$, so

$$\sup\{|\tau_\rho(x) - \tau_\rho(y)| : x, y \in [a, x_2, \dots, x_n]\} \le K_w D_w CA^\beta e^{-\alpha\beta(n-1)}.$$

Since there are only finitely many a where $r(a) \le S+2$ and only finitely many choices of w, our bounds are uniform over $\mathcal{A}$ and so τ_ρ is locally Hölder continuous.

It remains to check that τ_ρ is eventually positive. Since

$$d\big(\rho(\gamma_n)(b_0), \overrightarrow{b_0\xi_\rho(\omega(x))}\big) \le M$$

for all $n \in \mathbb{N}$, we see that

$$|S_n\tau_\rho(x) - d(b_0, G(x_1)\cdots G(x_n)(b_0))| \le 2M = R_\rho.$$

Since the set

$$\mathcal{B} = \{\gamma \in \Gamma \mid d(\rho(\gamma)(b_0), b_0) \le 2R_\rho\}$$

is finite, there exists $\widehat{N}$ such that, if γ has word length at least $\widehat{N}$ (in the generators, given $\mathcal{S}$), then γ does not lie in $\mathcal{B}$. Therefore, if $n \ge \widehat{N}$ and $x \in \Sigma^+$, then $S_n\tau_\rho(x) > R_\rho > 0$. Thus, τ_ρ is eventually positive. □

It is a standard feature of the thermodynamic formalism that one may replace an eventually positive roof function by a roof function which is strictly positive and cohomologous to the original roof function. (For a statement and proof which includes the current situation, see [8, Lemma 3.3].)

Corollary 3.2 *If $\rho \in \mathrm{QC}(\Gamma)$, there exists a locally Hölder continuous function $\hat{\tau}_\rho$ and $c > 0$ such that $\hat{\tau}_\rho(x) \ge c$ for all $x \in \Sigma^+$ and $\hat{\tau}_\rho$ is cohomologous to τ_ρ.*

4 Phase transition analysis

We begin by extending Kao's phase transition analysis — see Kao [23, Theorem 4.1] — which characterizes which linear combinations of a pair of roof functions have finite pressure. The primary use of this analysis will be in the case of a single roof function, ie when $a = 1$ and $b = 0$. However, we will use the full force of this result in the proof of our Manhattan curve theorem; see Theorem 6.1.

Theorem 4.1 *If $\rho, \eta \in \mathrm{QC}(\Gamma)$, $t \in \mathbb{R}$ and $a + b > 0$, then $P(-t(a\tau_\rho + b\tau_\eta))$ is finite if and only if $t > 1/2(a+b)$. Moreover, $P(-t(a\tau_\rho + b\tau_\eta))$ is monotone decreasing and analytic in t on $(1/2(a+b), \infty)$, and*

$$\lim_{t\to 1/2(a+b)^+} P(-t(a\tau_\rho + b\tau_\eta)) = +\infty.$$

If, in addition, $a, b \ge 0$, then

$$\lim_{t\to\infty} P(-t(a\tau_\rho + b\tau_\eta)) = -\infty.$$

Riquelme and Velozo [34, Theorem 1.4] previously established results closely related to Theorem 4.1 in the more general setting of negatively curved manifolds with bounded geometry.

Proof Recall from Theorem 2.2 that, since $-t(a\tau_\rho+b\tau_\eta)$ is locally Hölder continuous and (Σ^+,σ) is a one-sided, topologically mixing countable Markov shift with BIP, $P(-t(a\tau_\rho+b\tau_\eta))$ is finite if and only if $Z_1(-t(a\tau_\rho+b\tau_\eta))<+\infty$. Since there exists $D\in\mathbb{N}$ such that $\#r^{-1}(n)\le D$ for all $n\in\mathbb{N}$, Proposition 3.1 implies that

$$Z_1(-t(a\tau_\rho+b\tau_\eta))\le D\sum_{n=1}^{\infty}e^{-t(a+b)(2\log n-\max\{C_\rho,C_\eta\})},$$

so $P(-t(a\tau_\rho+b\tau_\eta))<+\infty$ if $t>1/2(a+b)$. Similarly, since $r^{-1}(n)$ is nonempty if $n\ge 1$, we see that

$$Z_1(-t(a\tau_\rho+b\tau_\eta))\ge \sum_{n=1}^{\infty}e^{-t(a+b)(2\log n+\max\{C_\rho,C_\eta\})},$$

so $P(-t(a\tau_\rho+b\tau_\eta))=+\infty$ if $t\le 1/2(a+b)$ and

$$\lim_{t\to 1/2(a+b)^+} Z_1(-t(a\tau_\rho+b\tau_\eta))=+\infty.$$

It follows from the definition that $P(-t(a\tau_\rho+b\tau_\eta))$ is monotone decreasing in t and Theorem 2.5 implies that it is analytic in t on $(1/2(a+b),\infty)$. In the proof of [29, Theorem 2.1.9], Mauldin and Urbański show that, given a locally Hölder continuous function f on a one-sided countable Markov shift which is topologically mixing and has property BIP, there exist constants $q,s,M,m>0$ such that, for any $n\in\mathbb{N}$,

$$\sum_{i=n}^{n+s(n-1)} Z_i(f)\geqslant \frac{e^{-M+(M-m)n}}{q^{n-1}}Z_1(f)^n,$$

where, if E^n is the set of allowable words of length n in $\mathcal{A}$, then

$$Z_n(f)=\sum_{w\in E^n}e^{\sup\{S_nf(x)\mid x_i=w_i \text{ for all } 1\le i\le n\}}\quad\text{and}\quad \lim\frac{1}{n}\log Z_n(f)=P(f).$$

It follows that, for all n, there exists $A>0$ and $\hat{n}\in[n,n+s(n-1)]$ such that $Z_{\hat{n}}\ge A^nZ_1(f)^n$, so $P(f)\ge(1/(1+s))Z_1(f)-\log A$. Therefore,

$$\lim_{t\to 1/2(a+b)^+} P(-t(a\tau_\rho+b\tau_\eta))=+\infty.$$

If $a, b \geq 0$ and $x \in \mathrm{Fix}^n$, then $S_n(a\tau_\rho + b\tau_\eta)(x) > 0$, so, if $t > 1$, then

$$\sum_{\substack{x \in \mathrm{Fix}^n \\ x_1 = a}} e^{S_n(-t(a\tau_\rho + b\tau_\eta))(x)} \leq \frac{1}{t} \sum_{\substack{x \in \mathrm{Fix}^n \\ x_1 = a}} e^{S_n(-a\tau_\rho - b\tau_\eta)(x)}$$

since $c^t \leq (1/t)c$ if $0 \leq c \leq 1$ and $t > 1$. Therefore,

$$P(-t(a\tau_\rho + b\tau_\eta)) \leq P(-a\tau_\rho - b\tau_\eta) - \log t,$$

so $\lim_{t\to\infty} P(-t(a\tau_\rho + b\tau_\eta)) = -\infty$. □

5 Entropy and Hausdorff dimension

Theorem 4.1 implies that, if $\rho \in \mathrm{QC}(\Gamma)$, then there is a unique solution $h(\rho) > \frac{1}{2}$ to $P(-h(\rho)\tau_\rho) = 0$. This unique solution $h(\rho)$ is the *topological entropy* of ρ; see the discussion in Kao [23, Section 5]. Theorem 2.5 and the implicit function theorem then imply that $h(\rho)$ varies analytically over $\mathrm{QC}(\Gamma)$, generalizing a result of Ruelle [36] in the convex cocompact case. Since the entropy $h(\rho)$ is invariant under conjugation, we obtain analyticity of entropy over $\mathrm{QF}(S)$. We recall that Schapira and Tapie [40, Theorem 6.2] previously established that the entropy is C^1 on $\mathrm{QF}(S)$.

Theorem 5.1 *If S is a compact hyperbolic surface with nonempty boundary, then the topological entropy varies analytically over* $\mathrm{QF}(S)$.

Sullivan [45] showed that the topological entropy $h(\rho)$ agrees with the Hausdorff dimension of the limit set $\Lambda(\rho(\Gamma))$.

Theorem 5.2 (Sullivan [45; 47]) *If $\rho \in \mathrm{QC}(\Gamma)$, then its topological entropy $h(\rho)$ is the exponential growth rate of the number of closed geodesics of length less than T in $N_\rho = \mathbb{H}^3/\rho(\Gamma)$. Moreover, $h(\rho)$ is the Hausdorff dimension of the limit set $\Lambda(\rho(\Gamma))$.*

Theorems 5.1 and 5.2 together imply that the Hausdorff dimension of the limit set varies analytically.

Corollary 5.3 *The Hausdorff dimension of $\Lambda(\rho(\Gamma))$ varies analytically over* $\mathrm{QC}(\Gamma)$.

Remarks (1) Sullivan [47] also showed that $h(\rho)$ is the critical exponent of the Poincaré series

$$Q_\rho(s) = \sum_{\gamma \in \Gamma} e^{-sd(b_0, \rho(\gamma)(b_0))},$$

ie $Q_\rho(s)$ diverges if $s < h(\rho)$ and converges if $s > h(\rho)$.

(2) Bowen [6] showed that, if $\rho \in \mathrm{QF}(S)$ and S is a closed surface, then $h(\rho) \geq 1$, with equality if and only if ρ is fuchsian. Sullivan [44, page 66] — see also Xie [49] — observed that Bowen's rigidity result extends to the case when $\mathbb{H}^2/\Gamma$ has finite area.

6 Manhattan curves

If $\rho, \eta \in \mathrm{QC}(\Gamma)$, we define, following Burger [12], the *Manhattan curve*

$$\mathcal{C}(\rho, \eta) = \{(a, b) \in D \mid P(-a\tau_\rho - b\tau_\eta) = 0\},$$

where $D = \{(a, b) \in \mathbb{R}^2 \mid a, b \geq 0 \text{ and } (a, b) \neq (0, 0)\}$. Notice that, since the Gurevich pressure is defined in terms of lengths of closed geodesics, if $\hat{\rho}$ is conjugate (or complex conjugate) to ρ and $\hat{\eta}$ is conjugate (or complex conjugate) to η, then $\mathcal{C}(\rho, \eta) = \mathcal{C}(\hat{\rho}, \hat{\eta})$.

One may give an alternative characterization by noticing that $P(-ab_\rho - b\tau_\eta) = 0$ if and only if

$$h^{a,b}(\rho, \eta) = \lim \frac{1}{T} \log \#\{[\gamma] \in [\Gamma] \mid 0 < a\ell_\rho(\gamma) + b\ell_\eta(\gamma) \leq T\} = 1,$$

where $[\Gamma]$ is the collection of conjugacy classes in Γ. Moreover, $h^{a,b}(\rho, \eta)$ is also the critical exponent of

$$Q^{a,b}_{\rho,\eta}(s) = \sum_{\gamma \in \Gamma} e^{-s\left(ad(0,\rho(\gamma)(0)) + bd(0,\eta(\gamma)(0))\right)}$$

(see Kao [22, Theorem 4.8, Remark 4.9 and Lemma 4.10]).

Theorem 6.1 *If* $\rho, \eta \in \mathrm{QC}(\Gamma)$, *then* $\mathcal{C}(\rho, \eta)$

(1) *is a closed subsegment of an analytic curve*,

(2) *has endpoints* $(h(\rho), 0)$ *and* $(0, h(\eta))$, *and*

(3) *is strictly convex*, *unless* ρ *and* η *are conjugate in* $\mathrm{Isom}(\mathbb{H}^3)$.

Moreover, *the tangent line to* $\mathcal{C}(\rho, \eta)$ *at* $(h(\rho), 0)$ *has slope*

$$-\frac{\int \tau_\eta \, dm_{-h(\rho)\tau_\rho}}{\int \tau_\rho \, dm_{-h(\rho)\tau_\rho}}.$$

Burger [12] established Theorem 6.1 for convex cocompact fuchsian groups, with the exception of the analyticity of the Manhattan curve, which was established by Sharp [41].

Notice that, if ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$, then $\tau_\rho = \tau_\eta$, so $\mathcal{C}(\rho,\eta)$ is a straight line. We will need the following technical result in the proof of Theorem 6.1:

Lemma 6.2 *If $\rho,\eta,\theta \in \mathrm{QC}(\Gamma)$, $2(a+b) > 1$ and $P(-a\tau_\rho - b\tau_\eta) = 0$, then there exists a unique equilibrium state $m_{-a\tau_\rho - b\tau_\eta}$ for $-a\tau_\rho - b\tau_\eta$ and*

$$0 < \int_{\Sigma^+} \tau_\theta \, dm_{-a\tau_\rho - b\tau_\eta} < +\infty.$$

Proof Notice that, since $P(-a\tau_\rho - b\tau_\eta) = 0$, there exists a unique shift-invariant Gibbs state $m_{-a\tau_\rho - b\tau_\eta}$ for $-a\tau_\rho - b\tau_\eta$; see Theorem 2.3. However, by [29, Lemma 2.2.8],

$$\int_{\Sigma^+} a\tau_\rho + b\tau_\eta \, dm_{-a\tau_\rho - b\tau_\eta} < +\infty$$

if and only if

$$\sum_{s\in\mathcal{A}} I(a\tau_\rho + b\tau_\eta, s) e^{I(-a\tau_\rho - b\tau_\eta, s)} < \infty,$$

where $I(f,s) = \inf\{f(x) \mid x \in \Sigma,\ x_1 = s\}$. But, by Proposition 3.1,

$$\begin{aligned}\sum_{a\in\mathcal{A}} &\inf(a\tau_\rho + b\tau_\eta|_{[a]}) e^{\inf(-a\tau_\rho - b\tau_\eta|_{[a]})} \\ &\le D \sum_{n\in\mathbb{N}} (|a|C_\rho + |b|C_\eta + 2(a+b)\log n) e^{|a|C_\rho + |b|C_\eta - 2(a+b)\log n} \\ &= D e^{|a|C_\rho + |b|C_\eta} \sum_{n\in\mathbb{N}} \frac{|a|C_\rho + |b|C_\eta + 2(a+b)\log n}{n^{2(a+b)}},\end{aligned}$$

which converges, since $2(a+b) > 1$. Theorem 2.4 then implies that $dm_{-a\tau_\rho - b\tau_\eta}$ is the unique equilibrium state for $-a\tau_\rho - b\tau_\eta$.

Proposition 3.1 implies that there exists $B > 1$ such that, if n is large enough, then

$$\frac{1}{B} \le \frac{\tau_\theta(x)}{a\tau_\rho(x) + b\tau_\eta(x)} \le B$$

for all $x \in \Sigma^+$ such that $r(x_1) > n$. (For example, if $\log n > 4\max\{aC_\rho + bC_\eta, C_\theta, 1\}$, then we may choose $B = 8(a+b)$.) Since τ_θ is locally Hölder continuous, it is bounded on the remainder of Σ^+. Therefore, since $\int_{\Sigma^+} a\tau_\rho + b\tau_\eta \, dm_{-a\tau_\rho - b\tau_\eta} < +\infty$, we see that

$$\int_{\Sigma^+} \tau_\theta \, dm_{-a\tau_\rho - b\tau_\eta} < +\infty.$$

Now notice that, since τ_θ is cohomologous to a positive function $\hat{\tau}_\theta$, by Corollary 3.2,

$$\int_{\Sigma^+} \tau_\theta \, dm_{-a\tau_\rho - b\tau_\eta} = \int_{\Sigma^+} \hat{\tau}_\theta \, dm_{-a\tau_\rho - b\tau_\eta} > 0. \qquad \square$$

Proof of Theorem 6.1 Recall that $t = h(\rho)$ is the unique solution to $P(-t\tau_\rho) = 0$ (see the discussion at the beginning of Section 5). So, the intersection of the Manhattan curve with the boundary of D consists of the points $(h(\rho), 0)$ and $(0, h(\eta))$.

Let
$$\widehat{D} = \left\{(a,b) \in \mathbb{R}^2 \mid a+b > \tfrac{1}{2}\right\}.$$

Theorem 4.1 implies that P is finite on $\widehat{D}$. Lemma 6.2 implies that, if $a, b \in \widehat{D}$ and $P(-a\tau_\rho - b\tau_\eta) = 0$, then there is an equilibrium state $m_{-a\tau_\rho - b\tau_\eta}$ for $-a\tau_\rho - b\tau_\eta$ and that $\int_{\Sigma^+} \tau_\theta \, dm_{-a\tau_\rho - b\tau_\eta}$ is finite for all $\theta \in \mathrm{QC}(\Gamma)$. Theorem 2.5 then implies that
$$\frac{\partial}{\partial a} P(-a\tau_\rho - b\tau_\eta) = \int_{\Sigma^+} -\tau_\rho \, dm_{-a\tau_\rho - b\tau_\eta},$$
$$\frac{\partial}{\partial b} P(-a\tau_\rho - b\tau_\eta) = \int_{\Sigma^+} -\tau_\eta \, dm_{-a\tau_\rho - b\tau_\eta}.$$
Since $\int_{\Sigma^+} -\tau_\rho \, dm_{-a\tau_\rho - b\tau_\eta}$ and $\int_{\Sigma^+} -\tau_\eta \, dm_{-a\tau_\rho - b\tau_\eta}$ are both nonzero, P is a submersion on $\widehat{D}$. Since P is analytic on $\widehat{D}$, the implicit function theorem then implies that
$$\widehat{\mathcal{C}}(\rho,\eta) = \{(a,b) \in \widehat{D} \mid P(-a\tau_\rho - b\tau_\eta) = 0\}$$
is an analytic curve and that, if $(a,b) \in \mathcal{C}(\rho,\eta)$, then the slope of the tangent line to $\mathcal{C}(\rho,\eta)$ at (a,b) is given by
$$c(a,b) = -\frac{\int_{\Sigma^+} \tau_\eta \, dm_{-a\tau_\rho - b\tau_\eta}}{\int_{\Sigma^+} \tau_\rho \, dm_{-a\tau_\rho - b\tau_\eta}}.$$
Notice that $\mathcal{C}(\rho,\eta)$ is the lower boundary of the region
$$\widehat{\mathcal{C}}(\rho,\eta) = \{(a,b) \mid Q^{a,b}_{\rho,\eta}(1) < \infty\}.$$
The Hölder inequality implies that, if $(a,b), (c,d) \in \widehat{\mathcal{C}}(\rho,\eta)$ and $t \in [0,1]$, then
$$Q^{ta+(1-t)c,\,tb+(1-t)d}_{\rho,\eta} \le Q(a,b)^t \, Q(c,d)^{1-t},$$
so $\widehat{\mathcal{C}}(\rho,\eta)$ is convex. Therefore, $\mathcal{C}(\rho,\eta)$ is convex.

A convex analytic curve is strictly convex if and only if it is not a line, so it remains to show that ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$ if $\mathcal{C}(\rho,\eta)$ is a straight line. So suppose that $\mathcal{C}(\rho,\eta)$ is a straight line with slope $c = -h(\rho)/h(\eta)$. In particular,

$$\frac{h(\rho)}{h(\eta)} = -c = -c(h(\rho),0) = \frac{\int_{\Sigma^+} \tau_\eta \, dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+} \tau_\rho \, dm_{-h(\rho)\tau_\rho}} = -c(0,h(\eta)) = \frac{\int_{\Sigma^+} \tau_\eta \, dm_{-h(\eta)\tau_\eta}}{\int_{\Sigma^+} \tau_\rho \, dm_{-h(\eta)\tau_\eta}}. \tag{1}$$

By definition,

$$h(m_{-h(\eta)\tau_\eta}) - h(\eta)\int_{\Sigma^+} \tau_\eta \, dm_{-h(\eta)\tau_\eta} = 0,$$

so, applying (1), we see that

$$\begin{aligned} h(m_{-h(\eta)\tau_\eta}) - h(\rho)\int_{\Sigma^+} \tau_\rho \, dm_{-h(\eta)\tau_\eta} \\ &= h(\eta)\int_{\Sigma^+} \tau_\eta \, dm_{-h(\eta)\tau_\eta} - h(\rho)\int_{\Sigma^+} \tau_\rho \, dm_{-h(\eta)\tau_\eta} \\ &= 0. \end{aligned}$$

Since $P(-h(\rho)\tau_\rho) = 0$, this implies that $m_{-h(\eta)\tau_\eta}$ is an equilibrium measure for $-h(\rho)\tau_\rho$. Therefore, by uniqueness of equilibrium measures, $m_{-h(\eta)\tau_\eta} = m_{-h(\eta)\tau_\rho}$. Sarig [39, Theorem 4.8] showed that this only happens when $-h(\rho)\tau_\rho$ and $-h(\eta)\tau_\eta$ are cohomologous, so the Livsic theorem [39, Theorem 1.1] (see also Mauldin and Urbański [29, Theorem 2.2.7]) implies that

$$\ell_\rho(\gamma) = \frac{h(\eta)}{h(\rho)}\ell_\eta(\gamma)$$

for all $\gamma \in \Gamma$. Kim [25, Theorem 3] proved that, if $\ell_\rho(\gamma) = c\ell_\eta(\gamma)$ for all $\gamma \in \Gamma$, then ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$. □

As a nearly immediate corollary, one obtains a generalization of the rigidity results of Bishop and Steger [3] and Burger [12]:

Corollary 6.3 *If $\rho, \eta \in \mathrm{QC}(\Gamma)$ and $(a,b) \in D$, then*

$$h^{a,b}(\rho,\eta) \le \frac{h(\rho)h(\eta)}{bh(\rho) + ah(\eta)}$$

with equality if and only if ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$*.*

7 Pressure intersection

We define the *pressure intersection* on $\mathrm{QC}(\Gamma) \times \mathrm{QC}(\Gamma)$, given by

$$I(\rho,\eta) = \frac{\int_{\Sigma^+} \tau_\eta \, dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+} \tau_\rho \, dm_{-h(\rho)\tau_\rho}}.$$

It follows from Lemma 6.2 that $I(\rho,\eta)$ is well defined. We also define a *renormalized pressure intersection*

$$J(\rho,\eta) = \frac{h(\eta)}{h(\rho)} I(\rho,\eta).$$

We notice that the pressure intersection and renormalized pressure intersection vary analytically in ρ and η.

Proposition 7.1 *Both $I(\rho,\eta)$ and $J(\rho,\eta)$ vary analytically over* $\mathrm{QC}(\Gamma)\times\mathrm{QC}(\Gamma)$.

Proof By Theorem 4.1, Proposition 3.1 and Theorem 2.5, $P = P(-a\tau_\rho - b\tau_\eta)$ is analytic on

$$R = \{(\rho,\eta,(a,b),t) \in \mathrm{QC}(\Gamma)\times\mathrm{QC}(\Gamma)\times\widehat{D}\}.$$

Since we observed, in the proof of Theorem 6.1, that the restriction of P to $\{\rho\}\times\{\eta\}\times\widehat{D}$ is a submersion for all $\rho,\eta\in\mathrm{QC}(\Gamma)$, P itself is a submersion, and $V = P^{-1}(0)\cap R$ is an analytic submanifold of R of codimension one. Then $-I(\rho,\eta)$ is the slope of the tangent line to $V\cap\{(\rho,\eta)\times\widehat{D}\}$ at the point $\big(\rho,\eta,(h(\rho),0)\big)$, so $I(\rho,\eta)$ is analytic. Theorem 5.1 then implies that $J(\rho,\eta)$ is analytic. □

We obtain the following rigidity theorem as a consequence of Theorem 6.1. The inequality portion of this result was previously established by Schapira and Tapie [40, Corollary 3.17].

Corollary 7.2 *If $\rho,\eta\in\mathrm{QC}(\Gamma)$, then*

$$J(\rho,\eta)\ge 1$$

with equality if and only if ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$.

Proof Recall that the slope $c = c(h(\rho),0)$ of $\mathcal{C}(\rho,\eta)$ at $(h(\rho),0)$ is given by

$$c = -\frac{\int_{\Sigma^+}\tau_\eta\,dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+}\tau_\rho\,dm_{-h(\rho)\tau_\rho}} = -I(\rho,\eta).$$

However, by Theorem 6.1,

$$c \le -\frac{h(\rho)}{h(\eta)}$$

with equality if and only if ρ and η are conjugate in $\mathrm{Isom}(\mathbb{H}^3)$. Our corollary follows immediately. □

8 The pressure form

We may define an analytic section $s\colon \mathrm{QF}(S)\to\mathrm{QC}(\Gamma)$ so that $s([\rho])$ is an element of the conjugacy class of ρ. Choose coprime hyperbolic elements α and β in Γ and let $s(\rho)$ be the unique element of $[\rho]$ such that $s(\rho)(\alpha)$ has attracting fixed point 0 and repelling fixed point ∞ and $s(\rho)(\beta)$ has attracting fixed point 1. This will allow us to abuse notation and regard $\mathrm{QF}(S)$ as a subset of $\mathrm{QC}(\Gamma)$.

Following Bridgeman [9] and McMullen [30], we define an analytic pressure form $\mathbb{P}$ on the tangent bundle $T\mathrm{QF}(S)$ of $\mathrm{QF}(S)$, by letting

$$\mathbb{P}_{T_{[\rho]}\mathrm{QF}(S)} = s^*\big(\mathrm{Hess}\big(J(s(\rho),\cdot)\big)\big|_{T_{s(\rho)}s(\mathrm{QF}(S))}\big),$$

which we rewrite with our abuse of notation as

$$\mathbb{P}_{T_\rho\mathrm{QF}(S)} = \mathrm{Hess}(J(\rho,\cdot)).$$

Corollary 7.2 implies that $\mathbb{P}$ is nonnegative, ie $\mathbb{P}(v,v)\geq 0$ for all $v\in T\mathrm{QF}(S)$.

Since $\mathbb{P}$ is nonnegative, we can define a path pseudometric on $\mathrm{QF}(S)$ by setting

$$d_{\mathbb{P}}(\rho,\eta) = \inf\Big\{\int_0^1 \sqrt{\mathbb{P}(\gamma'(t),\gamma'(t))}\,dt\Big\},$$

where the infimum is taken over all smooth paths in $\mathrm{QF}(S)$ joining ρ to η.

We now derive a standard criterion for when a tangent vector is degenerate with respect to $\mathbb{P}$; see also [11, Corollary 2.5; 10, Lemma 9.3].

Lemma 8.1 *If $v\in T_\rho\mathrm{QF}(S)$, then $\mathbb{P}(v,v)=0$ if and only if*

$$D_v(h\ell_\gamma)=0$$

for all $\gamma\in\Gamma$.

Proof Let $\mathcal{H}_0$ denote the space of pressure-zero, locally Hölder continuous functions on Σ^+. We have a well-defined thermodynamic mapping $\psi\colon \mathrm{QF}(S)\to\mathcal{H}_0$ given by $\psi(\rho)=-h(s(\rho))\tau_{s(\rho)}$. Notice that, by Proposition 3.1 and Theorem 5.1, $\psi(\mathrm{QF}(S))$ is a real analytic family.

Suppose that $\{\rho_t\}_{t\in(-\epsilon,\epsilon)}$ is a one-parameter analytic family in $\mathrm{QF}(S)$ and $v=\dot\rho_0$. Then

$$\frac{d^2}{dt^2}\bigg|_{t=0} J(\rho_0,\rho_t) = \frac{d^2}{dt^2}\left(\frac{\int_{\Sigma^+}\psi(\rho_t)\,dm_{\psi(\rho_0)}}{\int_{\Sigma^+}\psi(\rho_0)\,dm_{\psi(\rho_0)}}\right) = \frac{\int_{\Sigma^+}\ddot\psi_0\,dm_{\psi(\rho_0)}}{\int_{\Sigma^+}\psi(\rho_0)\,dm_{\psi(\rho_0)}},$$

where

$$\ddot\psi_0 = \frac{d^2}{dt^2}\bigg|_{t=0}\psi(\rho_t).$$

Theorem 2.5 implies that

$$0=\frac{d^2}{dt^2}\bigg|_{t=0} P(\psi(t)) = \mathrm{Var}(\dot\psi_0, m_{\psi(0)}) + \int_{\Sigma^+}\ddot\psi_0\,dm_{\psi(\rho_0)},$$

where

$$\dot\psi_0 = \frac{d}{dt}\bigg|_{t=0}\psi(\rho_t),$$

so

$$\frac{d^2}{dt^2}\bigg|_{t=0} J(\rho_0,\rho_t) = -\frac{\mathrm{Var}(\dot\psi_0, m_{\psi(0)})}{\int_{\Sigma^+}\psi(\rho_0)\,dm_{\psi(\rho_0)}}.$$

Recall — see Sarig [39, Theorem 5.12] — that $\mathrm{Var}(\dot\psi_0, m_{\psi(0)}) = 0$ if and only if $\dot\psi_0$ is cohomologous to a constant function C. On the other hand, since $P(\psi_t) = 0$ for all t, the formula for the derivative of the pressure function gives that

$$0 = \frac{d}{dt}\bigg|_{t=0} P(\psi_t) = \int_{\Sigma^+}\dot\psi_0\,dm_{\psi(\rho_0)},$$

so C must equal 0. However, $\dot\psi_0$ is cohomologous to 0 if and only if, for all $x\in\mathrm{Fix}^n$ and all n,

$$0 = S_n\dot\psi_0(x) = \frac{d}{dt}\bigg|_{t=0} S_n\psi_t(x) = \frac{d}{dt}\bigg|_{t=0}(h(\rho_t)\ell_{G(x_1)\cdots G(x_n)}(\rho_t))$$

(see [39, Theorem 1.1]). Moreover, for every hyperbolic element $\gamma\in\Gamma$, there exists $x\in\mathrm{Fix}^n$ (for some n) such that γ is conjugate to $G(x_1)\cdots G(x_n)$, so $\ell_\gamma(\rho_t) = \ell_{G(x_1)\cdots G(x_n)}(\rho_t)$ for all t. If $\gamma\in\Gamma$ is not hyperbolic, then $\ell_\gamma(\rho_t) = 0$ for all t, so

$$\frac{d}{dt}\bigg|_{t=0}(h(\rho_t)\ell_\gamma(\rho_t)) = 0$$

in every case. Therefore, $\dot\psi_0$ is cohomologous to 0 if and only if

$$\frac{d}{dt}\bigg|_{t=0}(h(\rho_t)\ell_\gamma(\rho_t)) = 0$$

for all $\gamma\in\Gamma$. □

9 Main theorem

We recall that a quasifuchsian representation $\rho\colon\Gamma\to\mathsf{PSL}(2,\mathbb{C})$ is said to be *fuchsian* if it is conjugate into $\mathsf{PSL}(2,\mathbb{R})$, ie there exists $A\in\mathsf{PSL}(2,\mathbb{C})$ such that $A\rho(\gamma)A^{-1}\in\mathsf{PSL}(2,\mathbb{R})$ for all $\gamma\in\Gamma$. The fuchsian locus $F(S)\subset\mathrm{QF}(S)$ is the set of (conjugacy classes of) fuchsian representations.

We say that $v\in T_\rho\mathrm{QF}(S)$ is a *pure bending* vector if $v = \partial\rho_t/\partial t$, $\rho=\rho_0$ is fuchsian and ρ_{-t} is the complex conjugate of ρ_t for all t. Since the fuchsian locus $F(S)$ is the fixed-point set of the action of complex conjugation on $\mathrm{QF}(S)$ and the collection of pure bending vectors at a point in $F(S)$ is half-dimensional, one gets a decomposition

$$T_\rho\mathrm{QF}(S) = T_\rho F(S)\oplus B_\rho,$$

where B_ρ is the space of pure bending vectors at ρ. If v is a pure bending vector at $\rho \in F(S)$, then v is tangent to a path obtained by bending ρ by a (signed) angle t along some measured lamination λ (see Bonahon [5, Section 2] for details).

We are finally ready to show that our pressure form is degenerate only along pure bending vectors.

Theorem 9.1 *If S is a compact hyperbolic surface with nonempty boundary, then the pressure form $\mathbb{P}$ defines an $\mathrm{Mod}(S)$–invariant path metric $d_{\mathbb{P}}$ on $\mathrm{QF}(S)$ which is an analytic Riemannian metric except on the fuchsian locus.*

Moreover, if $v \in T_\rho(\mathrm{QF}(S))$, then $\mathbb{P}(v,v) = 0$ if and only if ρ is fuchsian and v is a pure bending vector.

Proof If v is a pure bending vector, then we may write $v = \dot\rho_0$, where ρ_{-t} is the complex conjugate of ρ_t for all t, so $h\ell_\gamma(\rho_t)$ is an even function for all $\gamma \in \Gamma$. Therefore, $D_v h\ell_\gamma = 0$ for all $\gamma \in \Gamma$, so Lemma 8.1 implies that $\mathbb{P}(v,v) = 0$.

Our main work is the following converse:

Proposition 9.2 *Suppose that $v \in T_\rho \mathrm{QF}(S)$. If $\mathbb{P}(v,v) = 0$ and $v \neq 0$, then v is a pure bending vector.*

Recall — see [10, Lemma 13.1] — that, if a Riemannian metric on a manifold M is nondegenerate on the complement of a submanifold N of codimension at least one and the restriction of the Riemannian metric to TN is nondegenerate, then the associated path pseudometric is a metric. We will see in Corollary 10.4 that the pressure metric is mapping class group–invariant. Our theorem then follows from Proposition 9.2 and the fact, established by Kao [23], that $\mathbb{P}$ is nondegenerate on the tangent space to the fuchsian locus. □

Proof of Proposition 9.2 Now suppose that $v \in T_\rho \mathrm{QF}(S)$ and $\mathbb{P}(v,v) = 0$. One first observes, following Bridgeman [9], that, since, by Lemma 8.1, $D_v(h\ell_\gamma) = 0$ for all $\gamma \in \Gamma$,

$$D_v \ell_\gamma = k \ell_\gamma(\rho) \tag{2}$$

for all $\gamma \in \Gamma$, where $k = -D_v h / h(\rho)$.

If $\gamma \in \Gamma$, then one can locally define analytic functions $\mathrm{tr}_\gamma(\rho)$ and $\lambda_\gamma(\rho)$ which are the trace and eigenvalue of largest modulus of (some lift of) $\rho(\gamma)$. Notice that

$\ell_\gamma(\rho) = 2\log|\lambda_\gamma(\rho)|$, so we can express our degeneracy criterion (2) as

$$D_v \log|\lambda_\gamma| = k \log|\lambda_\gamma(\rho)| \tag{3}$$

for all $\gamma \in \Gamma$.

We observe that Lemma 7.4 of Bridgeman [9] goes through nearly immediately in our setting. We state the portion of his lemma we will need and provide a brief sketch of the proof for the reader's convenience.

Lemma 9.3 [9, Lemma 7.4] *If $\mathbb{P}(v,v) = 0$, $v \in T_\rho \mathrm{QF}(S)$, $v \neq 0$ and $\gamma \in \Gamma$, then $\lambda_\gamma(\rho)^2$ and $\mathrm{tr}_\gamma(\rho)^2$ are both real.*

Moreover, if $D_v \mathrm{tr}_\alpha \neq 0$, then $\mathrm{Re}(D_v\lambda_\alpha/\lambda_\alpha(\rho)) = 0$.

Proof Suppose first that $D_v\,\mathrm{tr}_\alpha \neq 0$. Since

$$D_v(\mathrm{tr}_\alpha) = D_v\lambda_\alpha\left(\frac{\lambda_\alpha^2 - 1}{\lambda_\alpha^2}\right),$$

we may conclude that $D_v\lambda_\alpha \neq 0$. Choose $\gamma \in \Gamma$ so that γ is hyperbolic and does not commute with α. Bridgeman then normalizes (the lifts) so that

$$\rho(\alpha) = \begin{bmatrix} \lambda_\alpha & 0 \\ 0 & \lambda_\alpha^{-1} \end{bmatrix} \quad \text{and} \quad \rho(\gamma) = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

where a, b, c and d are all functions defined on a neighborhood of ρ such that a and d are nonzero. He then computes that

$$\log|\lambda_{\alpha^n\gamma}| = n\log|\lambda_\gamma| + \log|a| + \mathrm{Re}\left(\lambda_\alpha^{-2n}\left(\frac{ad-1}{a^2}\right)\right) + O(|\lambda_\alpha^{-4n}|).$$

He differentiates this equation and applies (3) to conclude that

$$\mathrm{Re}\left(\frac{D_v\lambda_\alpha}{\lambda_\alpha(\rho)}\left(\frac{a(\rho)\,d(\rho)-1}{a(\rho)^2}\right)\right) = 0. \tag{4}$$

A final analysis, which breaks down into the consideration of the cases where the argument of $\lambda_\alpha^2(\rho)$ is rational or irrational, yields that $\lambda_\alpha(\rho)^2$ is real. Since $\mathrm{tr}_\alpha^2 = \lambda_\alpha^2 + 2 + \lambda_\alpha^{-2}$, we conclude that $\mathrm{tr}_\alpha^2(\rho)$ is real.

One may further differentiate the equation

$$\mathrm{tr}_{\alpha^n\gamma} = a\lambda_\alpha^n + d\lambda_a^{-n}$$

to conclude that

$$\lim \frac{D_v\,\mathrm{tr}_{\alpha^n\gamma}}{n\lambda_\alpha(\rho)^n} = \frac{a(\rho)D_v\lambda_\alpha}{\lambda_\alpha(\rho)},$$

so $D_v \operatorname{tr}_{\alpha^n\gamma} \neq 0$ is nonzero for all large enough n. Therefore, by the above paragraph,

$$\operatorname{tr}^2_{\alpha^n\gamma}(\rho) = a(\rho)^2\lambda_\alpha(\rho)^{2n} + 2ad(\rho) + d(\rho)^2\lambda_\alpha(\rho)^{-2n}$$

is real for all large enough n. Taking limits allows one to conclude that $a(\rho)^2$, $d(\rho)^2$ and $a(\rho)d(\rho)$ are real. Equation (4) then yields that $\operatorname{Re}(D_v\lambda_\alpha/\lambda_\alpha(\rho)) = 0$. This completes the proof when $D_v \operatorname{tr}_\alpha \neq 0$.

Now suppose that $D_v \operatorname{tr}_\gamma = 0$. If γ is parabolic, $\lambda_\gamma(\rho)^2 = 1$ and $\operatorname{tr}^2_\gamma(\rho) = 4$, which are both real, so we may suppose that γ is hyperbolic. Since there are finitely many elements $\{\alpha_1, \dots, \alpha_n\}$ of Γ such that $\rho \in \mathrm{QF}(S)$ is determined by $\{\operatorname{tr}_{\alpha_1}(\rho)^2, \dots, \operatorname{tr}_{\alpha_n}(\rho)^2\}$ — see [13, Lemma 2.5] — and trace functions are analytic, there exists $\alpha \in \Gamma$ such that $D_v \operatorname{tr}_\alpha \neq 0$. The above analysis then yields that $a(\rho)^2$, $d(\rho)^2$ and $a(\rho)d(\rho)$ are all real. Therefore,

$$\operatorname{tr}_\gamma(\rho)^2 = a(\rho)^2 + 2a(\rho)d(\rho) + d(\rho)^2 = \lambda_\gamma(\rho)^2 + 2 + \lambda_\gamma(\rho)^{-2}$$

is real. So, we may conclude that $\lambda_\gamma(\rho)^2$ is real in this case as well. □

Since $v \neq 0$, there exists $\alpha \in \Gamma$ such that $D_v \operatorname{tr}_\alpha \neq 0$ and

$$\operatorname{Re}\left(\frac{D_v\lambda_\alpha}{\lambda_\alpha(\rho)}\right) = \frac{D_v|\lambda_\alpha|}{|\lambda_\alpha(\rho)|} = D_v \log|\lambda_\alpha|,$$

equation (3) and Lemma 9.3 imply that

$$k = \frac{D_v \log|\lambda_\alpha|}{\log|\lambda_\alpha(\rho)|} = 0.$$

Therefore, $D_v\ell_\gamma = 0$ for all $\gamma \in \Gamma$.

Notice that, since $\operatorname{tr}_\gamma(\rho)^2$ is real for all $\gamma \in \Gamma$, $\rho(\Gamma)$ lies in a proper (real) Zariski closed subset of $\mathsf{PSL}(2,\mathbb{C})$, so is not Zariski dense. However, since the Zariski closure of $\rho(\Gamma)$ is a Lie subgroup, it must be conjugate to a subgroup of either $\mathsf{PSL}(2,\mathbb{R})$ or the index two extension of $\mathsf{PSL}(2,\mathbb{R})$ obtained by appending $z \to -z$. Since ρ is quasifuchsian, its limit set $\Lambda(\rho(\Gamma))$ is a Jordan curve and no element of $\rho(\Gamma)$ can exchange the two components of its complement. Therefore, ρ is fuchsian. (We note that this is the only place where our argument differs significantly from Bridgeman's. It replaces his rather technical [9, Lemma 15].)

We can then write $v = v_1 + v_2$, where $v_1 \in T_\rho F(S)$ and v_2 is a pure bending vector. Since v_2 is a pure bending vector,

$$0 = D_v\ell_\gamma = D_{v_1}\ell_\gamma + D_{v_2}\ell_\gamma = D_{v_1}\ell_\gamma$$

for all $\gamma \in \Gamma$. But, since $v_1 \in T_\rho F(S)$ and there are finitely many curves whose length functions provide analytic parameters for $F(S)$, this implies that $v_1 = 0$. Therefore, $v = v_2$ is a pure bending vector. □

10 Patterson–Sullivan measures

In this section, we observe that the equilibrium state $m_{-h(\rho)\tau_\rho}$ is a normalized pullback of the Patterson–Sullivan measure on $\Lambda(\rho(\Gamma))$. We use this to give a more geometric interpretation of the pressure intersection of two quasifuchsian representations, and hence a geometric formulation of the pressure form.

Sullivan [43; 45] generalized Patterson's construction [31] for fuchsian groups to define a probability measure μ_ρ supported on $\Lambda(\rho(\Gamma))$, called the *Patterson–Sullivan measure*. This measure satisfies the quasi-invariance property

(5) $$d\mu(\rho(\gamma)(z)) = e^{h(\rho)B_z(b_0,\rho(\gamma)^{-1}(b_0))}\, d\mu_\rho(z)$$

for all $z \in \Lambda(\rho(\Gamma))$ and $\gamma \in \Gamma$. Sullivan showed that μ_ρ is a scalar multiple of the $h(\rho)$–dimensional Hausdorff measure on $\partial\mathbb{H}^3$ (with respect to the metric obtained from its identification with $T^1_{b_0}(\mathbb{H}^3)$).

Let $\hat{\mu}_\rho = (\xi_\rho \circ \omega)^*\mu_\rho$ be the pullback of the Patterson–Sullivan measure to Σ^+. Our normalization will involve the Gromov product with respect to b_0, which is defined to be

(6) $$\langle z, w\rangle = \tfrac{1}{2}(B_z(b_0, p) + B_w(b_0, p))$$

for any pair z and w of distinct points in $\partial\mathbb{H}^3$, where p is some (any) point on the geodesic joining z to w. One may check that, for all $\alpha \in \rho(\Gamma)$ and $z, w \in \Lambda(\rho(\Gamma))$,

$$\langle \alpha(z), \alpha(w)\rangle = \langle z, w\rangle - \tfrac{1}{2}\big(B_z(b_0, \alpha^{-1}(b_0)) + B_w(b_0, \alpha^{-1}(b_0))\big).$$

If $x \in \Sigma^+$, let

$$\Lambda(\rho(\Gamma))_x = \{\xi_\rho(\omega(y^-)) \mid y \in \Sigma,\ y^+ = x\},$$

where Σ is the two-sided Markov shift associated to Σ^+ and $y^- = (y^{-1}_{1-i})_{i\in\mathbb{N}}$. Notice that each $\Lambda(\rho(\Gamma))_x$ is open in $\Lambda(\rho(\Gamma))$. Furthermore, there are only finitely many different sets which arise as $\Lambda(\rho(\Gamma))_x$ for some $x \in \Sigma^+$, since $\Lambda(\rho(\Gamma))_x$ depends only on x_1 and, if $r(x_1) \geq 3$ and $x_1 = (b_0, w^s, w_1, \dots, w_{k-1})$, then $\Lambda(\rho(\Gamma))_x$ depends only on b_0 and w. Let $H_\rho\colon \Sigma^+ \to (0,\infty)$ be defined by

$$H_\rho(x) = \int_{\Lambda(\rho(\Gamma))_x} e^{2h(\rho)\langle \xi_\rho(\omega(x)), z\rangle}\, d\mu_\rho(z).$$

Notice that $\Lambda(\rho(\Gamma))_x$ is disjoint from $\xi_\rho(I_x)$, where I_x is the component of $\partial\mathbb{H}^2-\partial D_0$ containing $\omega(x)$, so $e^{2h(\rho)\langle\xi_\rho(\omega(x)),z\rangle_{b_0}}$ is bounded on $\Lambda(\rho(\Gamma))_x$. In particular, $H_\rho(x)$ is finite for all x. Since ω is locally Hölder continuous and ξ_ρ is Hölder, H_ρ is locally Hölder continuous.

We now show that H_ρ is the normalization of the pullback $\hat{\mu}_\rho$ of Patterson–Sullivan measure which gives the equilibrium measure for $-h(\rho)\tau_\rho$. Dal'bo and Peigné [16, Proposition V.3] obtain an analogous result for negatively curved manifolds whose fundamental groups "act like" geometrically finite fuchsian groups of coinfinite area (see also Dal'bo and Peigné [15, Corollary II.5]).

Proposition 10.1 *If S is a compact surface with nonempty boundary and $\rho\in \mathrm{QF}(S)$, then the equilibrium state of $-h(\rho)\tau_\rho$ on Σ^+ is a scalar multiple of $H_\rho\hat{\mu}_\rho$.*

Proof Let $\alpha(\rho,x)=\rho(G(x_1))^{-1}$ and notice that

$$\alpha(\rho,x)\big(\xi_\rho(\omega(x))\big)=\xi_\rho\big(\omega(\sigma(x))\big)\quad\text{and}\quad \alpha(\rho,x)\big(\Lambda(\rho(\Gamma))_x\big)=\Lambda(\rho(\Gamma))_{\sigma(x)}.$$

The quasi-invariance of Patterson–Sullivan measure implies that

$$\frac{d\hat{\mu}(\sigma(y))}{d\hat{\mu}(y)}=\frac{d\mu_\rho\big(\alpha(\rho,x)\big(\xi_\rho(\omega(y))\big)\big)}{d\mu_\rho\big(\xi_\rho(\omega(y))\big)}=e^{h(\rho)B_{(\xi_\rho(\omega)(y))}(b_0,\alpha(\rho,x)^{-1}(b_0))}.$$

We first check that $H_\rho\hat{\mu}_\rho$ is shift-invariant:

$$\begin{aligned}
&H_\rho(\sigma(x))d\hat{\mu}_\rho(\sigma(x))\\
&\quad=\Big(\int_{\Lambda(\rho(\Gamma))_{\sigma(x)}} e^{2h(\rho)\langle\xi_\rho(\omega(\sigma(x))),w\rangle}\,d\mu_\rho(w)\Big)\,d\mu_\rho\big(\xi_\rho\big(\omega(\sigma(x))\big)\big)\\
&\quad=\Big(\int_{\Lambda(\rho(\Gamma))_{\sigma(x)}} e^{2h(\rho)\langle\alpha(\rho,x)(\xi_\rho(\omega(x))),\alpha(\rho,x)(v)\rangle}\,d\mu_\rho(\alpha(\rho,x)(v))\Big)\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\cdot d\mu_\rho\big(\alpha(\rho,x)\big(\xi_\rho(\omega(x))\big)\big)\\
&\quad=\Big(\int_{\Lambda(\rho(\Gamma))_x} e^{2h(\rho)\langle\xi_\rho(\omega(x)),v\rangle}e^{-h(\rho)\left(B_{\xi_\rho(\omega(x))}(b_0,\alpha(\rho,x)^{-1}(b_0))+B_v(b_0,\alpha(\rho,x)^{-1}(b_0))\right)}\\
&\qquad\quad\cdot e^{h(\rho)B_v(b_0,\alpha(\rho,x)^{-1}(b_0))}\,d\mu_\rho(v)\Big)e^{h(\rho)B_{\xi_\rho(\omega(x))}(b_0,\alpha(\rho,x)^{-1}(b_0))}\,d\mu_\rho\big(\xi_\rho(\omega(x))\big)\\
&\quad=\Big(\int_{\Lambda(\rho(\Gamma))_x} e^{2h(\rho)\langle\xi_\rho(\omega(x)),v\rangle}\,d\mu_\rho(v)\Big)\,d\mu_\rho\big(\xi_\rho(\omega(x))\big)\\
&\quad=H_\rho(x)\,d\hat{\mu}_\rho(x).
\end{aligned}$$

So $H_\rho\hat{\mu}_\rho$ is shift-invariant.

Now we check that $\hat{\mu}_\rho$ is a (scalar multiple of a) Gibbs state for $-h(\rho)\tau_\rho$. We recall, from [29, Theorem 2.3.3], that it suffices to check that $\hat{\mu}_\rho$ is an eigenmeasure for the dual of the transfer operator $\mathcal{L}_{-h(\rho)\tau_\rho}$. If $g\colon \Sigma^+ \to \mathbb{R}$ is bounded and continuous, then

$$\begin{aligned}\int_{\Sigma^+} \mathcal{L}_{-h(\rho)\tau_\rho}(g)(x)\, d\hat{\mu}_\rho(x) &= \int_{\Sigma^+}\Big(\sum_{y\in\sigma^{-1}(x)} e^{-h(\rho)\tau_\rho(y)} g(y)\Big)\, d\hat{\mu}_\rho(x) \\ &= \int_{\Sigma^+} e^{-h(\rho)\tau_\rho(y)} g(y)\, d\hat{\mu}_\rho(\sigma(y)) \\ &= \int_{\Sigma^+} g(y)\, d\hat{\mu}_\rho(y).\end{aligned}$$

Therefore, $\hat{\mu}_\rho$ is a (scalar multiple of a) Gibbs state for $-h(\rho)\tau_\rho$.

Finally, we observe that H_ρ is bounded above. If p is a vertex of D_0, then, by construction, there exists a neighborhood U_p of p such that, if $\omega(x) \in U_p$, then there exists $w \in \mathcal{C}^*$ such that $x_1 = (b, \omega^s, w_1, \dots, w_{k-1}, c)$ for some $s \geq 2$. Recall that we require that $b \neq w_{2N}$ and $c \neq w_k$. Observe that w_1 is the face pairing of the edge of D_0 associated to I_x and that w_{2N} is the inverse of the face pairing associated to the other edge E of ∂D_0 which ends at p. So, if I is the interval in $\partial\mathbb{H}^2 - \partial D_0$ bounded by E, then $\Lambda(\rho(\Gamma))_x$ is disjoint from $\xi_\rho(I_x \cup I)$. Therefore, H_ρ is uniformly bounded on $\omega^{-1}(U_p)$ (since $e^{2h(\rho)\langle \xi_\rho(\omega(x)), z\rangle_{b_0}}$ is uniformly bounded for all $z \in \Lambda(\rho(\Gamma))_x \subset \Lambda(\rho(\Gamma)) - \xi_\rho(I \cup I_x)$). However, D_0 has finitely many vertices $\{p_1, \dots, p_n\}$ and H_ρ is clearly bounded above if $\omega(x) \in \partial\mathbb{H}^2 - \bigcup U_{p_i}$ (since again $e^{2h(\rho)\langle \xi_\rho(\omega(x)), z\rangle_{b_0}}$ is uniformly bounded for all $z \in \Lambda(\rho(\Gamma))_x \subset \Lambda(\rho(\Gamma)) - I_x$). Therefore, H_ρ is bounded above on Σ^+.

Since every multiple of a Gibbs state for $-h(\rho)\tau_\rho$ by a continuous function which is bounded between positive constants is also a (scalar multiple of a) Gibbs state for $-h(\rho)\tau_\rho$ (see [29, Remark 2.2.1]), we see that $H_\rho\hat{\mu}_\rho$ is a shift-invariant Gibbs state and hence an equilibrium measure for $-h(\rho)\tau_\rho$ (see Theorem 2.4). □

If $\rho \in \mathrm{QC}(\Gamma)$, let $N_\rho = \mathbb{H}^3/\rho(\Gamma)$ be the quasifuchsian 3–manifold and let $T^1(N_\rho)^{\mathrm{nw}}$ denote the nonwandering portion of its geodesic flow. The Hopf parametrization provides a homeomorphism

$$\mathcal{H}\colon T^1(N_\rho)^{\mathrm{nw}} \to \Omega = \big(\big(\Lambda(\rho(\Gamma)) \times \Lambda(\rho(\Gamma)) - \Delta\big) \times \mathbb{R}\big)/\,\Gamma.$$

Let

$$\Sigma^{\hat{\tau}_\rho} = \{(x,t) : x \in \Sigma,\ 0 \leq t \leqslant \hat{\tau}_\rho(x^+)\}/\sim,$$

where $(x, \tau_\rho(x^+)) \sim (\sigma(x), 0)$, be the suspension flow over Σ with roof function $\hat{\tau}_\rho$. Recall that $\hat{\tau}_\rho\colon \Sigma^+ \to (0,\infty)$ is a positive function cohomologous to τ_ρ.

The Stadlbauer–Ledrappier–Sarig coding map ω for Σ^+ extends to a continuous injective coding map

$$\widehat{\omega}\colon \Sigma \to \Lambda(\Gamma) \times \Lambda(\Gamma)$$

given by $\widehat{\omega}(x) = (\omega(x^+), \omega(x^-))$, where $x^+ = (x_i)_{i\in\mathbb{N}}$ and $x^- = (x_{1-i}^{-1})_{i\in\mathbb{N}}$. One then has a continuous injective map

$$\kappa\colon \Sigma^{\widehat{\tau}_\rho} \to \Omega$$

which is the quotient of the map $\widetilde{\kappa}\colon \Sigma \times \mathbb{R} \to \big(\Lambda(\rho(\Gamma)) \times \Lambda(\rho(\Gamma)) - \Delta\big) \times \mathbb{R}$ given by

$$\widetilde{\kappa}(x,t) = ((\xi_\rho \times \xi_\rho)\widehat{\omega}(x), t).$$

(The image of κ is the complement of all flow lines which do not exit cusps of N_ρ and has full measure in Ω.) The map κ conjugates the suspension flow to the geodesic flow on its image, ie $\kappa \circ \phi_t = \phi_t \circ \kappa$ for all $t \in \mathbb{R}$ on $\kappa(\Sigma^{\widehat{\tau}_\rho})$.

The Bowen–Margulis–Sullivan measure m^ρ_{BM} on Ω can be described by its lift to $\widetilde{\Omega}$, which is given by

$$\widetilde{m}^\rho_{\mathrm{BM}}(z,w,t) = e^{2h(\rho)\langle z,w\rangle_{b_0}}\, d\mu_\rho(z)\, d\mu_\rho(w)\, dt.$$

The Bowen–Margulis–Sullivan measure m^ρ_{BM} is finite and ergodic (see Sullivan [45, Theorem 3]) and equidistributed on closed geodesics (see Roblin [35, théorème 5.1.1] or Paulin, Pollicott and Schapira [32, Theorem 9.11].)

Corollary 10.2 *Suppose that $F\colon (\Sigma^+)^{\widehat{\tau}_\rho} \to \mathbb{R}$ is a bounded continuous function and $\widehat{F}\colon \Sigma^{\widehat{\tau}_\rho} \to \mathbb{R}$ is given by $\widehat{F}(x,t) = F(x^+,t)$. Then*

$$\frac{\int_\Omega \widehat{F} \circ \kappa^{-1}\, dm^\rho_{\mathrm{BM}}}{\int_\Omega dm^\rho_{\mathrm{BM}}} = \frac{\int_{\Sigma^+} \big(\int_0^{\widehat{\tau}_\rho(x^+)} F(x,t)\, dt\big)\, dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+} \tau_\rho(x^+)\, dm_{-h(\rho)\widehat{\tau}_\rho}}.$$

Proof Let

$$\widehat{R} = \{(\widehat{\omega}(x), t) \in \Lambda(\rho(\Gamma)) \times \Lambda(\rho(\Gamma)) \times \mathbb{R} \mid x \in \Sigma,\ t \in [0, \widehat{\tau}_\rho(x^+)]\}$$

be a fundamental domain for the action of Γ on $\big(\Lambda(\rho(\Gamma)) \times \Lambda(\rho(\Gamma)) - \Delta\big) \times \mathbb{R}$ and let

$$R = \{(\omega(x^+), t) \in \Lambda(\rho(\Gamma)) \times \mathbb{R} \mid x^+ \in \Sigma^+,\ t \in [0, \widehat{\tau}_\rho(x^+)]\}.$$

By Proposition 10.1,

$$\begin{aligned}\int_\Omega \widehat{F}\circ\kappa^{-1}\,dm^\rho_{\mathrm{BM}} &= \int_{\widehat{R}} \widehat{F}\circ\kappa^{-1} e^{h(\rho)2\langle z,w\rangle_{b_0}}\,d\mu_\rho(z)\,d\mu_\rho(w)\,dt \\ &= \int_R F(\omega^{-1}(z),t)\Big(\int_{\Lambda(\rho(\Gamma))} e^{h(\rho)2\langle z,w\rangle_{b_0}}\,d\mu_\rho(w)\Big)\,d\mu_\rho(z)\,dt \\ &= \int_R F(\omega^{-1}(z),t)H_\rho(z)\,d\mu_\rho(z)\,dt \\ &= \int_{\Lambda(\rho(\Gamma))}\Big(\int_0^{\hat{\tau}_\rho(\omega^{-1}(z))} F(\omega^{-1}(z),t)\,dt\Big)H_\rho(z)\,d\mu_\rho(z) \\ &= \int_{\Sigma^+}\Big(\int_0^{\hat{\tau}_\rho(x_+)} F(x^+,t)\,dt\Big)\,dm_{-h(\rho)\tau_\rho}(x_+).\end{aligned}$$

In particular, if we consider $F\equiv 1$, then we see that

$$\begin{aligned}\|dm^\rho_{\mathrm{BM}}\| = \int_\Omega dm^\rho_{\mathrm{BM}} &= \int_{\Sigma^+}\Big(\int_0^{\hat{\tau}_\rho(x_+)} dt\Big)\,dm_{-h(\rho)\tau_\rho}(x_+) \\ &= \int_{\Sigma^+}\tau_\rho(x^+)\,dm_{-h(\rho)\tau_\rho},\end{aligned}$$

so our result follows. $\square$

Let

$$\mu_T(\rho) = \frac{1}{|R_T(\rho)|}\sum_{[\gamma]\in R_T(\rho)}\frac{\delta_{[\gamma]}}{\ell_\rho(\gamma)},$$

where $\delta_{[\gamma]}$ is the Dirac measure on the closed orbit associated to $[\gamma]$ and

$$R_T(\rho) = \{[\gamma]\in[\pi_1(S)] \mid 0<\ell_\rho(\gamma)\le T\}.$$

(If $\gamma=\beta^n$ for $n>1$ and β is indivisible, then $\delta_{[\gamma]}/\ell_\rho(\gamma) = n\delta_{[\beta]}/\ell_\rho(\beta^n) = \delta_{[\beta]}/\ell_\rho(\beta)$.) Since the Bowen–Margulis measure m^ρ_{BM} is equidistributed on closed geodesics, $\{\mu_T(\rho)\}$ converges to $m^\rho_{\mathrm{BM}}/\|m^\rho_{\mathrm{BM}}\|$ weakly (in the dual to the space of bounded continuous functions) as $T\to\infty$.

We finally obtain the promised geometric form for the pressure intersection. We may thus think of the pressure intersection, in the spirit of Thurston, as the Hessian of the length of a random geodesic.

Theorem 10.3 *Suppose that S is a compact surface with nonempty boundary, $X=\mathbb{H}^2/\Gamma$ is a finite-area surface homeomorphic to the interior of S, and $\rho\in\mathrm{QF}(S)$. If $\{\gamma_n\}\subset\Gamma$ and $\{\delta_{\rho(\gamma_n)}/\ell_\rho(\gamma_n)\}$ converges weakly to $m^\rho_{\mathrm{BM}}/\|m^\rho_{\mathrm{BM}}\|$, then*

$$I(\rho,\eta) = \lim_{n\to\infty}\frac{\ell_\eta(\gamma_n)}{\ell_\rho(\gamma_n)}.$$

Moreover,

$$I(\rho,\eta)=\lim_{T\to\infty}\frac{1}{|R_T(\rho)|}\sum_{[\gamma]\in R_T(\rho)}\frac{\ell_\eta(\gamma)}{\ell_\rho(\gamma)}.$$

Proof Let $\{\Gamma_n\}$ be a sequence of finite collections of elements of $[\Gamma]$ such that

$$\left\{\mu(\Gamma_n)=\frac{1}{|\Gamma_n|}\sum_{[\gamma]\in\Gamma_n}\frac{\delta_{[\gamma]}}{\ell_\rho(\gamma)}\right\}$$

converges weakly to $m^\rho_{\mathrm{BM}}/\|m^\rho_{\mathrm{BM}}\|$. As in [23, Definition 3.9], consider the bounded continuous function $\psi\colon\Sigma^{\hat\tau_\rho}\to\mathbb{R}$ given by

$$\psi(x,t)\mapsto\frac{\hat\tau_\eta(x)}{\hat\tau_\rho(x)}f\left(\frac{t}{\hat\tau_\rho(x)}\right)\quad\text{for all } t\in[0,\hat\tau_\rho(x)],$$

where $f\colon[0,1]\to\mathbb{R}$ is a smooth function such that $f(0)=f(1)=0$, $f(t)>0$ for $0<t<1$, and $\int_0^1 f(t)\,dt=1$. Then

$$\int_\Omega\hat\psi\circ\kappa^{-1}\,d\mu(\Gamma_n)=\frac{1}{|\Gamma_n|}\sum_{[\gamma]\in\Gamma_n}\frac{\ell_\eta(\gamma)}{\ell_\rho(\gamma)},$$

where $\hat\psi(x,t)=\psi(x^+,t)$ for all $x\in\Sigma$. So, by Corollary 10.2,

$$\left\{\frac{1}{|\Gamma_n|}\sum_{[\gamma]\in\Gamma_n}\frac{\ell_\eta(\gamma_n)}{\ell_\rho(\gamma_n)}\right\}$$

converges to

$$\begin{aligned}\frac{\int_\Omega\hat\psi\circ\kappa^{-1}\,dm^\rho_{\mathrm{BM}}}{\|m^\rho_{\mathrm{BM}}\|}&=\frac{\int_{\Sigma^+}(\hat\tau_\eta(x)/\hat\tau_\rho(x))\big(\int_0^{\hat\tau_\rho(x)}f(t/\hat\tau_\rho(x))\,dt\big)\,dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+}\hat\tau_\rho(x)\,dm_{-h(\rho)\tau_\rho}}\\&=\frac{\int_{\Sigma^+}\hat\tau_\eta\,dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+}\hat\tau_\rho\,dm_{-h(\rho)\tau_\rho}}=\frac{\int_{\Sigma^+}\tau_\eta\,dm_{-h(\rho)\tau_\rho}}{\int_{\Sigma^+}\tau_\rho\,dm_{-h(\rho)\tau_\rho}}.\end{aligned}$$ □

As a consequence, we obtain a geometric presentation of the pressure form which allows us to easily see that the pressure metric is mapping class group–invariant.

Corollary 10.4 *If S is a compact surface with nonempty boundary and $\rho_0\in\mathrm{QF}(S)$, then*

$$\mathbb{P}|_{T_{\rho_0}\mathrm{QF}(S)}=\mathrm{Hess}(J(\rho_0,\rho))=\mathrm{Hess}\left(\frac{h(\rho)}{h(\rho_0)}\lim_{T\to\infty}\frac{1}{|R_T(\rho_0)|}\sum_{[\gamma]\in R_T(\rho_0)}\frac{\ell_\rho(\gamma)}{\ell_{\rho_0}(\gamma)}\right).$$

Moreover, the pressure metric is mapping class group–invariant.

Proof The expression for the pressure form follows immediately from the definition and Theorem 10.3. Now observe that, if $\phi \in \mathrm{Mod}(S)$ and $\rho \in \mathrm{QF}(S)$, then $\phi(\rho) = \rho \circ \phi_*$, so $\ell_\rho(\gamma) = \ell_{\phi(\rho)}(\phi_*(\gamma))$. Therefore, $R_T(\phi(\rho)) = \phi_*(R_T(\rho))$, so $|R_T(\rho)| = |R_T(\phi(\rho))|$ for all T, which implies that $h(\rho) = h(\phi(\rho))$. We can also check that

$$\begin{aligned} I(\rho_0,\rho) &= \lim_{T\to\infty} \frac{1}{|R_T(\rho_0)|} \sum_{[\gamma]\in R_T(\rho_0)} \frac{\ell_\rho(\gamma)}{\ell_{\rho_0}(\gamma)} \\ &= \lim_{T\to\infty} \frac{1}{|R_T(\rho_0)|} \sum_{[\gamma]\in R_T(\rho)} \frac{\ell_{\phi(\rho)}(\phi_*(\gamma))}{\ell_{\phi(\rho_0)}(\phi_*(\gamma))} \\ &= \lim_{T\to\infty} \frac{1}{|R_T(\phi(\rho_0))|} \sum_{[\gamma]\in R_T(\phi(\rho_0))} \frac{\ell_{\phi(\rho)}(\gamma)}{\ell_{\phi(\rho_0)}(\gamma)} \\ &= I(\phi(\rho_0),\phi(\rho)). \end{aligned}$$

Therefore, $J(\rho_0,\rho) = J(\phi(\rho_0),\phi(\rho))$ for all $\phi \in \mathrm{Mod}(S)$ and $\rho_0, \rho \in \mathrm{QF}(S)$, so the renormalized pressure intersection is mapping class group–invariant, so the pressure metric is mapping class group–invariant. □

References

[1] **K Astala**, **T Iwaniec**, **G Martin**, *Elliptic partial differential equations and quasiconformal mappings in the plane*, Princeton Mathematical Series 48, Princeton Univ. Press (2009) MR Zbl

[2] **L Bers**, *Simultaneous uniformization*, Bull. Amer. Math. Soc. 66 (1960) 94–97 MR Zbl

[3] **C Bishop**, **T Steger**, *Representation-theoretic rigidity in* $\mathrm{PSL}(2,\mathbb{R})$, Acta Math. 170 (1993) 121–149 MR Zbl

[4] **F Bonahon**, *The geometry of Teichmüller space via geodesic currents*, Invent. Math. 92 (1988) 139–162 MR Zbl

[5] **F Bonahon**, *Kleinian groups which are almost Fuchsian*, J. Reine Angew. Math. 587 (2005) 1–15 MR Zbl

[6] **R Bowen**, *Hausdorff dimension of quasicircles*, Inst. Hautes Études Sci. Publ. Math. 50 (1979) 11–25 MR

[7] **R Bowen**, **C Series**, *Markov maps associated with Fuchsian groups*, Inst. Hautes Études Sci. Publ. Math. 50 (1979) 153–170 MR Zbl

[8] **H Bray**, **R Canary**, **L-Y Kao**, **G Martone**, *Counting, equidistribution and entropy gaps at infinity with applications to cusped Hitchin representations*, J. Reine Angew. Math. 791 (2022) 1–51 MR Zbl

[9] **M Bridgeman**, *Hausdorff dimension and the Weil–Petersson extension to quasifuchsian space*, Geom. Topol. 14 (2010) 799–831 MR Zbl

[10] **M Bridgeman**, **R Canary**, **F Labourie**, **A Sambarino**, *The pressure metric for Anosov representations*, Geom. Funct. Anal. 25 (2015) 1089–1179 MR Zbl

[11] **M Bridgeman**, **R Canary**, **A Sambarino**, *An introduction to pressure metrics for higher Teichmüller spaces*, Ergodic Theory Dynam. Systems 38 (2018) 2001–2035 MR Zbl

[12] **M Burger**, *Intersection, the Manhattan curve, and Patterson–Sullivan theory in rank 2*, Int. Math. Res. Not. (1993) 217–225 MR Zbl

[13] **R D Canary**, **M Culler**, **S Hersonsky**, **P B Shalen**, *Approximation by maximal cusps in boundaries of deformation spaces of Kleinian groups*, J. Differential Geom. 64 (2003) 57–109 MR Zbl

[14] **F Dal'bo**, **J-P Otal**, **M Peigné**, *Séries de Poincaré des groupes géométriquement finis*, Israel J. Math. 118 (2000) 109–124 MR Zbl

[15] **F Dal'bo**, **M Peigné**, *Groupes du ping-pong et géodésiques fermées en courbure -1*, Ann. Inst. Fourier (Grenoble) 46 (1996) 755–799 MR Zbl

[16] **F Dal'bo**, **M Peigné**, *Some negatively curved manifolds with cusps, mixing and counting*, J. Reine Angew. Math. 497 (1998) 141–169 MR Zbl

[17] **A Douady**, **C J Earle**, *Conformally natural extension of homeomorphisms of the circle*, Acta Math. 157 (1986) 23–48 MR Zbl

[18] **A Fathi**, **L Flaminio**, *Infinitesimal conjugacies and Weil–Petersson metric*, Ann. Inst. Fourier (Grenoble) 43 (1993) 279–299 MR Zbl

[19] **G Iommi**, **T Jordan**, *Phase transitions for suspension flows*, Comm. Math. Phys. 320 (2013) 475–498 MR Zbl

[20] **G Iommi**, **F Riquelme**, **A Velozo**, *Entropy in the cusp and phase transitions for geodesic flows*, Israel J. Math. 225 (2018) 609–659 MR Zbl

[21] **L-Y Kao**, *Pressure type metrics on spaces of metric graphs*, Geom. Dedicata 187 (2017) 151–177 MR Zbl

[22] **L-Y Kao**, *Manhattan curves for hyperbolic surfaces with cusps*, Ergodic Theory Dynam. Systems 40 (2020) 1843–1874 MR Zbl

[23] **L-Y Kao**, *Pressure metrics and Manhattan curves for Teichmüller spaces of punctured surfaces*, Israel J. Math. 240 (2020) 567–602 MR Zbl

[24] **M Kapovich**, *Hyperbolic manifolds and discrete groups*, Progr. Math. 183, Birkhäuser, Boston, MA (2001) MR Zbl

[25] **I Kim**, *Marked length rigidity of rank one symmetric spaces and their product*, Topology 40 (2001) 1295–1323 MR Zbl

[26] **G Knieper**, *Volume growth, entropy and the geodesic stretch*, Math. Res. Lett. 2 (1995) 39–58 MR Zbl

[27] **F Ledrappier**, **O Sarig**, *Fluctuations of ergodic sums for horocycle flows on $\mathbb{Z}^d$–covers of finite volume surfaces*, Discrete Contin. Dyn. Syst. 22 (2008) 247–325 MR Zbl

[28] **B Maskit**, *On boundaries of Teichmüller spaces and on Kleinian groups, II*, Ann. of Math. 91 (1970) 607–639 MR Zbl

[29] **R D Mauldin**, **M Urbański**, *Graph directed Markov systems: geometry and dynamics of limit sets*, Cambridge Tracts in Mathematics 148, Cambridge Univ. Press (2003) MR Zbl

[30] **C T McMullen**, *Thermodynamics, dimension and the Weil–Petersson metric*, Invent. Math. 173 (2008) 365–425 MR Zbl

[31] **S J Patterson**, *The limit set of a Fuchsian group*, Acta Math. 136 (1976) 241–273 MR Zbl

[32] **F Paulin**, **M Pollicott**, **B Schapira**, *Equilibrium states in negative curvature*, Astérisque 373, Soc. Math. France, Paris (2015) MR Zbl

[33] **M Pollicott**, **R Sharp**, *A Weil–Petersson type metric on spaces of metric graphs*, Geom. Dedicata 172 (2014) 229–244 MR Zbl

[34] **F Riquelme**, **A Velozo**, *Escape of mass and entropy for geodesic flows*, Ergodic Theory Dynam. Systems 39 (2019) 446–473 MR Zbl

[35] **T Roblin**, *Ergodicité et équidistribution en courbure négative*, Mém. Soc. Math. Fr. 95, Soc. Math. France, Paris (2003) MR Zbl

[36] **D Ruelle**, *Repellers for real analytic maps*, Ergodic Theory Dynam. Systems 2 (1982) 99–107 MR Zbl

[37] **O M Sarig**, *Thermodynamic formalism for countable Markov shifts*, Ergodic Theory Dynam. Systems 19 (1999) 1565–1593 MR Zbl

[38] **O Sarig**, *Existence of Gibbs measures for countable Markov shifts*, Proc. Amer. Math. Soc. 131 (2003) 1751–1758 MR Zbl

[39] **O Sarig**, *Thermodynamic formalism for topological Markov shifts*, lecture notes, Pennsylvnia State University (2009) Available at `http://www.weizmann.ac.il/math/sarigo/lecture-notes`

[40] **B Schapira**, **S Tapie**, *Regularity of entropy, geodesic currents and entropy at infinity*, Ann. Sci. Éc. Norm. Supér. 54 (2021) 1–68 MR Zbl

[41] **R Sharp**, *The Manhattan curve and the correlation of length spectra on hyperbolic surfaces*, Math. Z. 228 (1998) 745–750 MR Zbl

[42] **M Stadlbauer**, *The return sequence of the Bowen–Series map for punctured surfaces*, Fund. Math. 182 (2004) 221–240 MR Zbl

[43] **D Sullivan**, *The density at infinity of a discrete group of hyperbolic motions*, Inst. Hautes Études Sci. Publ. Math. 50 (1979) 171–202 MR Zbl

[44] **D Sullivan**, *Discrete conformal groups and measurable dynamics*, Bull. Amer. Math. Soc. 6 (1982) 57–73 MR Zbl

[45] **D Sullivan**, *Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups*, Acta Math. 153 (1984) 259–277 MR Zbl

[46] **D Sullivan**, *Quasiconformal homeomorphisms and dynamics, II: Structural stability implies hyperbolicity for Kleinian groups*, Acta Math. 155 (1985) 243–260 MR Zbl

[47] **D Sullivan**, *Related aspects of positivity in Riemannian geometry*, J. Differential Geom. 25 (1987) 327–351 MR Zbl

[48] **S A Wolpert**, *Thurston's Riemannian metric for Teichmüller space*, J. Differential Geom. 23 (1986) 143–174 MR Zbl

[49] **X Xie**, *A Bowen type rigidity theorem for non-cocompact hyperbolic groups*, Math. Z. 259 (2008) 249–253 MR Zbl

Department of Mathematics, George Mason University
Fairfax, VA, United States

Department of Mathematics, University of Michigan
Ann Arbor, MI, United States

Department of Mathematics, George Washington University
Washington, DC, United States

hbray@gmu.edu, canary@umich.edu, lkao@gwu.edu

Received: 6 August 2021 Revised: 13 December 2021

Algebraic & Geometric Topology **23**:8 (2023) 3655–3706
DOI: 10.2140/agt.2023.23.3655
Published: 5 November 2023

The $\mathrm{Sp}_{k,n}$–local stable homotopy category

DREW HEARD

We study the category of $(K(k) \vee K(k+1) \vee \cdots \vee K(n))$–local spectra, following a suggestion of Hovey and Strickland. When $k = 0$, this is equivalent to the category of $E(n)$–local spectra, while for $k = n$, this is the category of $K(n)$–local spectra, both of which have been studied in detail by Hovey and Strickland. Based on their ideas, we classify the localizing and colocalizing subcategories, and give characterizations of compact and dualizable objects. We construct an Adams-type spectral sequence and show that when $p \gg n$ it collapses with a horizontal vanishing line above filtration degree $n^2 + n - k$ at the E_2–page for the sphere spectrum. We then study the Picard group of $(K(k) \vee K(k+1) \vee \cdots \vee K(n))$–local spectra, showing that this group is algebraic, in a suitable sense, when $p \gg n$. We also consider a version of Gross–Hopkins duality in this category. A key concept throughout is the use of descent.

55P42, 55P60; 55T15

1 Introduction

In their memoir [32] Hovey and Strickland studied the categories of $K(n)$–local and E_n–local spectra in great detail. Here $K(n)$ is the n^{th} Morava K–theory; the spectrum whose homotopy groups are the graded field

$$K(n)_* \cong \mathbb{F}_p[v_n^{\pm 1}], \quad |v_n| = 2(p^n - 1)$$

and E_n is the n^{th} Lubin–Tate spectrum, or Morava E–theory, with

$$(E_n)_* \cong W(\mathbb{F}_{p^n})[\![u_1, \ldots, u_{n-1}]\!][u^{\pm 1}], \quad |u_i| = 0, |u| = 2.$$

As explained in the introduction of [32], the Morava K–theories are the prime field objects in the stable homotopy category — for a way to make that precise, see Hopkins and Smith [27], or more specifically, Balmer [4, Corollary 9.5] — and are one of the fundamental objects in the chromatic approach to stable homotopy theory.

A deep result of Hopkins and Ravenel [48] is that Bousfield localization with respect to E_n is smashing, which simplifies the study of the category of E_n–local spectra considerably. On the other hand, localization with respect to $K(n)$ is not smashing [32, Lemma 8.1], and the monoidal unit $L_{K(n)}S^0 \in \mathrm{Sp}_{K(n)}$ is dualizable, but not compact. In the language of tensor-triangulated geometry, $\mathrm{Sp}_{K(n)}$ is a nonrigidly compactly generated category. Because of this, much of the work in [32] is therefore dedicated to understanding the more complicated category of $K(n)$–local spectra.

By a Bousfield class argument, the category of E_n–local spectra is equivalent to the category of $(K(0) \vee \cdots \vee K(n))$–local spectra. In this paper we study the categories of $(K(k) \vee \cdots \vee K(n))$–local spectra for $0 \le k \le n$, which were suggested as "interesting to investigate" by Hovey and Strickland; see the remark after Corollary B.9 in [32]. We write $L_{k,n}$ for the associated Bousfield localization functor. As we shall see, when $k \neq 0$, the category $\mathrm{Sp}_{k,n}$ of $(K(k) \vee \cdots \vee K(n))$–local spectra behaves much like the category $\mathrm{Sp}_{K(n)} = \mathrm{Sp}_{n,n}$ of $K(n)$–local spectra. For example, it is an example of a nonrigidly compactly generated category; as soon as $k \neq 0$, the monoidal unit $L_{k,n}S^0 \in \mathrm{Sp}_{k,n}$ is dualizable, but not compact. However, the categories $\mathrm{Sp}_{k,n}$ for $k \neq n$ are in some sense more complicated than the case $k = n$; for example, $\mathrm{Sp}_{K(n)}$ has no nontrivial (co)localizing subcategories, while this is not true for $\mathrm{Sp}_{k,n}$ as long as $k \neq n$.

1A Contents of the paper

We now describe the contents of the paper in more detail. We begin with a study of Bousfield classes, constructing some other spectra which are Bousfield equivalent to $K(k) \vee \cdots \vee K(n)$. In particular, we show that there is a Bousfield equivalence between a localized quotient of BP, denoted by $E(n, J_k)$, and $K(k) \vee \cdots \vee K(n)$. For this reason, as well as brevity, we often say that X is $E(n, J_k)$–local, instead of $(K(k) \vee \cdots \vee K(n))$–local.

As was already noted by Hovey and Strickland [32, Corollary B.9], $\mathrm{Sp}_{k,n}$ is an algebraic stable homotopy theory in the sense of Hovey, Palmieri and Strickland [30] with compact generator $L_{k,n}F(k)$, the localization of a finite spectrum of type k. We investigate some consequences of this; for example, analogous to Hovey and Strickland's formulas for $L_{K(n)}X$, in Proposition 2.24, we prove some formulas for $L_{k,n}X$ in terms of towers of finite type k Moore spectra. Some of these results had previously been obtained by the author and Barthel and Valenzuela [11].

In Section 3 we investigate the tensor-triangulated geometry of $\mathrm{Sp}_{k,n}$. We begin by characterizing the compact objects in $\mathrm{Sp}_{k,n}$, culminating in Theorem 3.8 which is a natural extension of Hovey and Strickland's results in the cases $k = 0, n$. A classification of the thick ideals of $\mathrm{Sp}_{k,n}^{\omega}$ is an almost immediate consequence of this classification; see Theorem 3.16 for the precise result. Of course, here we rely on the deep thick subcategory theorem in stable homotopy [27] and its consequences. Finally, we classify the localizing and colocalizing subcategories of $\mathrm{Sp}_{k,n}$ in Theorem 3.33. We obtain the following.

Theorem 1.1 *There is an order-preserving bijection between (co)localizing subcategories of $\mathrm{Sp}_{k,n}$ and subsets of $\{k, \dots, n\}$. Moreover, the map that sends a localizing subcategory $\mathcal{C}$ of $\mathrm{Sp}_{k,n}$ to its left orthogonal $\mathcal{C}^{\perp}$ induces a bijection between the set of localizing and colocalizing subcategories of $\mathrm{Sp}_{k,n}$. The inverse map sends a colocalizing subcategory $\mathcal{U}$ to its right orthogonal ${}^{\perp}\mathcal{U}$.*

We also compute the Bousfield lattice of $\mathrm{Sp}_{k,n}$ (Proposition 3.39) and show that a form of the telescope conjecture holds (Theorem 3.46).

In Section 4 we show that, as a consequence of the Hopkins–Ravenel smash product theorem, the commutative algebra object $E_n \in \mathrm{Sp}_{k,n}$ is descendable, in the sense of Mathew [41]. This has a number of immediate consequences. For example, it implies the existence of a strongly convergent Adams-type spectral sequence, which we call the $E(n, J_k)$–local E_n–Adams spectral sequence, computing $\pi_*(L_{k,n}X)$ for any spectrum X. Moreover, descendability implies this collapses with a horizontal vanishing line at a finite stage (independent of X). In the case of $X = S^0$ it is known that, in the cases $k = 0$ and $k = n$, this vanishing line already occurs on the E_2–page so long as $p \gg n$. In order to generalize this result, we first show that when $X = S^0$, the E_2–term of the $E(n, J_k)$–local E_n–Adams spectral sequence spectral sequence can be given as the inverse limit of certain Ext groups computed in the category of $(E_n)_*E_n$–comodules; see Proposition 4.16 for the precise result. We are then able to utilize a chromatic spectral sequence and Morava's change of rings theorem to show the following (Theorem 4.24):

Theorem 1.2 *Suppose $p-1$ does not divide $k+s$ for $0 \le s \le n-k$ (for example, if $p > n+1$), then in the E_2–term of the $E(n, J_k)$–local E_n–Adams spectral sequence converging to $L_{k,n}S^0$, we have $E_2^{s,t} = 0$ for $s > n^2+n-k$.*

In the case $k = 0$, this recovers a result of Hovey and Sadofsky [31, Theorem 5.1].

As noted previously, so long as $k \neq 0$, the categories of dualizable and compact spectra do not coincide in $\mathrm{Sp}_{k,n}$; every compact spectrum is dualizable, but the converse does not hold, with the unit $L_{k,n}S^0$ being an example. In Section 5 we study the category of dualizable objects in $\mathrm{Sp}_{k,n}$. As a consequence of descendability, we show that $X \in \mathrm{Sp}_{k,n}$ is dualizable if and only if $L_{k,n}(E_n \wedge X)$ is dualizable in the category of $E(n, J_k)$–local E_n–modules. In turn, we show that this holds if and only if $L_{k,n}(E_n \wedge X)$ is dualizable (equivalently, compact) in the category of E_n–modules. We deduce that $X \in \mathrm{Sp}_{k,n}$ is dualizable if and only if its *Morava module* $(E_{k,n})^\vee_*(X) := \pi_*L_{k,n}(E_n \wedge X)$ is finitely generated as an $(E_n)_*$–module; see Theorem 5.11. This generalizes a result of Hovey and Strickland, but even in this case our proof differs from theirs.

It is an observation of Hopkins that the Picard group of invertible $K(n)$–local spectra is an interesting object to study; see Hopkins, Mahowald and Sadofsky [26]. Likewise, Hovey and Sadofsky [31] have studied the Picard group of $E(n)$–local spectra. In Section 6 we study the Picard group $\mathrm{Pic}_{k,n}$ of $E(n, J_k)$–local spectra. Our first result, which is a consequence of descent, is that $X \in \mathrm{Sp}_{k,n}$ is invertible if and only if its Morava module $(E_{k,n})^\vee_*(X)$ is free of rank 1. We then study the Picard spectrum — see Mathew and Stojanoska [44] — of the category $\mathrm{Sp}_{k,n}$. Using descent again, we construct a spectral sequence whose abutment for π_0 is exactly $\mathrm{Pic}_{k,n}$. The existence of this spectral sequence in the case $k = n$ is folklore. We say that this spectral sequence is algebraic if the only nonzero terms in the spectral sequence occur in filtration degree 0 and 1. Using Theorem 1.2 we deduce the following result (Theorem 6.8). In the case $k=n$, this is a theorem of Pstrągowski [45].

Theorem 1.3 *If $2p-2 \geq n^2+n-k$ and $p-1$ does not divide $k+s$ for $0 \leq s \leq n-k$, then $\mathrm{Pic}_{k,n}$ is algebraic. For example, this holds if $2p-2 > n^2+n$.*

There is an interesting element in the $K(n)$–local Picard group, namely the Brown–Comenetz dual of the monochromatic sphere [32, Theorem 10.2]. In Section 7 we extend Brown–Comenetz duality to the $E(n, J_k)$–local category. We do not know when the Brown–Comenetz dual of the monochromatic sphere defines an element of $\mathrm{Pic}_{k,n}$; this is not true when $k=0$, and we provide a series of equivalent conditions for the general case in Proposition 7.10.

The case $n=2$ and $k=1$

The first example that has essentially not been studied in the literature is when $n=2$ and $k=1$, ie the category of $(K(1) \vee K(2))$–local spectra. In Section 5B we give a computation of the Balmer spectrum of $(K(1) \vee K(2))$–locally dualizable spectra. For this, we recall that Hovey and Strickland have conjectured a description of the Balmer spectrum $\mathrm{Spc}(\mathrm{Sp}_{K(n)}^{\mathrm{dual}})$ of dualizable objects in $K(n)$–local spectra [32, page 61]. This was investigated by the author, along with Barthel and Naumann, in [10]. This admits a natural generalization to $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$. For $i \leq n$, let $\mathcal{D}_i$ denote the category of $X \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ such that X is a retract of $Y \wedge Z$ for some $Y \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ and some finite spectrum Z of type at least i. We also set $\mathcal{D}_{n+1}=(0)$. The conjecture is that these exhaust all the thick tensor-ideals of $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$. We show in Theorem 5.21 that if this holds $K(n)$–locally (ie in $\mathrm{Sp}_{n,n}^{\mathrm{dual}}$), then it holds for all $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$. In particular, since it is known to hold $K(2)$–locally by [10, Theorem 4.15], we obtain the following; see Corollary 5.22.

Theorem 1.4 *The Balmer spectrum of $K(1) \vee K(2)$–locally dualizable spectra*

$$\mathrm{Spc}(\mathrm{Sp}_{1,2}^{\mathrm{dual}}) = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$$

with topology determined by $\overline{\{\mathcal{D}_j\}} = \{\mathcal{D}_i \mid i \geq j\}$. In particular, if $\mathcal{C}$ is a thick tensor-ideal of $\mathrm{Sp}_{1,2}^{\mathrm{dual}}$, then $\mathcal{C}=\mathcal{D}_k$ for $0 \leq k \leq 3$.

Conventions and notation

We let $\langle X \rangle$ denote the Bousfield class of a spectrum X. The smallest thick tensor-ideal containing an object A will be denoted by $\mathrm{thick}_{\otimes}\langle A \rangle$ (it will always be clear in which category this thick subcategory should be taken in). Likewise, the smallest thick (resp. localizing) subcategory containing an object A will be written as $\mathrm{Thick}(A)$ (resp. $\mathrm{Loc}(A)$).

Acknowledgements

It goes without saying that this paper owes a tremendous intellectual debt to Hovey and Strickland, in particular for the wonderful manuscript [32]. We also thank Neil Strickland for a helpful conversation, as well as his comments on a draft version of this document.

2 The category of $\mathrm{Sp}_{k,n}$–local spectra

2A Chromatic spectra

We begin by introducing some of the main spectra that we will be interested in.

Definition 2.1 Let BP denote the Brown–Peterson homotopy ring spectrum with coefficient ring

$$BP_* \cong \mathbb{Z}_{(p)}[v_1, v_2, \dots]$$

with $|v_i| = 2(p^i - 1)$.

Remark 2.2 The classes v_i are not intrinsically defined, and so the definition of BP depends on a choice of sequence of generators; for example, they could be the Hazewinkel generators or the Araki generators. However, the ideals $I_n = (p, v_1, \dots, v_{n-1})$ for $0 \le n \le \infty$ do not depend on this choice.

By taking quotients and localizations of BP — for example, using the theory of structured ring spectra [19, Chapter V] — we can form new homotopy ring spectra. In particular, let J_k denote a fixed invariant regular sequence $p^{i_0}, v_1^{i_1}, \dots, v_{k-1}^{i_{k-1}}$ of length k. Then we can form the homotopy associative ring spectrum BPJ_k with

$$(BPJ_k)_* \cong BP_*/J_k.$$

These were first studied by Johnson and Yosimura [35]. A detailed study on the product structure one obtains via this method can be found in [51].

Definition 2.3 We let $E(n, J_k)$ for $n \ge k$ denote the Landweber exact spectrum with

$$E(n, J_k)_* \cong v_n^{-1}((BPJ_k)_*/(v_{n+1}, v_{n+2}, \dots)).$$

Here Landweber exact means over BPJ_k (as studied by Yosimura [56]), that is, there is an isomorphism

$$E(n,J_k)_*(X) \cong (BPJ_k)_*(X) \otimes_{BP_*/J_k} E(n,J_k)_*.$$

Example 2.4 If $k = 0$ (so that J_k is the trivial sequence), then $E(n,J_0) \simeq E(n)$, Johnson–Wilson theory. For the other extreme, if $J_n = p, v_1, \dots, v_{n-1}$, then BPJ_n is the spectrum known as $P(n)$, and $E(n,J_n) \simeq K(n)$ is Morava K–theory [32].

Definition 2.5 For $k \le n < \infty$, we let $\mathrm{Sp}_{k,n} \subseteq \mathrm{Sp}$ denote the full subcategory of $(K(k) \vee K(k+1) \vee \cdots \vee K(n))$–local spectra.

Lemma 2.6 *The inclusion* $\mathrm{Sp}_{k,n} \hookrightarrow \mathrm{Sp}$ *has a left adjoint* $L_{k,n}$, *and* $\mathrm{Sp}_{k,n}$ *is a presentable, stable* ∞*–category.*

Proof This is a consequence of [39, Proposition 5.5.4.15]. □

Remark 2.7 The category $\mathrm{Sp}_{k,n}$ and localization functor $L_{k,n}$ only depend on the Bousfield class $\langle K(k) \vee \cdots \vee K(n) \rangle$.

Notation 2.8 We will follow standard conventions and write $\mathrm{Sp}_{0,n}$ as Sp_n and $\mathrm{Sp}_{n,n}$ as $\mathrm{Sp}_{K(n)}$. Similarly, the corresponding Bousfield localization functors will be denoted by $L_{0,n} = L_n$ and $L_{n,n} = L_{K(n)}$, respectively.

Remark 2.9 By [46, Theorem 2.1] we have $\langle E(n) \rangle = \langle K(0) \vee \cdots \vee K(n) \rangle$. In fact, let E be a BP–module spectrum that is Landweber exact over BP, and is v_n–periodic, in the sense that $v_n \in BP_*$ maps to a unit in $E_*/(p, v_1, \dots, v_{n-1})$. Then Hovey has shown that $\langle E \rangle = \langle K(0) \vee \cdots \vee K(n) \rangle$ [28, Corollary 1.12]. In particular, this applies to the Lubin–Tate E–theory spectrum E_n — see [49] — with

$$(E_n)_* \cong W(\mathbb{F}_{p^n})[\![u_1, \dots, u_{n-1}]\!][u^{\pm 1}]$$

or the completed version of E–theory used in [32] with

$$E_* \cong (E(n)_*)^\wedge_{I_n} \cong \mathbb{Z}_p[v_1, \dots, v_{n-1}, v_n^{\pm 1}]^\wedge_{I_n}.$$

2B Bousfield decomposition

In the previous section we introduced the spectra $E(n,J_k)$ for $n \ge k$ and an invariant regular sequence $p^{i_0}, \dots, v_{k-1}^{i_{k-1}}$ of length k. We now give Bousfield decompositions for $E(n,J_k)$–local spectra.

Proposition 2.10 *There are equivalences of Bousfield classes*:

(1) (*Johnson–Yosimura*) $\langle v_n^{-1} BPJ_k \rangle = \langle E(n, J_k) \rangle$.

(2) (*Yosimura*) $\langle E(n, J_k) \rangle = \bigvee_{i=k}^{n} \langle K(i) \rangle$.

(3) $\langle E_n / I_k \rangle = \bigvee_{i=k}^{n} \langle K(i) \rangle$.

Proof Part (1) is [35, Corollary 4.11]. Part (2) can be deduced from [57] as we now explain. First, by [57, Corollary 1.3 and Proposition 1.4] along with (1),

$$\langle E(n, J_k) \rangle = \langle v_n^{-1} BPJ_k \rangle = \langle L_n BPJ_k \rangle = \bigvee_{i=k}^{n} \langle v_i^{-1} P(i) \rangle.$$

By [57, Corollary 1.8] we have $\langle v_i^{-1} P(i) \rangle = \langle K(i) \rangle$, and hence (2) follows. For (3), we first note that by the thick subcategory theorem $\langle E_n / I_k \rangle = \langle E_n \wedge F(k) \rangle$ for some finite type k spectrum. Since $E_n = \bigvee_{i=0}^{n} K(i)$, (3) then follows from the definition of a type k spectrum. $\square$

Remark 2.11 In other words, the category of $E(n, J_k)$–local spectra is equivalent to the category of $K(k) \vee \cdots K(n)$–local spectra. Note that this implies this category only depends on the length of the sequence, and not the integers $i_0, \dots, i_{n-1}$. We will therefore sometimes say that a spectrum X is $E(n, J_k)$–local if $X \in \mathrm{Sp}_{k,n}$.

2C Algebraic stable homotopy categories

We now begin by recalling the basics on algebraic stable homotopy theories; see [30] in the triangulated setting.

Definition 2.12 A stable homotopy theory is a presentable, symmetric monoidal stable ∞–category $(\mathcal{C}, \otimes, \mathbf{1})$ where the tensor product commutes with all colimits. It is algebraic if there is a set $\mathcal{G}$ of compact objects such that the smallest localizing subcategory of $\mathcal{C}$ containing all $G \in \mathcal{G}$ is $\mathcal{C}$ itself.

Remark 2.13 The assumptions on $\mathcal{C}$ imply that it the functor $- \wedge Y$ has a right adjoint $F(Y, -)$, ie the symmetric monoidal structure on $\mathcal{C}$ is closed.

Remark 2.14 The associated homotopy category $\mathrm{Ho}(\mathcal{C})$ is then an algebraic stable homotopy theory in the sense of [30]. We note that compactness can be checked at the level of the homotopy category; see [40, Remark 1.4.4.3].

Applying [32, Corollary B.9; 30, Theorem 3.5.1], we have the following.

Proposition 2.15 (Hovey, Palmieri and Strickland) *$\mathrm{Sp}_{k,n}$ is an algebraic stable homotopy category with compact generator $L_{k,n}F(k)$. The symmetric monoidal structure in $\mathrm{Sp}_{k,n}$ is given by*

$$X \mathbin{\overline{\wedge}} Y := L_{k,n}(X \wedge Y).$$

Colimits are computed by taking the colimit in spectra and then applying $L_{k,n}$, while function objects and limits are computed in the category of spectra.

Remark 2.16 The most difficult part of the above proposition is that $L_{k,n}F(k)$ is a compact generator of $\mathrm{Sp}_{k,n}$. Indeed, one must show that the conditions of [32, Proposition B.7] are satisfied and to do this, one at some point needs to invoke the thick subcategory theorem [27], or one its consequences (such as the Hopkins–Ravenel smash product theorem [48]).

Remark 2.17 The localization $L_n = L_{0,n}$ is smashing (that is $L_n X \simeq L_n S^0 \wedge X$) by the Hopkins–Ravenel smash product theorem [48] and in this case $X \mathbin{\overline{\wedge}} Y \simeq X \wedge Y$. However, if $k \neq 0$, then localization $L_{k,n}$ is not smashing as the following lemma shows, and so $X \mathbin{\overline{\wedge}} Y \not\simeq X \wedge Y$ in general.

Lemma 2.18 *If $k \neq 0$, then $L_{k,n}$ is not smashing, and $L_{k,n}S^0$ is not compact in $\mathrm{Sp}_{k,n}$.*

Proof We first claim that $\langle L_{k,n}S^0\rangle = \langle E(n)\rangle$. To see this, note that we have ring maps $L_nS^0 \to L_{k,n}S^0 \to L_{K(n)}S^0$, so $\langle L_{K(n)}S^0\rangle \le \langle L_{k,n}S^0\rangle \le \langle L_nS^0\rangle$. However, $\langle L_{K(n)}S^0\rangle = \langle L_nS^0\rangle = \langle E(n)\rangle$ [32, Corollary 5.3], so these inequalities are actually equalities, and all three are Bousfield equivalent to $E(n)$.

Suppose now that $L_{k,n}$ were smashing, so that $\langle L_{k,n}S^0\rangle = \bigvee_{i=k}^n \langle K(i)\rangle$; see [46, Proposition 1.27]. Then, since $\langle E(n)\rangle \gneq \bigvee_{i=k}^n \langle K(i)\rangle$ as soon as $k \neq 0$, we have obtained a contradiction.

The second part is then a consequence of [30, Theorem 3.5.2]. $\square$

Remark 2.19 Using the periodicity theorem of Hopkins and Smith [27], Hovey and Strickland [32, Section 4] constructed a sequence of ideals $\{I_j\}_j \subseteq \mathfrak{m} \subseteq E_0$ and type k spectra $\{M_k(j)\}_j$ with the following properties (see also [7, Remark 2.1]):

(1) $I_{j+1} \subseteq I_j$ and $\cap_j I_j = 0$;

(2) E_0/I_j is finite; and

(3) $E_*(M_k(j)) \cong E_*/I_j$ and there are spectrum maps $q\colon M_k(j+1) \to M_k(j)$ realizing the quotient $E_*/M_k(j+1) \to E_*/M_k(j)$.

We call such a tower $\{M_k(j)\}_j$ a tower of generalized Moore spectrum of type k.

Remark 2.20 The tower as above is constructed in the homotopy category of spectra. However, as explained in [27, page 9, equation (15)], such sequential diagrams can always be lifted to a sequence of cofibrations between cofibrant objects, and in particular to a diagram in the ∞–category of spectra (the point is that such diagrams have no nontrivial homotopy coherence data). Then, the (co)limit in the ∞–categorical sense, agrees with the homotopy (co)limit used in [30, Definitions 2.2.3 and 2.2.10].

Notation 2.21 We write $M_{k,n}X$ for the fiber of the localization map $L_nX \to L_{k-1}X$. By definition, we set $M_{0,n} = L_n$.

Lemma 2.22 *We have an equality of Bousfield classes*

$$\langle M_{k,n}S^0\rangle = \bigvee_{i=k}^{n} \langle K(i)\rangle.$$

Proof Recall that, by definition, there is a cofiber sequence

$$C_{k-1}S^0 \to S^0 \to L_{k-1}S^0.$$

Applying L_n to this and using $L_iL_j \simeq L_{\min(i,j)}$ we see that

$$M_{k,n}S^0 \simeq L_nC_{k-1}S^0 \simeq C_{k-1}L_nS^0,$$

where the last equivalence follows as both functors are smashing. It follows from [32, Proposition 5.3] that $\langle M_{k,n}S^0\rangle = \bigvee_{i=k}^{n}\langle K(i)\rangle$ as claimed. □

Remark 2.23 In [32, Proposition 7.10(e)] Hovey and Strickland give a formula for $L_{K(n)}X$ in terms of towers of generalized Moore spectra. We show now that their proof extends to $L_{k,n}X$.

Proposition 2.24 *There are equivalences*

$$L_{k,n}X \simeq L_{F(k)}L_nX \simeq \varprojlim_j (L_nX \wedge M_k(j)) \simeq F(M_{k,n}S^0, L_nX),$$

where the limit is taken over a tower $\{M_k(j)\}$ of generalized Moore spectra of type k.

Proof We first note that $L_{k-1}X \simeq L_nL_{k-1}X \simeq L_nL_{k-1}^fX$, the latter by Corollary 6.10 of [32]. It follows that $M_{k,n}X \simeq L_nC_{k-1}^fX$, where C_{k-1}^f is the acyclization functor associated to L_{k-1}^f. By [32, Proposition 7.10(a)] (and Remark 2.20) $C_{k-1}^f X \simeq \varinjlim_j D(M_k(j)) \wedge X$, so $M_{k,n}X \simeq \varinjlim_j D(M_k(j)) \wedge L_n X$. It follows that

$$\varprojlim_j (L_n X \wedge M_k(j)) \simeq F(M_{k,n}S^0, L_n X).$$

Moreover, by [32, Proposition 7.10(a)] this is equivalent to $L_{F(k)} L_n X$.

To finish the proof, we will show that $L_{k,n}X \simeq L_{F(k)} L_n X$. First, note that $X \to L_n X$ is an $L_n S^0$–equivalence, and $L_n X \to L_{F(k)} L_n X$ is an $F(k)$–equivalence, so $X \to L_{F(k)} L_n X$ is an $L_n S^0 \wedge F(k)$–equivalence. But $L_n S^0 \wedge F(k) \simeq L_n F(k)$ and $\langle L_n F(k) \rangle = \bigvee_{i=k}^n \langle K(i) \rangle$ [32, Proposition 5.3]. Therefore $X \to L_{F(k)} L_n X$ is a $(K(k) \vee \cdots \vee K(n))$–equivalence, and we only need show that $L_{F(k)} L_n X$ is $(K(k) \vee \cdots \vee K(n))$–local. But $L_{F(k)} L_n X \simeq F(M_{k,n}S^0, L_n X)$ and so it follows from Lemma 2.22 that $L_{F(k)} L_n X$ is $(K(k) \vee \cdots \vee K(n))$–local. We conclude that $L_{k,n}X \simeq L_{F(k)} L_n X$, as required. □

Remark 2.25 The equivalence

$$L_{k,n}X \simeq \varprojlim_j (L_n X \wedge M_k(j))$$

has also been obtained in [11, Proposition 6.21] using the theory and complete and torsion objects in a stable ∞–category. The next result is also contained in [11, Corollary 6.17].

Proposition 2.26 *For any spectrum X there is a pullback square*

$$\begin{tikzcd} L_n X \arrow[r] \arrow[d] & L_{k,n} X \arrow[d] \\ L_{k-1} X \arrow[r] & L_{k-1} L_{k,n} X \end{tikzcd}$$

Proof This is a standard consequence of the Bousfield decomposition

$$\langle E(n) \rangle = \langle E(k-1) \rangle \vee \langle E(n, J_k) \rangle$$

using for example [13, Proposition 2.2] (or, in the ∞–categorical setting, see [1]). □

Remark 2.27 Use [13, Proposition 2.2] one can deduce various other chromatic fracture squares. For example, we have a pullback square

$$\begin{tikzcd} L_{k,n} X \arrow[r] \arrow[d] & L_{k,h} X \arrow[d] \\ L_{h+1,n} X \arrow[r] & L_{k,h} L_{h+1,n} X \end{tikzcd}$$

for $k \le h \le n-1$.

Remark 2.28 These types of iterated chromatic localizations have been investigated by Bellumat and Strickland [53]. Results such as the chromatic fracture square can be recovered from their work; however we do not investigate this in detail.

Corollary 2.29 *Suppose M_j is a generalized Moore spectrum of type at least k. Then $L_n M_j \simeq L_{k,n} M_j$.*

Proof By definition, $L_{k-1} M_j \simeq *$, and so by the pullback square of Proposition 2.26 we must show that $L_{k-1} L_{k,n} M_j$ is contractible. Because M_j is a finite complex, this is equivalent to $L_{k-1}((L_{k,n} S^0) \wedge M_j) \simeq *$, and the result follows. □

Definition 2.30 Let $\mathcal{M}_{k,n}$ denote the essential image of the functor $M_{k,n} \colon \mathrm{Sp} \to \mathrm{Sp}$.

Theorem 2.31 *For any spectrum X, we have natural equivalences*

$$M_{k,n} L_{k,n} X \simeq M_{k,n} X, \quad L_{k,n} X \simeq L_{k,n} M_{k,n} X.$$

It follows that there is an equivalence of categories $\mathcal{M}_{k,n} \simeq \mathrm{Sp}_{k,n}$ given by $L_{k,n}$, with inverse given by $M_{k,n}$.

Proof The proof of Hovey and Strickland in the case $k = n$ generalizes essentially without change.

By definition, $M_{k,n} X$ fits into a cofiber sequence

$$M_{k,n} X \to L_n X \to L_{k-1} X,$$

so applying $L_{k,n}$ gives a cofiber sequence

$$L_{k,n} M_{k,n} X \to L_{k,n} L_n X \to L_{k,n} L_{k-1} X.$$

But $\langle E(k-1) \rangle = \bigvee_{i=0}^{k-1} \langle K(k) \rangle$, so $L_{k,n} L_{k-1} X \simeq 0$, while clearly $L_{k,n} L_n X \simeq L_{k,n} X$. It follows that $L_{k,n} M_{k,n} X \simeq L_{k,n} X$.

Using Proposition 2.24, $L_{k,n} X \simeq F(M_{k,n} S^0, L_n X)$, and so applying $F(-, L_n X)$ to the defining cofiber sequence for $M_{k,n} S^0$ we obtain a cofiber sequence

$$F(L_{k-1} S^0, L_n X) \to F(L_n S^0, L_n X) \to F(M_{k,n} S^0, L_n X),$$

or, equivalently,

$$F(L_{k-1} S^0, L_n X) \to L_n X \to L_{k,n} X.$$

It is easy to check that $F(L_{k-1} S^0, L_n X)$ is $E(k-1)$–local, and so by Lemma 2.22 we have $M_{k,n} F(L_{k-1} S^0, L_n X) \simeq *$. It follows that $M_{k,n} L_n X \simeq M_{k,n} X \simeq M_{k,n} L_{k,n} X$ as claimed. □

Remark 2.32 Once again, this result was obtained (by different methods) in [11, Proposition 6.21].

3 Thick subcategories and (co)localizing subcategories

In this section we compute the thick subcategories of compact objects in $\mathrm{Sp}_{k,n}$ and (co)localizing subcategories of $\mathrm{Sp}_{k,n}$. When $k=0$ or $k=n$ both results have been obtained by Hovey and Strickland. Along the way we give a classification of the compact objects in $\mathrm{Sp}_{k,n}$.

3A Compact objects in $\mathrm{Sp}_{k,n}$

In this section we characterize the compact objects in $\mathrm{Sp}_{k,n}$. We will use this in the next section to compute the thick subcategories of $\mathrm{Sp}_{k,n}^{\omega}$.

We begin by recalling the notions of thick and (co)localizing subcategories.

Definition 3.1 Let $(\mathcal{C}, \wedge, \mathbf{1})$ be an algebraic stable homotopy category, and let $\mathcal{D}$ be a full, stable subcategory.

1. $\mathcal{D}$ is called thick if it is closed under extensions and retracts.
2. $\mathcal{D}$ is called localizing if it is thick and closed under arbitrary colimits.
3. $\mathcal{D}$ is called colocalizing if it is thick and closed under arbitrary limits.
4. $\mathcal{D}$ is a tensor-ideal if $X \in \mathcal{C}$ and $Y \in \mathcal{D}$ implies $X \wedge Y \in \mathcal{D}$.
5. $\mathcal{D}$ is a coideal if $X \in \mathcal{C}$ and $Y \in \mathcal{D}$ implies $F(X,Y) \in \mathcal{D}$.

We will also speak of localizing (or thick) tensor-ideals and colocalizing coideals.

Remark 3.2 In Sp_n the dualizable and compact objects coincide, and are precisely those that lie in the thick subcategory generated by the tensor unit $L_n S^0$. In categories whose tensor unit is not compact, such as $\mathrm{Sp}_{k,n}$ for $k \neq 0$, the dualizable and compact objects do not coincide — for example, the tensor unit is always dualizable, but is not compact (Lemma 2.18). In [32] Hovey and Strickland gave numerous characterizations of compact objects in $\mathrm{Sp}_{K(n)}$. In this section we extend some of these characterizations to $\mathrm{Sp}_{k,n}$. We first recall the concept of a nilpotent object in a symmetric monoidal category.

Definition 3.3 We say that X is R–nilpotent if X lies in the thick $\otimes$–ideal generated by R, ie $X \in \mathrm{Thick}_{\otimes}\langle R\rangle$.

Lemma 3.4 *The category of E_n/I_k–nilpotent spectra is the same in $\mathrm{Sp}_{k,n}$, Sp_n and Sp.*

Proof Using that $\langle E_n/I_k\rangle = \langle E(n, J_k)\rangle$ (Proposition 2.10) we see that $E_n/I_k \wedge X$ is always $E(n, J_k)$–local, from which the result easily follows. $\square$

Remark 3.5 In other words, we can talk unambiguously about the category of E_n/I_k–nilpotent spectra.

We will also need the following generalization of [32, Lemma 6.15].

Lemma 3.6 *If X is a finite spectrum of type at least k, then $L_nX \simeq L_{k,n}X$ is E_n/I_k–nilpotent.*

Proof The argument is only a slight adaptation of that given by Hovey and Strickland. By a thick subcategory argument, we can assume that $X = M_k$ is a generalized Moore spectrum of type k. By [48], $L_nS^0 \in \mathrm{thick}_{\otimes}\langle E_n\rangle$, and it follows that $L_nS^0 \wedge M_k \simeq L_nM_k \in \mathrm{thick}_{\otimes}\langle E_n \wedge M_k\rangle$. But it is easy to see that $\mathrm{thick}_{\otimes}\langle E_n \wedge M_k\rangle \simeq \mathrm{thick}_{\otimes}\langle E_n/I_k\rangle$ and we are done. $\square$

Remark 3.7 The fact that $L_nS^0 \in \mathrm{thick}_{\otimes}\langle E_n\rangle$ is equivalent to the claim that $E_n \in \mathrm{Sp}_n$ is *descendable*, a condition we investigate further in Section 4A.

The compact objects in $\mathrm{Sp}_{k,n}$ can be characterized in the following ways, partially generalizing [32, Theorem 8.5]. We note that every compact object in $\mathrm{Sp}_{k,n}$ is automatically dualizable by [30, Theorem 2.1.3]; we investigate the dualizable objects in $\mathrm{Sp}_{k,n}$ in more detail in Section 5.

Theorem 3.8 *The following are equivalent for $X \in \mathrm{Sp}_{k,n}$:*

1. *X is compact.*
2. *$X \in \mathrm{thick}\langle L_nF(k)\rangle$.*
3. *X is a retract of $L_nX' \simeq L_{k,n}X'$ for a finite spectrum X' of type at least k.*
4. *X is a retract of $Y \wedge X'$, where Y is dualizable and X' is a finite spectrum of type at least k.*
5. *X is dualizable and E_n/I_k–nilpotent.*

The category $\mathrm{Sp}_{k,n}^{\omega} \subseteq \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ is thick. Moreover, if $X \in \mathrm{Sp}_{k,n}^{\omega}$ and $Y \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$, then $X \wedge Y$, $F(X, Y)$ and $F(Y, X)$ lie in $\mathrm{Sp}_{k,n}^{\omega}$. In particular, $F(X, L_{k,n}S^0) \in \mathrm{Sp}_n^{\omega}$.

Proof The equivalence of (1) and (2) is [30, Theorem 2.1.3] along with Proposition 2.15. Item (3) implies (2) because every finite spectrum of type at least k lies in the thick subcategory generated by $F(k)$. That (1) implies (3) is the same as given by Hovey and Strickland [32, Theorem 8.5]. Namely, suppose that $X \in \mathrm{Sp}_{k,n}$ is compact. By Proposition 2.24, $X \simeq \varinjlim_j (X \wedge DM_k(j))$, so $[X, X] \simeq \varinjlim_j [X, X \wedge DM_k(j)]$. In particular, X is a retract of $Y := X \wedge DM_k(j)$. We claim that such a Y is compact in Sp_n. Indeed, let $\{Z_i\}$ be a filtered diagram of E_n–local spectra. Then we have equivalences

$$\begin{aligned}[Y, \varinjlim_i Z_i]_* &\simeq [X, L_n M_k(j) \wedge \varinjlim_i Z_i]_* \\ &\simeq \big[X, L_{k,n} \varinjlim_i (M_k(j) \wedge Z_i)\big]_* \\ &\simeq \varprojlim_i [X, M_k(j) \wedge Z_i]_* \\ &\simeq \varprojlim_i [Y, Z_i]_*.\end{aligned}$$

The first and last equivalence follow by adjunction, the second because $\langle L_n M_k(j)\rangle = \bigvee_{i=k}^n \langle K(i)\rangle$ [32, Proposition 5.3], so

$$L_n M_k(j) \wedge \varinjlim_i Z_i \simeq L_n M_k(j) \wedge L_{k,n} \varinjlim_i Z_i \simeq L_{k,n} \varinjlim_i (M_k(j) \wedge Z_i),$$

while the third equivalence follows because $X \in \mathrm{Sp}_{k,n}^{\omega}$ by assumption and because $L_{k,n} \varinjlim_i$ is the colimit in $\mathrm{Sp}_{k,n}$. We have $K(i)_* Y = 0$ for $i < k$ and so Corollary 6.11 of [32] implies that Y, and hence X, is a retract of $L_n Z \simeq L_{k,n} Z$ for a finite spectrum Z of type at least k. This shows that (1), (2) and (3) are equivalent.

Assume now that (4) holds. Note that $Y \wedge X'$ is $E(n, J_k)$–local, and moreover $Y \wedge X' \simeq Y \wedge L_{k,n} X'$, where $L_{k,n} X' \in \mathrm{Sp}_{k,n}^{\omega}$. By [30, Theorem 2.1.3] the smash product of a dualizable and compact object is compact, and so X is a retract of a compact $E(n, J_k)$–local spectrum, and thus is also compact; ie (1) holds.

To see that (3) implies (5), we use a thick subcategory argument to reduce to the case that $X = L_n M_k \simeq L_{k,n} M_k$ is a localized generalized Moore spectrum of type k. Such an X is clearly dualizable and is additionally E_n/I_k–nilpotent by Lemma 3.6.

Now suppose that X satisfies (5). Following Hovey and Strickland [32, Proof of Corollary 12.16] let $\mathcal{J}$ be the collection of spectra $Z \in \mathrm{Sp}_{k,n}$ such that Z is a module over a generalized Moore spectrum of type i (for a fixed i with $k \le i \le n$). By [32, Proposition 4.17], $\mathcal{J}$ forms an ideal. Because $K(i) \wedge Z$ is nonzero and a wedge of suspensions of $K(i)$, $\mathcal{J}$ contains the ideal of $K(i)$–nilpotent spectra. Moreover, it follows from the Bousfield decomposition $\langle E_n/I_k\rangle = \bigvee_{i=k}^n \langle K(i)\rangle$ that $K(i)\wedge E_n/I_k \neq 0$, and so $\mathrm{thick}_{\otimes}\langle K(i)\rangle \subseteq \mathrm{thick}_{\otimes}\langle E_n/I_k\rangle$; ie every $K(i)$–nilpotent spectrum (for $k \le i \le n$) is also E_n/I_k–nilpotent. In particular, $X \in \mathcal{J}$, so X is retract of a spectrum of the form $Y \wedge X$ where Y is a generalized Moore spectrum of type i, and thus (4) holds.

Finally, we prove the subsidiary claims. It is immediate from (2) that $\mathrm{Sp}_{k,n}^{\omega} \subseteq \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ is thick, and it is an ideal by [30, Theorem 2.1.3(a)]. Because generalized Moore spectra are self-dual — see [32, Proposition 4.18] — (c) implies that $\mathrm{Sp}_{k,n}^{\omega}$ is closed under Spanier–Whitehead duality. Therefore, $F(X,Y) \simeq F(X, L_{k,n}S^0) \barwedge Y$ and $F(Y,X) \simeq X \barwedge F(Y, L_{k,n}S^0)$ lie in $\mathrm{Sp}_{k,n}^{\omega}$. ◻

Remark 3.9 When $k=0$, then X is compact if and only if X is dualizable [32, Theorem 6.2]. To reconcile this with (5) of the previous theorem, we note that every spectrum $X \in \mathrm{Sp}_n$ is $E_n/I_0 \simeq E_n$–nilpotent [48, Theorem 5.3].

3B The thick subcategory theorem

We now give a thick subcategory theorem for $\mathrm{Sp}_{k,n}^{\omega}$. As we shall see, given Theorem 3.8 this is an immediate consequence of the classification of thick subcategories of Sp_n^{ω}, which ultimately relies on the Devinatz–Hopkins–Smith nilpotence theorem.

Definition 3.10 For $0 \le j \le n+1$ let $\mathcal{C}_j$ denote the thick subcategory of Sp_n consisting of all compact spectra X such that $K(i)_*X = 0$ for all $i < j$; ie

$$\mathcal{C}_j = \{X \in \mathrm{Sp}_n^{\omega} \mid K(i)_*X = 0 \text{ for all } i < j\}.$$

Remark 3.11 By [32, Proposition 6.8], we equivalently have

$$\mathcal{C}_j = \{X \in \mathrm{Sp}_n^{\omega} \mid K(j-1)_*X = 0\}.$$

Remark 3.12 We have

$$\mathcal{C}_0 \supsetneq \mathcal{C}_1 \supsetneq \cdots \supsetneq \mathcal{C}_{n+1} = (0),$$

and moreover $L_nF(j)$ is in $\mathcal{C}_j$, but not $\mathcal{C}_{j+1}$.

We now present the result of Hovey and Strickland [32, Theorem 6.9].

Theorem 3.13 (Hovey–Strickland) *If $\mathcal{C}$ is a thick subcategory of Sp_n^{ω}, then $\mathcal{C} = \mathcal{C}_j$ for some j such that $0 \le j \le n+1$.*

Remark 3.14 This result can be restated in terms of the Balmer spectrum of Sp_n^{ω} [3]. In particular,

$$\mathrm{Spc}(\mathrm{Sp}_n^{\omega}) \cong \{\mathcal{C}_1, \ldots, \mathcal{C}_{n+1}\}$$

with topology determined by the closure operator $\overline{\{\mathcal{C}_j\}} = \{\mathcal{C}_i \mid i \geq j\}$. This is in fact equivalent to Theorem 3.13, essentially by the same argument as in [10, Proposition 3.5].

The Balmer support of $X \in \mathrm{Sp}_n^\omega$, defined in [3, Definition 2.1], is given by

$$\mathrm{supp}(X) = \{\mathcal{P} \in \mathrm{Spc}(\mathrm{Sp}_n^\omega) \mid X \notin \mathcal{P}\}.$$

Note that $X \notin \mathcal{C}_j$ if and only if $K(j-1)_*X \neq 0$. Therefore, by Theorem 3.13,

$$\mathrm{supp}(X) = \{i \in \{0,\dots,n\} \mid K(i)_*X \neq 0\} = \{i \in \{0,\dots,n\} \mid K(i) \wedge X \neq 0\}.$$

For a thick subcategory $\mathcal{J}$, we define $\mathrm{supp}(\mathcal{J}) = \bigcup_{X \in \mathcal{J}} \mathrm{supp}(X)$. Then, Balmer's classification result [3, Theorem 4.10] shows that there is a bijection

$$\{\text{thick subcategories of } \mathrm{Sp}_n^\omega\} \xrightarrow[\text{supp}]{\sim} \{\text{specialization closed subsets of } \{0,\dots,n\}\}.$$

with the topology on $\{0,\dots,n\}$ determined by $\overline{\{k\}} = \{k, k+1, \dots, n\}$, with inverse given by sending a specialization closed subset Y to $\{X \in \mathrm{Sp}_n^\omega \mid \mathrm{supp}(X) \subseteq Y\}$. Note that there are exactly $n+2$ such specialization closed subsets, namely $\varnothing$ and the subsets $\{k,\dots,n\}$ for $k = 0,\dots,n$. The thick subcategory $\mathcal{C}_{n+1}$ corresponds to $\varnothing$ under this bijection, while $\mathcal{C}_k$ corresponds to $\{k,\dots,n\}$ for $0 \leq k \leq n$.

Given the classification of compact $E(n, J_k)$–local spectra in Theorem 3.8, we deduce the following.

Lemma 3.15 *The category of compact $E(n, J_k)$–local spectra, $\mathrm{Sp}_{k,n}^\omega$, is equivalent to the thick subcategory $\mathcal{C}_k \subseteq \mathrm{Sp}_n^\omega$.*

Proof By [32, Corollary 6.11] if $X \in \mathcal{C}_k$, then X is a retract of $L_nY \simeq L_{k,n}Y$ for some finite spectrum Y of type of least k. Then X is a compact $E(n, J_k)$–local spectrum by Theorem 3.8. Conversely, if X is a compact $E(n, J_k)$–local spectrum, then X is a retract of $L_nY \simeq L_{k,n}Y$ for Y a finite spectrum Y of type of least k, again by Theorem 3.8. Therefore $K(i)_*X = 0$ for $i < k$ and $X \in \mathcal{C}_k$. □

Theorem 3.16 (thick subcategory theorem) *There is a bijection*

$$\{\textit{thick subcategories of } \mathrm{Sp}_{k,n}^\omega\} \xrightarrow[\text{supp}]{\sim} \{\textit{specialization closed subsets of } \{k,\dots,n\}\},$$

with inverse given by sending a specialization closed subset Y to

$$\{X \in \mathrm{Sp}_{k,n}^\omega \mid \mathrm{supp}(X) \subseteq Y\}.$$

In particular, if $\mathcal{C}$ is a thick subcategory of $\mathrm{Sp}_{k,n}^\omega$, then $\mathcal{C} = \mathcal{C}_j$ for some j such that $k \leq j \leq n+1$.

Proof This follows by combining Theorem 3.13 and Lemma 3.15. □

Remark 3.17 Note that $\mathrm{Sp}_{k,n}^{\omega}$ is not a tensor-triangulated category when $k \neq 0$, as it does not have a tensor unit. Therefore, we cannot speak of the Balmer spectrum of $\mathrm{Sp}_{k,n}^{\omega}$.

We also have a nilpotence theorem.

Proposition 3.18 *Let $X \in \mathrm{Sp}_{k,n}^{\omega}$, and $u\colon \Sigma^d X \to X$ a self-map such that $K(i)_* u$ is nilpotent for $k \le i \le n$. Then u is nilpotent; ie the j–fold composite*

$$u \circ \cdots \circ u\colon \Sigma^{jd} X \to X$$

is trivial for large enough j.

Proof In light of Lemma 3.15, this follows from [32, Corollary 6.6]. □

3C Localizing and colocalizing subcategories

In this section we calculate the (co)localizing (co)ideals of $\mathrm{Sp}_{k,n}$. We first observe that every (co)localizing subcategory is automatically a (co)ideal, so it suffices in fact to concentrate on (co)localizing subcategories.

Lemma 3.19 *Every (co)localizing subcategory of $\mathrm{Sp}_{k,n}$ is a (co)ideal.*

Proof We prove the case of localizing subcategories — the case of colocalizing subcategories is similar.[1] To that end, let $\mathcal{C} \subseteq \mathrm{Sp}_{k,n}$ be a localizing subcategory, and consider the collection $\mathcal{D} = \{X \in \mathrm{Sp} \mid X \mathbin{\overline{\wedge}} \mathcal{C} \subseteq \mathcal{C}\}$. This is a localizing subcategory of Sp containing $\mathbb{S}$, and hence $\mathcal{D} = \mathrm{Sp}$ itself. It follows that $\mathcal{C}$ is a localizing ideal. □

Remark 3.20 We remind the reader that $\mathbf{1}$ is not compact in $\mathrm{Sp}_{k,n}$ unless $k = 0$ (see Lemma 2.18). Therefore, in all other cases, $\mathbf{1}$ is a noncompact generator of $\mathrm{Sp}_{k,n}$.

Notation 3.21 Throughout this section we let $\mathcal{Q} = \{k, \dots, n\}$.

We begin by defining a notion of support and cosupport in $\mathrm{Sp}_{k,n}$, extending the notion of support defined previously for Sp_{n}^{ω}.

Definition 3.22 For a spectrum $X \in \mathrm{Sp}_{k,n}$, we define the support and cosupport of X by

$$\mathrm{supp}(X) = \{i \in \mathcal{Q} \mid K(i) \wedge X \neq 0\}, \quad \mathrm{cosupp}(X) = \{i \in \mathcal{Q} \mid F(K(i), X) \neq 0\}.$$

[1] We thank Neil Strickland for providing this argument, which simplifies a previous argument.

Example 3.23 Because $K(i)\wedge K(j)=0$ if $i\neq j$, and $K(i)\wedge K(i)\neq 0$ [46, Theorem 2.1], we have

$$\mathrm{supp}(K(i))=i$$

for $i\in\mathcal{Q}$. On the other hand, $K(i)^*K(j)=\mathrm{Hom}_{K(i)_*}(K(i)_*K(j),K(i)_*)$, and so

$$\mathrm{cosupp}(K(i))=i$$

as well.

Remark 3.24 The notion of support is slightly ambiguous, as objects can live in multiple categories. For example $L_{K(n)}S^0\in\mathrm{Sp}_{i,n}$ for all $0\le i\le n$, and in fact has different support in each category. However, it should also be clear in which category we are considering the support.

Remark 3.25 Because $K(i)\wedge X$ is always $K(i)$–local, we equivalently have

$$\mathrm{supp}(X)=\{i\in\mathcal{Q}\mid K(i)\barwedge X\neq 0\}.$$

Remark 3.26 In [32, Definition 6.7] Hovey and Strickland define the support of an object by

$$\mathrm{supp}_{\mathrm{HS}}(X)=\{i\mid K(i)\wedge X\neq 0\}.$$

By definition then, $\mathrm{supp}(X)=\mathrm{supp}_{\mathrm{HS}}(X)\cap\mathcal{Q}$.

Support and cosupport are well behaved with respect to products and function objects in $\mathrm{Sp}_{k,n}$.

Lemma 3.27 *For any* $X,Y\in\mathrm{Sp}_{k,n}$ *there are equalities*

$$\mathrm{supp}(X\barwedge Y)=\mathrm{supp}(X)\cap\mathrm{supp}(Y),\quad \mathrm{cosupp}(F(X,Y))=\mathrm{supp}(X)\cap\mathrm{cosupp}(Y).$$

Proof Because $K(i)\wedge X$ is always $K(i)$–local, $K(i)\wedge(X\barwedge Y)\simeq K(i)\wedge X\wedge Y$, and it is clear that $\mathrm{supp}(X\barwedge Y)\subseteq\mathrm{supp}(X)\cap\mathrm{supp}(Y)$. The converse follows because $K(i)_*$ is a graded field; if $K(i)\wedge X\wedge Y\simeq *$ then either $K(i)\wedge X\simeq *$ or $K(i)\wedge Y\simeq *$.

For the cosupport, suppose $i\in\mathrm{cosupp}(F(X,Y))$, ie $F(K(i),F(X,Y))\neq 0$. By adjunction we must have $F(K(i)\wedge X,Y)\neq 0$ as well as $F(X,F(K(i),Y))\neq 0$, so that $K(i)\wedge X\neq 0$ and $F(K(i),Y)\neq 0$. This shows that $\mathrm{cosupp}(F(X,Y))\subseteq\mathrm{supp}(X)\cap\mathrm{cosupp}(Y)$. For the converse, let $i\in\mathrm{supp}(X)\cap\mathrm{cosupp}(Y)$, and consider $F(K(i),F(X,Y))\simeq F(K(i)\wedge X,Y)$. Because $i\in\mathrm{supp}(X)$, and $K(i)_*$ is a graded field, $K(i) \wedge X$ is a wedge of suspensions of $K(i)$, and it suffices to show that $F(K(i), Y) \not\simeq 0$, which is precisely the statement that $i \in \operatorname{cosupp}(Y)$. Therefore, $i \in \operatorname{cosupp}(F(X, Y))$, as required. □

Notation 3.28 For an arbitrary collection $\mathcal{C}$ of objects we set

$$\operatorname{supp}(\mathcal{C}) = \bigcup_{X \in \mathcal{C}} \operatorname{supp}(X), \quad \operatorname{cosupp}(\mathcal{C}) = \bigcup_{X \in \mathcal{C}} \operatorname{cosupp}(X).$$

For a subset $\mathcal{T} \subseteq \mathcal{Q}$ we also define

$$\operatorname{supp}^{-1}(\mathcal{T}) = \{X \in \mathrm{Sp}_{k,n} \mid \operatorname{supp}(X) \subseteq \mathcal{T}\},$$
$$\operatorname{cosupp}^{-1}(\mathcal{T}) = \{X \in \mathrm{Sp}_{k,n} \mid \operatorname{cosupp}(X) \subseteq \mathcal{T}\}.$$

Lemma 3.29 *For a subset $\mathcal{T} \subseteq \mathcal{Q}$, $\operatorname{supp}^{-1}(\mathcal{T})$ and $\operatorname{cosupp}^{-1}(\mathcal{T})$ are localizing and colocalizing subcategories of $\mathrm{Sp}_{k,n}$, respectively.*

Proof We simply note that

$$\operatorname{supp}^{-1}(\mathcal{T}) = \{X \in \mathrm{Sp}_{k,n} \mid K(i) \wedge X = 0 \text{ for all } i \in \mathcal{Q} \setminus \mathcal{T}\},$$
$$\operatorname{cosupp}^{-1}(\mathcal{T}) = \{X \in \mathrm{Sp}_{k,n} \mid F(K(i), X) \simeq 0 \text{ for all } i \in \mathcal{Q} \setminus \mathcal{T}\},$$

which are clearly (co)localizing subcategories of $\mathrm{Sp}_{k,n}$. □

We thus obtain maps

$$\{\text{localizing subcategories of } \mathrm{Sp}_{k,n}\} \underset{\operatorname{supp}^{-1}}{\overset{\operatorname{supp}}{\rightleftarrows}} \{\text{subsets of } \mathcal{Q}\}, \tag{3-1}$$

$$\{\text{colocalizing subcategories of } \mathrm{Sp}_{k,n}\} \underset{\operatorname{cosupp}^{-1}}{\overset{\operatorname{cosupp}}{\rightleftarrows}} \{\text{subsets of } \mathcal{Q}\}. \tag{3-2}$$

We will see that these are bijections. We need the following local–global principle, which is a slight variant of that given by Hovey and Strickland [32, Proposition 6.18].

Proposition 3.30 (local–global principle) *For any $X \in \mathrm{Sp}_{k,n}$,*

$$X \in \operatorname{Loc}_{\mathrm{Sp}_{k,n}}(X) = \operatorname{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \operatorname{supp}(X)),$$
$$X \in \operatorname{Coloc}_{\mathrm{Sp}_{k,n}}(X) = \operatorname{Coloc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \operatorname{cosupp}(X)).$$

Proof Because $X \in \mathrm{Sp}_{k,n} \subseteq \mathrm{Sp}_n$, applying [32, Proposition 6.18] we have

$$\begin{aligned} X &\in \operatorname{Loc}_{\mathrm{Sp}}(X) = \operatorname{Loc}_{\mathrm{Sp}}(K(i) \mid i \in \operatorname{supp}_{\mathrm{HS}}(X)), \\ X &\in \operatorname{Coloc}_{\mathrm{Sp}}(X) = \operatorname{Coloc}_{\mathrm{Sp}}(K(i) \mid i \in \operatorname{cosupp}(X)). \end{aligned} \tag{3-3}$$

The result for colocalizing subcategories is then clear, as we get the same result taking the colocalizing subcategories in $\mathrm{Sp}_{k,n}$. For localizing subcategories we apply [8, Lemma 2.5] to (3-3) with the colimit-preserving functor $F = L_{k,n}\colon \mathrm{Sp} \to \mathrm{Sp}_{k,n}$ to see that

$$L_{k,n}X \simeq X \in \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(X) = \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathrm{supp}_{\mathrm{HS}}(X) \cap \mathcal{Q}),$$

where we have used that

$$L_{k,n}K(i) \simeq \begin{cases} K(i) & \text{if } i \in \mathcal{Q}, \\ 0 & \text{if } i \notin \mathcal{Q}. \end{cases}$$

As noted in Remark 3.26, $\mathrm{supp}_{\mathrm{HS}}(X) \cap \mathcal{Q} = \mathrm{supp}(X)$, and the result follows. □

Remark 3.31 If follows from the local–global principle that both support and cosupport detect trivial objects:

$$\mathrm{supp}(X) = \varnothing \iff X \simeq 0 \iff \mathrm{cosupp}(X) = \varnothing.$$

Corollary 3.32 *We have*

$$\mathrm{supp}^{-1}(\mathcal{T}) = \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathcal{T}), \quad \mathrm{cosupp}^{-1}(\mathcal{T}) = \mathrm{Coloc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathcal{T}).$$

Proof Let $\mathcal{A} = \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathcal{T})$. Because $\mathrm{supp}(K(i)) = i$ (Example 3.23), it is clear that $\mathcal{A} \subseteq \mathrm{supp}^{-1}(\mathcal{T})$. Conversely, if $X \in \mathrm{supp}^{-1}(\mathcal{T})$, then Proposition 3.30 shows that

$$X \in \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathcal{T}) = \mathcal{A},$$

so $\mathrm{supp}^{-1}(\mathcal{T}) = \mathcal{A}$, as claimed. The argument for colocalizing categories is similar. □

We now give the promised classification of localizing and colocalizing subcategories.

Theorem 3.33 (1) *The maps* (3-1) *give an order-preserving bijection between localizing subcategories of* $\mathrm{Sp}_{k,n}$ *and subsets of* $\mathcal{Q} = \{k, \dots, n\}$.

(2) *The maps* (3-2) *give an order-preserving bijection between colocalizing subcategories of* $\mathrm{Sp}_{k,n}$ *and subsets of* $\mathcal{Q} = \{k, \dots, n\}$.

Proof Let $\mathcal{C} \subseteq \mathrm{Sp}_{k,n}$ be a localizing subcategory and $\mathcal{T} \subseteq \{k, \dots, n\}$ a subset. Then via Corollary 3.32 and basic properties of support,

$$\mathrm{supp}(\mathrm{supp}^{-1}(\mathcal{T})) = \bigcup_{i \in \mathcal{T}} \mathrm{supp}(K(i)) = \mathcal{T}.$$

Now suppose that $X \in \mathcal{C}$, so that $\mathrm{supp}(X) \subseteq \mathrm{supp}(\mathcal{C})$. It follows from the definitions that $X \in \mathrm{supp}^{-1}(\mathrm{supp}(\mathcal{C}))$, and so $\mathcal{C} \subseteq \mathrm{supp}^{-1}(\mathrm{supp}(\mathcal{C}))$. We are therefore reduced to showing that $\operatorname{supp}^{-1}(\operatorname{supp}(\mathcal{C})) \subseteq \mathcal{C}$. To that end, let $Y \in \operatorname{supp}^{-1}(\operatorname{supp}(\mathcal{C}))$, so that $\operatorname{supp}(Y) \subseteq \operatorname{supp}(C)$. Using the local–global principle, Proposition 3.30, we then have

$$Y \in \operatorname{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \operatorname{supp}(Y)) \subseteq \operatorname{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \operatorname{supp}(\mathcal{C})) = \mathcal{C},$$

where the last equality follows from Proposition 3.30 again. The proof for colocalizing subcategories is analogous. □

Notation 3.34 For the following, we recall that for $\mathcal{C} \subseteq \mathrm{Sp}_{k,n}$ the right orthogonal $\mathcal{C}^{\perp}$ is defined as

$$\mathcal{C}^{\perp} = \{Y \in \mathrm{Sp}_{k,n} \mid F(X,Y) = 0 \text{ for all } X \in \mathcal{C}\}.$$

Similarly, the left orthogonal ${}^{\perp}\mathcal{C}$ is

$${}^{\perp}\mathcal{C} = \{Y \in \mathrm{Sp}_{k,n} \mid F(Y,X) = 0 \text{ for all } X \in \mathcal{C}\}.$$

Moreover, the right orthogonal is a colocalizing subcategory, and the left orthogonal is a localizing subcategory.

Corollary 3.35 *The map that sends a localizing subcategory $\mathcal{C}$ of $\mathrm{Sp}_{k,n}$ to $\mathcal{C}^{\perp}$ induces a bijection*

$$\{\textit{Localizing subcategories of } \mathrm{Sp}_{k,n}\} \xrightarrow{\sim} \{\textit{Colocalizing subcategories of } \mathrm{Sp}_{k,n}\}. \tag{3-4}$$

The inverse map sends a colocalizing subcategory $\mathcal{U}$ to ${}^{\perp}\mathcal{U}$.

Proof We follow [14, Corollary 9.9]. Let $\mathcal{C}$ be a localizing subcategory; then, using Remark 3.31 and Lemma 3.27,

$$\begin{aligned} \mathcal{C}^{\perp} &= \{Y \in \mathrm{Sp}_{k,n} \mid F(X,Y) = 0 \text{ for all } X \in \mathcal{C}\} \\ &= \{Y \in \mathrm{Sp}_{k,n} \mid \operatorname{cosupp}(Y) \cap \operatorname{supp}(\mathcal{C}) = \varnothing\} \\ &= \{Y \in \mathrm{Sp}_{k,n} \mid \operatorname{cosupp}(Y) \subseteq \mathcal{Q} \setminus \operatorname{supp}(\mathcal{C})\} \\ &= \operatorname{cosupp}^{-1}(\mathcal{Q} \setminus \operatorname{supp}(\mathcal{C})). \end{aligned}$$

Similarly, if $\mathcal{U}$ is a colocalizing subcategory, then

$$\begin{aligned} {}^{\perp}\mathcal{U} &= \{X \in \mathrm{Sp}_{k,n} \mid F(X,Y) = 0 \text{ for all } Y \in \mathcal{U}\} \\ &= \{X \in \mathrm{Sp}_{k,n} \mid \operatorname{cosupp}(\mathcal{U}) \cap \operatorname{supp}(X) = \varnothing\} \\ &= \{X \in \mathrm{Sp}_{k,n} \mid \operatorname{supp}(X) \subseteq \mathcal{Q} \setminus \operatorname{cosupp}(\mathcal{U})\} \\ &= \operatorname{supp}^{-1}(\mathcal{Q} \setminus \operatorname{cosupp}(\mathcal{U})). \end{aligned}$$

It follows that under the equivalences of Theorem 3.33, the maps $\mathcal{C} \mapsto \mathcal{C}^{\perp}$ and $\mathcal{U} \mapsto {}^{\perp}\mathcal{U}$ correspond to the map $\mathcal{Q} \to \mathcal{Q}$ sending a subset to its complement, and are thus mutually inverse bijections. □

3D The Bousfield lattice

We recall the basics on the Bousfield lattice of an algebraic stable homotopy theory. In order to avoid confusion with the (localized) categories of spectra considered previously we let $(\mathcal{C}, \wedge, \mathbf{1})$ denote a tensor triangulated category.

Definition 3.36 The Bousfield class of an object $X \in \mathcal{C}$ is the full subcategory of objects
$$\langle X \rangle = \{W \in \mathcal{C} \mid X \wedge W = 0\}.$$

Remark 3.37 We always assume that our categories are compactly generated and hence there is a set of Bousfield classes [34, Theorem 3.1].

Remark 3.38 We let $\mathsf{BL}(\mathcal{C})$ denote the set of Bousfield classes of $\mathcal{C}$. As is known, this has a lattice structure, which we now describe. We say that $\langle X \rangle \leq \langle Y \rangle$ if $Y \wedge W = 0$ implies $X \wedge W = 0$. Hence, $\langle 0 \rangle$ is the minimum Bousfield class, and $\langle \mathbf{1} \rangle$ is the maximum. The join is defined by $\bigvee_{i \in I} \langle X_i \rangle = \langle \coprod_{i \in I} X_i \rangle$, and the meet is the join of all lower bounds.

Proposition 3.39 *The Bousfield lattice* $\mathsf{BL}(\mathrm{Sp}_{k,n})$ *is isomorphic to the lattice of subsets of* $\mathcal{Q}$ *via the map sending* $\langle X \rangle$ *to* $\mathrm{supp}(X)$.

Proof Define a map that sends $\mathcal{T} \subseteq \mathcal{Q}$ to $\langle \bigvee_{i \in \mathcal{T}} K(i) \rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$. We claim that this gives the necessary inverse map. By the local–global principle (Proposition 3.30),
$$\mathrm{Loc}_{\mathrm{Sp}_{k,n}}(X) = \mathrm{Loc}_{\mathrm{Sp}_{k,n}}(K(i) \mid i \in \mathrm{supp}(X)).$$
In particular, $X \mathbin{\overline{\wedge}} W \simeq 0$ if and only if $K(i) \mathbin{\overline{\wedge}} W \simeq 0$ for all $i \in \mathrm{supp}(X)$, so
$$\langle X \rangle = \Big\langle \bigvee_{i \in \mathrm{supp}(X)} K(i) \Big\rangle. \tag{3-5}$$
The result then follows by direct computation. □

3E The telescope conjecture and variants

We begin by considering variants of the telescope conjecture in the localized categories $\mathrm{Sp}_{k,n}$ using work of Wolcott [55].

Definition 3.40 For $i \in \mathcal{Q}$, let $l_i^f : \mathrm{Sp}_{k,n} \to \mathrm{Sp}_{k,n}$ denote finite localization away from $L_{k,n} F(i+1)$.

Remark 3.41 Because $L_{k,n}F(i+1)$ is in $\mathrm{Sp}^{\omega}_{k,n}$ by Theorem 3.8, this is a smashing localization.

Remark 3.42 By [55, Proposition 3.8] we have an equivalence of endofunctors of $\mathrm{Sp}_{k,n}$ (recall that $\langle \mathrm{Tel}(n)\rangle$ is the Bousfield class of a telescope of a finite type n spectrum),

$$l_i^f \simeq L_{L_{k,n}\mathrm{Tel}(0)\vee L_{k,n}\mathrm{Tel}(1)\vee\cdots\vee L_{k,n}\mathrm{Tel}(i)}.$$

We note that $L_{k,n}\,\mathrm{Tel}(j)$ is trivial when $j \notin \mathcal{Q}$ by [48, Proposition A.2.13]. In particular,

$$l_i^f \simeq L_{L_{k,n}\mathrm{Tel}(k)\vee\cdots\vee L_{k,n}\mathrm{Tel}(i)}.$$

We also consider the following Bousfield localization on $\mathrm{Sp}_{k,n}$.

Definition 3.43 For $i \in \mathcal{Q}$, let $l_i\colon \mathrm{Sp}_{k,n} \to \mathrm{Sp}_{k,n}$ denote Bousfield localization at $K(k)\vee K(k+1)\vee\cdots\vee K(i)$.

Remark 3.44 Following Wolcott [55], we consider the following variants of the telescope conjecture on $\mathrm{Sp}_{k,n}$ for $i \in \mathcal{Q}$:

- $\mathsf{LTC1_i}$ $\langle L_{k,n}\,\mathrm{Tel}(i)\rangle = \langle K(i)\rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$.
- $\mathsf{LTC2_i}$ $l_i^f X \xrightarrow{\ \simeq\ } l_i X$ for all X, or equivalently,
$$\left\langle \bigvee_{j=k}^{i} L_{k,n}\,\mathrm{Tel}(j)\right\rangle = \left\langle \bigvee_{j=k}^{i} K(j)\right\rangle$$
in $\mathsf{BL}(\mathrm{Sp}_{k,n})$.
- $\mathsf{LTC3_i}$ If X is a type i spectrum and f is a v_i self-map, $l_i(L_{k,n}X) \cong L_{k,n}(f^{-1}X)$.
- GSC Every smashing localization is generated by a set of compact objects.
- SDGSC Every smashing localization is generated by a set of dualizable objects.

Here LTC stands for the localized telescope conjecture, GSC is the generalized smashing conjecture, and SDGSC is the strongly dualizable generalized smashing conjecture. We emphasize the difference here because compact and dualizable objects do not coincide in $\mathrm{Sp}_{k,n}$ when $k \neq 0$.

Proposition 3.45 *On $\mathrm{Sp}_{k,n}$, we have that $\mathsf{LTC1_i}$, $\mathsf{LTC2_i}$ and $\mathsf{LTC3_i}$ hold for all $i \in \mathcal{Q}$.*

Proof By [55, Theorem 3.12] it suffices to prove that $\mathsf{LTC1_i}$ holds. By Proposition 3.39 this will follow if we show that $L_{k,n}\,\mathrm{Tel}(i)$ and $K(i)$ have the same support in $\mathrm{Sp}_{k,n}$. To see this, we have $\mathrm{supp}(K(i)) = \{i\}$ by Example 3.23, while $\mathrm{supp}(\mathrm{Tel}(i)) = \{i\}$ by [55, Lemmas 2.10 and 3.7]. □

We now classify all smashing localizations on $\mathrm{Sp}_{k,n}$ and show that all variants of the telescope conjecture hold.

Theorem 3.46 *Let L be a nontrivial smashing localization functor on $\mathrm{Sp}_{k,n}$. Then $L \simeq l_j^f \simeq l_j$ for some $j \in \mathcal{Q}$. In particular, the* GSC *and* SDGSC *both hold in* $\mathrm{Sp}_{k,n}$.

Proof We closely follow [55, Theorem 4.4]. Throughout the proof we let $\mathbf{1}$ denote $L_{k,n}S^0$, the monoidal unit in $\mathrm{Sp}_{k,n}$, so that $\langle L\rangle = \langle L\mathbf{1}\rangle$. By (3-5),

$$\langle L\mathbf{1}\rangle = \Big\langle \bigvee_{i\in \mathrm{supp}(L\mathbf{1})} K(i)\Big\rangle.$$

Note that $\mathrm{supp}(L\mathbf{1})$ is nonempty because we assume $L \neq 0$. Hence, we can fix $j \in \mathrm{supp}(L\mathbf{1})$ such that $\langle K(j)\rangle \le \langle L(\mathbf{1})\rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$. It follows that $L_{K(j)}L \simeq LL_{K(j)} \simeq L_{K(j)}$, and $\langle L_{K(j)}\mathbf{1}\rangle = \langle L_{K(j)}\mathbf{1} \,\overline{\wedge}\, L\mathbf{1}\rangle \le \langle L\mathbf{1}\rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$. We also note that $L_{K(j)}\mathbf{1} = L_{K(j)}L_{k,n}S^0 \simeq L_{K(j)}S^0$.

By [32, Proposition 5.3], $\langle L_{K(j)}S^0\rangle = \bigvee_{i=0}^{j}\langle K(i)\rangle$ in $\mathsf{BL}(\mathrm{Sp})$, and it follows easily that $\langle L_{K(j)}S^0\rangle = \bigvee_{i=k}^{j}\langle K(i)\rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$. It follows that $\langle L\mathbf{1}\rangle \ge \bigvee_{i=k}^{j}\langle K(i)\rangle$ in $\mathsf{BL}(\mathrm{Sp}_{k,n})$. We deduce that $\langle L\mathbf{1}\rangle = \bigvee_{i=k}^{j}\langle K(i)\rangle$, where $j = \max\{\mathrm{supp}(L\mathbf{1})\}$, and hence by Proposition 3.45 that $L \simeq l_j^f \simeq l_j$. Finally, because $L_nF(j+1)$ is compact and therefore also dualizable in $\mathrm{Sp}_{k,n}$, both the GSC and SDGSC hold in $\mathrm{Sp}_{k,n}$. □

Remark 3.47 Using [30, Proposition 3.8.3] and Theorems 3.33 and 3.46 one can reprove the thick subcategory theorem Theorem 3.16.

4 Descent theory and the $E(n, J_k)$–local Adams spectral sequence

In this section we use descent theory to construct an Adams-type spectral sequence in the $E(n, J_k)$–local category. Using descent, we shall see that this has a vanishing line at some finite stage. Moreover, for $p \gg n$, we show that the $E(n, J_k)$–local Adams spectral sequence computing $\pi_* L_{k,n}S^0$ has a horizontal vanishing line on the E_2–page, and there are no nontrivial differentials.

4A Descendability

We begin with the notion of a descendable object in an algebraic stable homotopy category.

Remark 4.1 We recall that in $\mathcal{C}$ there is an ∞–category $\mathrm{CAlg}(\mathcal{C})$ of commutative algebra objects; see [40, Chapter 3]. Moreover, given $A \in \mathrm{CAlg}(\mathcal{C})$ we can define a stable, presentable, symmetric monoidal ∞–category $\mathrm{Mod}_A(\mathcal{C})$ of A–modules internal to $\mathcal{C}$, with the relative A–linear tensor product [40, Section 4.5]. We will mainly focus on the case $A = E_n$ and $\mathcal{C} = \mathrm{Sp}_{k,n}$, so that $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ denotes the ∞–category of $E(n, J_k)$–local E_n–modules, that is E_n–modules whose underlying spectrum is $E(n, J_k)$–local, with monoidal structure $A \wedge B = L_{k,n}(A \wedge_E B)$.

Remark 4.2 Note that $E_n \in \mathrm{CAlg}(\mathrm{Sp})$ by the Goerss–Hopkins–Miller [21] theorem, and so $E_n \in \mathrm{CAlg}(\mathrm{Sp}_{k,n})$ as well. On the other hand, $E(n, J_k)$ will not, in general, be a commutative ring spectrum (for example, $K(n)$ is never a commutative ring spectrum).

Definition 4.3 [41, Definition 3.18] A commutative algebra object $A \in \mathrm{CAlg}(\mathcal{C})$ is said to be descendable if $\mathbf{1} \in \mathcal{C}$ is A–nilpotent (Definition 3.3), or equivalently $\mathcal{C} = \mathrm{thick}_{\otimes}\langle A\rangle$.

One reason to be interested in descendable objects is the following [41, Proposition 3.22].

Proposition 4.4 (Mathew) *Let $A \in \mathrm{CAlg}(\mathcal{C})$ be descendable. Then the adjunction $C \leftrightarrows \mathrm{Mod}_C(A)$ given by tensoring with A and forgetting is comonadic. In particular, the natural functor from C to the totalization*

$$\mathcal{C} \to \mathrm{Tot}\big(\mathrm{Mod}_A(\mathcal{C}) \rightrightarrows \mathrm{Mod}_{A\wedge A}(\mathcal{C}) \substack{\to\\ \to\\ \to} \big)$$

is an equivalence.

We also note the following [41, Proposition 3.19].

Proposition 4.5 (Mathew) *If $A \in \mathrm{CAlg}(\mathcal{C})$ is descendable, then the functor*

$$\mathcal{C} \to \mathrm{Mod}_A(\mathcal{C}), \quad M \mapsto M \wedge A,$$

is conservative.

4B Morava modules and L–complete comodules

The following theorem, essentially due to Hopkins–Ravenel [48], shows that the results of the previous section can be applied in $\mathrm{Sp}_{k,n}$. We note that $E_n \in \mathrm{Sp}$ is a commutative algebra object; this is the Goerss–Hopkins–Miller theorem [21]. It follows that $E_n \in \mathrm{CAlg}(\mathrm{Sp}_{k,n})$.

Theorem 4.6 *$E_n \in \mathrm{CAlg}(\mathrm{Sp}_{k,n})$ is descendable, and there is an equivalence of symmetric-monoidal stable ∞–categories*

$$\mathrm{Sp}_{k,n} \simeq \mathrm{Tot}\big(\,\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}) \rightrightarrows \mathrm{Mod}_{E_n \overline{\wedge} E_n}(\mathrm{Sp}_{k,n}) \,\substack{\rightarrow\\\rightarrow\\\rightarrow}\,\big) \tag{4-1}$$

Proof It is consequence of the Hopkins–Ravenel smash product theorem that $E_n \in \mathrm{CAlg}(\mathrm{Sp}_n)$ is descendable; see [41, Theorem 4.18]. It follows from [41, Corollary 3.21] that $L_{k,n}E_n \simeq E_n$ is descendable in $\mathrm{Sp}_{k,n}$. The equivalence then follows from Proposition 4.4. ◻

By Proposition 4.5 we deduce the following.

Corollary 4.7 *The functor $E_n \overline{\wedge} (-) \colon \mathrm{Sp}_{k,n} \to \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ is conservative.*

We therefore define the following.

Definition 4.8 For $X \in \mathrm{Sp}_{k,n}$ the Morava module of X is $(E_{k,n})^\vee_* X := \pi_*(E_n \overline{\wedge} X)$.

We recall that $L_{k,n}X \simeq \varprojlim_j (L_n X \wedge M_k(j))$ (Proposition 2.24). The Milnor sequence then gives the following.

Lemma 4.9 *There is a short exact sequence*

$$0 \to \varprojlim_j{}^1 (E_n)_{*+1}(X \wedge M_j(k)) \to (E_{k,n})^\vee_* X \to \varprojlim_j E_*(X \wedge M_j(k)) \to 0.$$

Example 4.10 If $(E_n)_*X$ is a free $(E_n)_*$–module, then the $\varprojlim^1$ term vanishes and it follows that $(E_{k,n})^\vee_* X = (E_*X)^\wedge_{I_k}$.

Remark 4.11 As the short exact sequence shows, $(E_{k,n})^\vee_* X$ is not always complete with respect to the I_k–adic topology. However, it is always $L_0^{I_k}$–complete in the sense of [32, Appendix A] — this the same argument as given in [32, Proposition 8.4(a)].

4C The $E(n, J_k)$–local E_n–Adams spectral sequence

In this section we construct an Adams-type spectral sequence in the $E(n, J_k)$–local category. When $k = 0$, this is the E_n–Adams spectral sequence, while when $k = n$ this is the $K(n)$–local E_n–Adams spectral sequence considered in [16, Appendix A].

To begin, we recall that the cobar (or Amitsur) complex for E_n in $\mathrm{Sp}_{k,n}$ is

$$CB^\bullet(E_n)\colon\ E_n \rightrightarrows E_n \mathbin{\overline{\wedge}} E_n \mathrel{\substack{\textstyle\rightarrow\\[-0.6ex]\textstyle\rightarrow\\[-0.6ex]\textstyle\rightarrow}} \cdots.$$

Definition 4.12 Let $\widehat{\mathrm{Ext}}^{s,*}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*,(E_n)_*) := H^s\bigl(\pi_*(CB^\bullet(E_n))\bigr)$, ie it is the cohomology of the complex

$$(E_n)_* \rightrightarrows (E_{k,n})^\vee_*(E_n) \mathrel{\substack{\textstyle\rightarrow\\[-0.6ex]\textstyle\rightarrow\\[-0.6ex]\textstyle\rightarrow}} \cdots.$$

More generally, we let

$$\widehat{\mathrm{Ext}}^{s,*}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*,(E_{k,n})^\vee_*(X)) := H^s\bigl(\pi_*(X \mathbin{\overline{\wedge}} CB^\bullet(E_n))\bigr).$$

Proposition 4.13 *For any spectrum X there is a strongly convergent spectral sequence*

$$E_2^{s,t} \cong \widehat{\mathrm{Ext}}^{s,*}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*,(E_{k,n})^\vee_*(X)) \Rightarrow \pi_*(L_{k,n}X)$$

which has a horizontal vanishing line at a finite stage (*independent of X*).

Proof This is the Bousfield–Kan spectral sequence associated to the tower

$$X \mathbin{\overline{\wedge}} CB^\bullet(E_n).$$

The claimed results are a consequence of descendability (Theorem 4.6); see [41, Corollary 4.4; 42, Example 2.11, Propositions 2.12 and 2.14]. □

Remark 4.14 For the $\mathrm{Sp}_{K(n)}$–local homotopy category, this completed Ext can be interpreted as an Ext group in the category of $L_0^{I_n}$–complete comodules [9]. In the case of $X = S^0$, Morava's change of rings theorem, in the form [9, Theorem 4.3], shows that

$$E_2^{s,t} \cong H_c^s(\mathbb{G}_n,(E_n)_t),$$

the continuous cohomology of the Morava stabilizer group $\mathbb{G}_n$, and this spectral sequence is isomorphic to that considered by Devinatz and Hopkins in [16, Appendix A]. The key point is the computation that

$$(E_n)^\vee_*(E_n) \cong \mathrm{Hom}^c(\mathbb{G}_n,(E_n)_*),$$

for which see [29]. We remark that we do not know what $(E_{k,n})^\vee_*(E_n)$ is for $k \ne n$. However, the same arguments as in [9] go through; the pair $((E_n)_*,(E_{k,n})^\vee_*(E_n))$ is an $L_0^{I_k}$–complete comodule, and if $(E_{k,n})^\vee_*(X)$ is either a finitely generated $(E_n)_*$–module, is the I_k–adic completion of a free-module, or has bounded I_k–torsion, then $(E_{k,n})^\vee_*(X)$ is a comodule over this Hopf algebroid — see [9, Lemma 1.17 and

Proposition 1.22]. The relative homological algebra studied in [9, Section 2] also goes through to see that $\widehat{\mathrm{Ext}}^{s,*}$ as used above is a relative Ext group in the category of $L_0^{I_k}$–complete comodules. We will not use this in what follows, so we leave the details to the interested reader.

Remark 4.15 In [26, Section 7] the authors construct the $K(n)$–local Adams spectral sequence for dualizable $K(n)$–local X as the inverse limit of the E_n–Adams spectral sequences for $X \wedge M_n(j)$. The following result recovers the identification of the E_2–term in the case $k = n$.

Proposition 4.16 *Let $M_k(j)$ be a tower of generalized Moore spectra of height k. Then there is an isomorphism*

$$\widehat{\mathrm{Ext}}^{s,t}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*, (E_n)_*) \cong \varprojlim_j \mathrm{Ext}^{s,t}_{(E_n)_*E_n}\big((E_n)_*, (E_n)_*(M_k(j))\big).$$

Proof By definition, $\widehat{\mathrm{Ext}}^{s,t}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*, (E_n)_*)$ is the cohomology of the complex

$$(E_n)_* \rightrightarrows (E_{k,n})^\vee_*(E_n) \overset{\longrightarrow}{\underset{\longrightarrow}{\longrightarrow}} \cdots .$$

The t^{th} term of this complex is the homotopy of $L_{k,n}(E_n^{\wedge t}) \simeq \varprojlim_j (E_n^{\wedge t} \wedge M_k(j))$ by Proposition 2.24, and there is a corresponding Milnor exact sequence of the form

$$0 \to \varprojlim_j{}^1 \pi_{q+1}(E_n^{\wedge t} \wedge M_k(j)) \to \pi_q(L_{k,n}(E_n^{\wedge t})) \to \varprojlim_j \pi_q(E_n^{\wedge t} \wedge M_k(j)) \to 0.$$

We note that $E_n^{\wedge t}$ is Landweber exact, as the smash product of Landweber exact spectra; see [12, Lemma 4.3]. It follows that $\pi_*(E_n^{\wedge t} \wedge M_k(j)) \cong \pi_*(E_n^{\wedge t})/(p^{i_0}, \dots, u_{k-1}^{i_{k-1}})$ for suitable integers $i_0, \dots, i_{k-1}$. In particular, the maps in the tower are surjections by the construction of the tower $\{M_k(j)\}$ (see Remark 2.19), and so the $\varprojlim_j^1$–term vanishes, and

$$(E_{k,n})^\vee_*(E_n^{t-1}) \cong \pi_*(L_{k,n}(E_n^{\wedge t})) \cong \varprojlim_j \pi_q(E_n^{\wedge t} \wedge M_k(j)).$$

Note that the cohomology of the complex $\{\pi_q(E_n^t \wedge M_k(j))\}_t$ is

$$\mathrm{Ext}^{*,*}_{(E_n)_*E_n}\big((E_n)_*, (E_n)_*(M_k(j))\big).$$

Therefore, there is an exact sequence

$$\begin{aligned} 0 \to \varprojlim_j{}^1 \mathrm{Ext}^{q-1,*}_{(E_n)_*E_n}\big((E_n)_*, (E_n)_*(M_k(j))\big) &\to \widehat{\mathrm{Ext}}^{q,*}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*, (E_n)_*) \\ &\to \varprojlim_j \mathrm{Ext}^{q,*}_{(E_n)_*E_n}\big((E_n)_*, (E_n)_*(M_k(j))\big) \to 0. \end{aligned}$$

We will see below in Corollary 4.22 that $\operatorname{Ext}^{q,*}_{(E_n)_*E_n}\big((E_n)_*,(E_n)_*(M_k(j))\big)$ is finite, and so the $\varprojlim^1_j$–term vanishes in the exact sequence, and the result follows. □

Remark 4.17 It follows that when $k \neq 0$, the groups $\widehat{\operatorname{Ext}}^{s,t}_{(E_{k,n})^{\vee}_*(E_n)}((E_n)_*,(E_n)_*)$ are profinite, ie either finite or uncountable. Contrast the case $k=0$, where

$$\operatorname{Ext}^{s,t}_{(E_n)_*E_n}((E_n)_*,(E_n)_*)$$

is countable [28, Proof of Lemma 5.4].

4D Vanishing lines in the $E(n,J_k)$–local E_n–Adams spectral sequence

In Proposition 4.13 we constructed a spectral sequence

$$\widehat{\operatorname{Ext}}^{s,t}_{(E_{k,n})^{\vee}_*(E_n)}((E_n)_*,(E_n)_*) \Rightarrow \pi_{t-s}(L_{k,n}S^0),$$

and showed that, as a consequence of descendability, this has a horizontal vanishing line at some finite stage. In the extreme cases of $k=0$ and $k=n$ it is known that when $p \gg n$ and $X=S^0$, this vanishing line occurs on the E_2–page, and occurs at $s=n^2+n$ and $s=n^2$, respectively; see [31, Theorem 5.1] and [47, Theorem 6.2.10]. In this section, we show (Theorem 4.24) that the analogous result occurs in general; for $p \gg n$ there is a vanishing line on the E_2–page of the spectral sequence of Proposition 4.13 above, and $s=n^2+n-k$ in the case $X=S^0$. The proof relies on a variant of the chromatic spectral sequence [47, Chapter 5], which we now construct. Along the way we prove Corollary 4.22, which also completes the proof of Proposition 4.16.

Remark 4.18 (the chromatic spectral sequence) Fix $k \le n$, and for $0 \le s \le n-k$ let M^s denote the $(E_n)_*(E_n)$–comodule

$$u_{k+s}^{-1}(E_n)_*/(p,u_1,\dots,u_{k-1},u_k^{\infty},\dots,u_{k+s-1}^{\infty}).$$

Arguing as in [47, Lemma 5.1.6], there is an exact sequence of $(E_n)_*(E_n)$–comodules

$$(E_n)_*/I_k \to M^0 \to M^1 \to \cdots \to M^{n-k} \to 0.$$

Applying [47, Theorem A.1.3.2], there is then a chromatic spectral sequence of the form

$$E_1^{s,r,*} \cong \operatorname{Ext}^{r,*}_{(E_n)_*(E_n)}((E_n)_*,M^s) \Rightarrow \operatorname{Ext}^{r+s,*}_{(E_n)_*(E_n)}((E_n)_*,(E_n)_*/I_k). \tag{4-2}$$

Proposition 4.19 *In the chromatic spectral sequence (4-2),*

$$E_1^{s,r,*} \cong \begin{cases} \operatorname{Ext}^{r,*}_{(E_{k+s})_*(E_{k+s})}((E_{k+s})_*,(E_{k+s})_*/(p,\dots,u_{k-1},u_k^{\infty},\dots,u_{k+s-1}^{\infty})) & \text{if } s \le n-k, \\ 0 & \text{if } s > n-k. \end{cases}$$

If particular, if $p-1$ does not divide $k+s$, we have $E_1^{s,r,*}=0$ for $r>(s+k)^2$. Thus, if $p-1$ does not divide $k+s$ for all $0 \le s \le n-k$,[2] *then*

$$\mathrm{Ext}^{s,*}_{(E_n)_*(E_n)}((E_n)_*,(E_n)_*/I_k)=0$$

for $s>n^2+n-k$.

Proof This is similar to the proof by Hovey and Sadofsky [31, Theorem 5.1], which is the case where $k=0$. We first recall the change of rings theorem of Hovey and Sadofsky [31, Theorem 3.1]; if M is a BP_*BP–comodule, on which v_j acts isomorphically, and $n \ge j$, then there is an isomorphism[3]

$$\mathrm{Ext}^{*,*}_{BP_*BP}(BP_*,M) \cong \mathrm{Ext}^{*,*}_{(E_n)_*(E_n)}((E_n)_*,(E_n)_* \otimes_{BP_*} M).$$

Applying this change of rings theorem twice to the BP_*BP–comodule

$$u_{k+s}^{-1}BP_*/(p,\dots,u_{k-1},u_k^{\infty},\dots,u_{k+s-1}^{\infty})$$

with $j=k+s$ and $j=n$ shows that the E_1–term has the claimed form.

For brevity, let us denote $I=(p,\dots,u_{k-1},u_k^{\infty},\dots,u_{k+s-1}^{\infty})$. By Morava's change of rings theorem,

$$\mathrm{Ext}^{r,*}_{(E_{k+s})_*(E_{k+s})}((E_{k+s})_*,(E_{k+s})_*/I) \cong H^r(\mathbb{G}_{k+s},(E_{k+s})_*/I).$$

Morava's vanishing theorem [47, Theorem 6.2.10] shows that if $p-1$ does not divide $k+s$, then

$$H^r(\mathbb{G}_{k+s},(E_{k+s})_*)=0$$

for $r>(k+s)^2$. Along with an argument similar to that given by Hovey and Sadofsky's, using standard exact sequences and taking direct limits we find that

$$\mathrm{Ext}^{r,*}_{(E_{k+s})_*(E_{k+s})}((E_{k+s})_*,(E_{k+s})_*/I)=0$$

for $r>(k+s)^2$ as well. □

Remark 4.20 Let M_k denote a generalized Moore spectrum of type k. Then there is an obvious analog of this spectral sequence, whose abutment is

$$\mathrm{Ext}^{r+s,*}_{(E_n)_*(E_n)}((E_n)_*,(E_n)_*(M_k)) \cong \mathrm{Ext}^{r+s,*}_{(E_n)_*(E_n)}((E_n)_*,(E_n)_*/(p^{i_0},\dots,u_{k-1}^{i_{k-1}}))$$

[2] Taking $p>n+1$ suffices, but may not be optimal.

[3] Hovey and Sadofsky work with $E(n)$ instead of E_n, but this does not change anything in light of [33, Theorem C].

with E_1–term of the form

$$E_1^{s,r,*} \cong \begin{cases} \mathrm{Ext}^{r,*}_{(E_{k+s})_*(E_{k+s})}((E_{k+s})_*, (E_{k+s})_*/(p^{i_0}, \dots, u_{k-1}^{i_{k-1}}, u_k^\infty, \dots, u_{k+s-1}^\infty)) & \text{if } s \le n-k, \\ 0 & \text{if } s > n-k. \end{cases}$$

Remark 4.21 The following completes the proof of Proposition 4.16.

Corollary 4.22 *Let M_k denote a generalized Moore spectrum of type k. Then the group $\mathrm{Ext}^{r,*}_{(E_n)_*(E_n)}((E_n)_*, (E_n)_*(M_k))$ is finite.*

Proof By taking appropriate exact sequences it suffices to show this for $(E_n)_*/I_k$ (alternatively, one can argue directly using the spectral sequence of Remark 4.20). Given the chromatic spectral sequence, we can reduce to showing that $H^r(\mathbb{G}_{k+s}, (E_{k+s})_*/I)$ is finite, with I as in the proof of the previous proposition. For this, see Proposition 4.2.2 of [54]. □

Corollary 4.23 *Let M_k denote a generalized Moore spectrum of type k. Then if $p-1$ does not divide $k+s$ for $0 \le s \le n-k$,*

$$\mathrm{Ext}^{s,*}_{(E_n)_*(E_n)}((E_n)_*, (E_n)_*(M_k)) = 0$$

for $s > n^2+n-k$.

Proof Recall that $(E_n)_*(M_k) \cong (E_n)_*/(p^{i_0}, \dots, u_{k-1}^{i_{k-1}})$ for a suitable sequence of integers $(i_0, \dots, i_{k-1})$. The result for the sequence $(1, \dots, 1)$ holds by Proposition 4.19, and therefore in general by taking appropriate exact sequences. □

Theorem 4.24 *Suppose $p-1$ does not divide $k+s$ for $0 \le s \le n-k$. Then*

$$\widehat{\mathrm{Ext}}^{s,t}_{(E_{k,n})^\vee_*(E_n)}((E_n)_*, (E_n)_*) = 0$$

for $s > n^2+n-k$.

Proof Combine Proposition 4.16 and Corollary 4.23. □

Remark 4.25 The condition on the prime is always satisfied if p is large enough compared to n (in fact $p > n+1$ suffices). This suggests the following, which we do not attempt to make precise: for large enough primes, the cohomological dimension of $(E_n)_*$ in a suitable category of (completed) $(E_{k,n})^\vee_*(E_n)$–comodules is finite, and equal to n^2+n-k.

We also have the following expected sparseness result.

Proposition 4.26 *Let $q = 2(p-1)$. Then*

$$\widehat{\mathrm{Ext}}^{s,t}_{(E_n)^\vee_*(E_n)}((E_n)_*, (E_n)_*) = 0$$

for all s and t unless $t \equiv 0 \mod q$. Consequently, in the spectral sequence of Proposition 4.13, d_r is nontrivial only if $r \equiv 1 \mod q$ and $E^{*,*}_{mq+2} = E^{*,*}_{mq+q+1}$ for all $m \geq 0$.

Proof Using Proposition 4.16 it suffices to show the first statement for the E_1–term of the chromatic spectral sequence of Proposition 4.19. Again using the Hovey–Sadofsky change of rings theorem, this E_1–term is isomorphic to

$$\mathrm{Ext}^{s,*,*}_{BP_*BP}\big(BP_*, u^{-1}_{k+s}BP_*/(p, \dots, u_{k-1}, u^\infty_k, \dots, u^\infty_{k+s-1})\big).$$

Now apply [47, Proposition 4.4.2]. □

5 Dualizable objects in $\mathrm{Sp}_{k,n}$

In this section we use descendability to characterize the dualizable objects in $\mathrm{Sp}_{k,n}$. As noted previously, as long as $k \neq 0$, these differ from the compact objects studied in Section 3A.

Definition 5.1 Let $(\mathcal{C}, \wedge, \mathbf{1})$ be a symmetric-monoidal ∞–category. Then $X \in \mathcal{C}$ is dualizable if there exists an object $D_{\mathcal{C}}X$ and a pair of morphisms

$$e\colon D_{\mathcal{C}}X \wedge X \to \mathbf{1}, \quad c\colon \mathbf{1} \to X \wedge D_{\mathcal{C}}X$$

such that the composites

$$X \xrightarrow{c\wedge \mathrm{id}} X \wedge D_{\mathcal{C}}X \wedge X \xrightarrow{\mathrm{id}\wedge e} X, \quad D_{\mathcal{C}}X \xrightarrow{\mathrm{id}\wedge c} D_{\mathcal{C}}X \wedge X \wedge D_{\mathcal{C}}X \xrightarrow{e\wedge \mathrm{id}} D_{\mathcal{C}}X$$

are the identity on X and $D_{\mathcal{C}}X$, respectively.

Remark 5.2 The definition makes it clear that $X \in \mathcal{C}$ is dualizable if and only if it is dualizable in the homotopy category of $\mathcal{C}$. Moreover, a formal argument shows that, if it exists, we must have $D_{\mathcal{C}}X \simeq F(X, \mathbf{1})$. Finally, for the equivalence with other definitions of dualizability the reader may have seen, see [17, Theorem 1.3].

Definition 5.3 We let $\mathcal{C}^{\mathrm{dual}} \subseteq \mathcal{C}$ denote the full subcategory consisting of the dualizable objects of $\mathcal{C}$.

Remark 5.4 The full subcategory $\mathcal{C}^{\mathrm{dual}}$ is a thick tensor ideal [30, Theorem A.2.5].

We have the following relationship between descent theory and dualizability.

Proposition 5.5 *Let $A \in \mathrm{CAlg}(\mathcal{C})$ be descendable. Then the adjunction $C \leftrightarrows \mathrm{Mod}_A(\mathcal{C})$ gives rise to an equivalence of symmetric monoidal ∞–categories*

$$\mathcal{C}^{\mathrm{dual}} \to \mathrm{Tot}\big(\mathrm{Mod}_A(\mathcal{C})^{\mathrm{dual}} \rightrightarrows \mathrm{Mod}_{A\wedge A}(\mathcal{C})^{\mathrm{dual}} \substack{\rightarrow\\\rightarrow\\\rightarrow} \big).$$

In particular, $M \in \mathcal{C}$ is dualizable if and only if $M \wedge A \in \mathrm{Mod}_A(\mathcal{C})$ is dualizable.

Proof The first claim follows from Proposition 4.4 because passing to dualizable objects commutes with limits of ∞–categories [40, Proposition 4.6.1.11]. The second is then an easy consequence, using that all the maps in the totalization are symmetric monoidal. ◻

5A Dualizable objects in the $E(n, J_k)$–local category

Using Theorem 4.6 and Proposition 5.5 we deduce the following.

Proposition 5.6 *The adjunction $\mathrm{Sp}_{k,n} \leftrightarrows \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ gives rise to an equivalence of symmetric monoidal ∞–categories*

$$\mathrm{Sp}_{k,n}^{\mathrm{dual}} \to \mathrm{Tot}\big(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})^{\mathrm{dual}} \rightrightarrows \mathrm{Mod}_{E_n \overline{\wedge} E_n}(\mathrm{Sp}_{k,n})^{\mathrm{dual}} \substack{\rightarrow\\\rightarrow\\\rightarrow} \big).$$

In particular, $X \in \mathrm{Sp}_{k,n}$ is dualizable if and only if $E_n \mathbin{\overline{\wedge}} X \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ is dualizable.

This proposition suggests we begin by studying dualizable objects in the category $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$. Fortunately, these have a nice characterization. We begin with the following.

Lemma 5.7 *If X is dualizable in $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ then the spectrum underlying X is $K(n)$–local.*

Proof We first note that for any $M \in \mathrm{Mod}_{E_n}$ (in particular, for $M = X$), the Bousfield localization $L_{K(n)}M$ is the spectrum underlying $L^{E_n}_{E_n \wedge X}$, where the latter denotes the Bousfield localization with respect to $E_n \wedge X$ internal to the category of E_n–modules. In particular, the localization map $M \to L_{K(n)}M$ is a map in Mod_{E_n}; see [19, Chapter VIII], particularly [19, Proposition VIII.1.8]. If follows that $K(n)$–localization defines a localization

$$L_{K(n)} \colon \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}) \to \mathrm{Mod}_{E_n}(\mathrm{Sp}_{K(n)}).$$

Because $X \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})^{\mathrm{dual}}$, using [30, Lemma 3.3.1], we see that there are equivalences

$$L_{K(n)}X \simeq L_{k,n}((L_{K(n)}E_n) \wedge_{E_n} X) \simeq L_{k,n}(E_n \wedge_{E_n} X) \simeq L_{k,n}X \simeq X. \qquad \square$$

Remark 5.8 For the following, we let $K_n \cong E_n/I_n$. This is a 2–periodic form of Morava K–theory; indeed,

$$(K_n)_*X \cong (K_n)_* \otimes_{K(n)_*} K(n)_*X,$$

and so $\langle K(n)\rangle = \langle K_n\rangle$. We use this only because K_n is naturally an E_n–module.

Proposition 5.9 *For $X \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ the following are equivalent:*

(1) *X is dualizable in $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$.*

(2) *X is compact (equivalently, dualizable) in $\mathrm{Mod}_{E_n}(\mathrm{Sp})$.*

(3) *The spectrum underlying X is $K(n)$–local and the homotopy groups $\pi_*(K_n \wedge_{E_n} X)$ are finite.*

Proof We first show that (2) implies (1). The compact objects in $\mathrm{Mod}_{E_n}(\mathrm{Sp})$ are precisely those in the thick subcategory generated by E_n; see, for example, [40, Proposition 7.2.4.2]. Since $E_n \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})^{\mathrm{dual}}$, and the collection of dualizable objects is thick, the implication (2) implies (1) follows.

Conversely, assume that (1) holds. As above, we have a symmetric-monoidal localization

$$L_{K(n)}\colon \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}) \to \mathrm{Mod}_{E_n}(\mathrm{Sp}_{K(n)}),$$

which preserves dualizable objects (as any symmetric-monoidal functor does). Using Lemma 5.7 it follows that $L_{K(n)}X \simeq X$ is dualizable in $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{K(n)})$, which implies by [41, Proposition 10.11] that X is compact in $\mathrm{Mod}_{E_n}(\mathrm{Sp})$.

Finally, the equivalence of (2) and (3) is well known; see for example Proposition 2.9.4 of [25].[4] $\square$

Remark 5.10 Suppose $X \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$, so that $L_{k,n}(E_n \wedge X) \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})^{\mathrm{dual}}$. The previous proposition then implies that

$$L_{k,n}(E_n \wedge X) \simeq L_{K(n)}L_{k,n}(E_n \wedge X) \simeq L_{K(n)}(E_n \wedge X).$$

In other words, for dualizable X, there is an isomorphism $(E_{k,n})^\vee_*(X) \cong (E_{n,n})^\vee_*(X)$.

[4] Lurie has confirmed via private communication that the cited proposition [25, Proposition 2.9.4] should additionally have the assumption X is $K(n)$–local in condition (3).

We now give our characterization of dualizable spectra in $\mathrm{Sp}_{k,n}$ — see [32, Theorem 8.6] for the case $k=n$. We note that even in this case our proof, which uses descendability, differs from that of Hovey and Strickland.

Theorem 5.11 *The following are equivalent for $X \in \mathrm{Sp}_{k,n}$:*

(1) *X is dualizable.*

(2) *X is F–small, ie for any collection of objects $\{Z_i\}$, the natural map*
$$L_{k,n}\Big(\bigvee_i F(X,Z_i)\Big) \to F\Big(X, L_{k,n}\Big(\bigvee_i Z_i\Big)\Big)$$
is an equivalence.

(3) *$E_n \mathbin{\overline{\wedge}} X \in \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$ is dualizable.*

(4) *$E_n \mathbin{\overline{\wedge}} X \in \mathrm{Mod}_{E_n}(\mathrm{Sp})$ is dualizable (equivalently, compact).*

(5) *$E_n \mathbin{\overline{\wedge}} X$ is $K(n)$–local and $(K_n)_* X$ is finite.*

(6) *$(E_{k,n})^\vee_*(X)$ is a finitely generated E_*–module.*

Proof The equivalences between the first five items come from [30, Theorem 2.1.3(c)] ((1) $\iff$ (2)), Proposition 5.5 ((1) $\iff$ (3)) and Proposition 5.9 ((3) $\iff$ (4) $\iff$ (5)). We note that if M is an E_n–module, then M is compact if and only if $\pi_* M$ is finitely generated over $(E_n)_*$; see [22, Lemma 10.2(i)]. Applying this with $M = E_n \mathbin{\overline{\wedge}} X$ gives the equivalence between (4) and (6). □

Finally, we show that there is only a set of isomorphism classes of dualizable objects.

Lemma 5.12 *There are at most $2^{\aleph_0}$ isomorphism classes of objects in $\mathrm{Sp}^{\mathrm{dual}}_{k,n}$.*

Proof This is the same as the argument given in [32, Propositon 12.17]. Namely, there are only countably many finite spectra X' of type at least k, and for each one $[L_n X', L_n X']$ is finite, so $L_n X'$ has only finitely many retracts. By Theorem 3.8 it follows that there is a countable set of isomorphism classes of objects in $\mathrm{Sp}^{\omega}_{k,n}$. If U and V are finite, then $[U,V]$ is finite, and so there are at most $\aleph_0^{\aleph_0} = 2^{\aleph_0}$ different towers of spectra in $\mathrm{Sp}^{\omega}_{k,n}$. For $X \in \mathrm{Sp}^{\mathrm{dual}}_{k,n}$, write $X \simeq \varprojlim_j X \wedge M_k(j)$, as in Proposition 2.24. Because X is dualizable and $M_k(j)$ is compact, $X \wedge M_k(j)$ is compact [30, Theorem 2.1.3(a)]. Therefore, X is the inverse limit of a tower of spectra in $\mathrm{Sp}^{\omega}_{k,n}$, and hence there are at most $2^{\aleph_0}$ isomorphism classes of objects in $\mathrm{Sp}^{\mathrm{dual}}_{k,n}$. □

5B The spectrum of dualizable objects

In Theorem 3.16 we computed the thick subcategories (equivalently, thick tensor-ideals) of compact objects in $\mathrm{Sp}_{k,n}$. One could also ask for a classification of the thick tensor-ideals of dualizable objects in $\mathrm{Sp}_{k,n}$, or equivalently a computation of the Balmer spectrum $\mathrm{Spc}(\mathrm{Sp}_{k,n}^{\mathrm{dual}})$ (which is well defined by Lemma 5.12). Based on a conjecture of Hovey and Strickland, the author, along with Barthel and Naumann, investigated $\mathrm{Spc}(\mathrm{Sp}_{K(n)}^{\mathrm{dual}})$ in detail in [10], showing that the Hovey–Strickland conjecture holds when $n = 1$ and 2, and that in general it is implied by a hope of Chai in arithmetic geometry. In this section, we make some general comments regarding $\mathrm{Spc}(\mathrm{Sp}_{k,n}^{\mathrm{dual}})$.

Remark 5.13 The following full subcategories were considered in the case $k = n$ by Hovey and Strickland [32, Definition 12.14].

Definition 5.14 For $i \le n$, let $\mathcal{D}_i$ denote the category of spectra $X \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ such that X is a retract of $Y \wedge Z$ for some $Y \in \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ and some finite spectrum Z of type at least i. It is also useful to set $\mathcal{D}_{n+1} = (0)$.

Remark 5.15 We note that $\mathcal{D}_k \simeq \mathrm{Sp}_{k,n}^{\omega}$; this is a consequence of the characterization of compact objects given in Theorem 3.8, and that $\mathcal{D}_0 = \mathrm{Sp}_{k,n}^{\mathrm{dual}}$.

The following is [32, Proposition 4.17].

Lemma 5.16 *X is in $\mathcal{D}_k$ if and only if X is a module over a generalized Moore spectrum of type k. Moreover, $\mathcal{D}_k \subseteq \mathrm{Sp}_{k,n}^{\mathrm{dual}}$ is a thick tensor ideal.*

Hovey and Strickland conjecture that in the case $k = n$ these exhaust the thick-tensor ideals of $\mathrm{Sp}_{K(n)}^{\mathrm{dual}}$. This has been investigated in detail in [10]. We conjecture this holds more generally in $\mathrm{Sp}_{k,n}$.

Conjecture 5.17 *If $\mathcal{C}$ is a thick tensor-ideal of $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$, then $\mathcal{C} = \mathcal{D}_i$ for some $0 \le i \le n+1$. Equivalently,*

$$\mathrm{Spc}(\mathrm{Sp}_{k,n}^{\mathrm{dual}}) = \{\mathcal{D}_1, \dots, \mathcal{D}_{n+1}\}$$

with topology determined by the closure operator $\overline{\{\mathcal{D}_i\}} = \{\mathcal{D}_j \mid j \ge i\}$.

In this section we show that if Conjecture 5.17 holds $K(n)$–locally, ie for $\mathrm{Sp}_{n,n}^{\mathrm{dual}}$, then it holds for all $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$. We first recall the following definition.

Definition 5.18 Suppose $F\colon \mathcal{K} \to \mathcal{L}$ is an exact tensor triangulated functor between tensor-triangulated categories. We say that F detects tensor-nilpotence of morphisms if every morphism $f\colon X \to Y$ in $\mathcal{K}$ such that $F(f) = 0$ satisfies $f^{\otimes m} = 0$ for some $m \geq 1$.

We will use the following.

Proposition 5.19 *Suppose $A \in \mathrm{CAlg}(\mathcal{C})$ is descendable. Then extension of scalars $\mathcal{C} \to \mathrm{Mod}_A(\mathcal{C})$ detects tensor-nilpotence of morphisms.*

Proof Let I denote the fiber of $\mathbf{1} \xrightarrow{\eta} A$, and let $\xi\colon I \to \mathbf{1}$ denote the induced map. By [43, Proposition 4.7] if A is descendable, then there exists $m \geq 1$ such that $I^{\otimes m} \to \mathbf{1}$ is null-homotopic, ie ξ is tensor-nilpotent. We can now argue as in (ii) implies (iii) of [5]: suppose we are given $f\colon X \to Y$, a morphism in $\mathcal{C}$, with $A \otimes f\colon A \otimes X \to A \otimes Y$ null-homotopic. Now consider the diagram of fiber sequences:

$$\begin{tikzcd} I \otimes X \arrow[r, "\xi \otimes \mathrm{id}_X"] \arrow[d, "\mathrm{id}_I \otimes f"'] & X \arrow[r, "\eta \otimes \mathrm{id}_X"] \arrow[d, "f"] & A \otimes X \arrow[d, "\mathrm{id}_A \otimes f"] \\ I \otimes Y \arrow[r, "\xi \otimes \mathrm{id}_Y"'] & Y \arrow[r, "\eta \otimes \mathrm{id}_Y"'] & A \otimes Y \end{tikzcd}$$

We see that $(\eta \otimes \mathrm{id}_Y) f$ is null-homotopic, so f factors through $\xi \otimes \mathrm{id}_Y$, which is tensor-nilpotent. $\square$

The following is our key observation.

Proposition 5.20 *If $i > k$, then the map induced by localization*

$$\mathrm{Spc}(\mathrm{Sp}_{i,n}^{\mathrm{dual}}) \to \mathrm{Spc}(\mathrm{Sp}_{k,n}^{\mathrm{dual}})$$

is surjective.

Proof By [6, Theorem 1.3] it suffices to show that the functor $L_{i,n}\colon \mathrm{Sp}_{k,n}^{\mathrm{dual}} \to \mathrm{Sp}_{i,n}^{\mathrm{dual}}$ detects tensor-nilpotence of morphisms. To that end, let $f\colon X \to Y$ be a morphism in $\mathrm{Sp}_{k,n}^{\mathrm{dual}}$ with $L_{i,n}(f) = 0$, so that we must show $f^{\overline{\wedge} m} = 0$ for some $m \geq 1$. Because $E_n \in \mathrm{Sp}_{k,n}$ is descendable, Proposition 5.19 shows that

$$L_{k,n}(E_n \wedge -)\colon \mathrm{Sp}_{k,n} \to \mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$$

detects tensor-nilpotence of morphisms, and hence so does its restriction to dualizable objects; ie $L_{k,n}(E_n \wedge f) = 0$ implies $f^{\overline{\wedge} m} = 0$ for some $m \geq 1$. In other words, it suffices to show that $L_{k,n}(E_n \wedge f)$ is trivial. By Lemma 5.7 however, this is a morphism in $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{n,n})$. In particular,

$$L_{k,n}(E_n \wedge f) \simeq L_{i,n}(E_n \wedge f) \simeq L_{i,n}(E_n \wedge L_{i,n}(f)) = 0$$

because $L_{i,n}(f) = 0$ by assumption. $\square$

Theorem 5.21 *Suppose Conjecture 5.17 holds for $\mathrm{Sp}^{\mathrm{dual}}_{n,n}$. Then it holds for all $\mathrm{Sp}^{\mathrm{dual}}_{k,n}$.*

Proof By [10, Proposition 3.5], Conjecture 5.17 holds for $\mathrm{Sp}^{\mathrm{dual}}_{n,n}$ if and only if $L_{n,n}\colon \mathrm{Sp}^{\mathrm{dual}}_{0,n} \to \mathrm{Sp}^{\mathrm{dual}}_{n,n}$ induces a homeomorphism on Balmer spectra. In other words, the composite, induced by the localization maps,

$$\mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{n,n}) \to \mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{n-1,n}) \to \cdots \to \mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{0,n})$$

is a homeomorphism. It follows that $\mathrm{Spc}(L_{n-1,n})\colon \mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{n,n}) \to \mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{n-1,n})$ is an injection, and hence a bijection by Proposition 5.20. Using that $\mathrm{Spc}(L_{n-1,n})$ is continuous and the topologies on each space, we see that it is fact a homeomorphism. It follows that $\mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{n-1,n}) \to \mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{0,n})$ is a homeomorphism, and we can now repeat the argument. $\square$

By [10, Theorem 4.15], Conjecture 5.17 holds for $\mathrm{Sp}^{\mathrm{dual}}_{2,2}$. Along with Theorem 5.21 we deduce the following.

Corollary 5.22 *The Balmer spectrum $\mathrm{Spc}(\mathrm{Sp}^{\mathrm{dual}}_{1,2}) = \{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$ with*

$$(0) = \mathcal{D}_3 \subsetneq \mathcal{D}_2 \subsetneq \mathcal{D}_1 = \mathrm{Sp}^{\omega}_{1,2}.$$

In particular, if $\mathcal{C}$ is a thick tensor-ideal of $\mathrm{Sp}^{\mathrm{dual}}_{1,2}$, then $\mathcal{C} = \mathcal{D}_k$ for $0 \le k \le 3$.

6 The Picard group of the $\mathrm{Sp}_{k,n}$–local category

In this section we study invertible objects in the $\mathrm{Sp}_{k,n}$–local category. We show that invertibility of an object can be detected by its Morava module. We construct a spectral sequence computing the homotopy groups of the Picard spectrum of $\mathrm{Sp}_{k,n}$ and use this to show that if p is large compared to n, then the Picard group of $\mathrm{Sp}_{k,n}$ is entirely algebraic, in a sense we make precise.

6A Invertible objects and Picard spectra

We recall that if $\mathcal{C}$ is a symmetric monoidal category, we denote by $\mathrm{Pic}(\mathcal{C})$ the group of isomorphism classes of invertible objects; a priori this could be a proper class, but if

$\mathcal{C}$ is a presentable stable ∞–category (which it will always be in our cases), then it is actually a set [44, Remark 2.1.4].

The following standard lemma will be useful for us. Here we write $D_{\mathcal{C}}(X)$ for the dual of an object in a category $\mathcal{C}$, ie $D_{\mathcal{C}}(X) = F(X, \mathbf{1})$. Note that an invertible object is always dualizable [30, Proposition A.2.8].

Lemma 6.1 *Let $F : \mathcal{C} \to \mathcal{D}$ be a symmetric-monoidal conservative functor between stable ∞–categories. Then $X \in \mathcal{C}$ is invertible if and only if $F(X) \in \mathcal{D}$ is invertible.*

Proof We first note that X is invertible if and only if the natural morphism

$$X \otimes_{\mathcal{C}} D_{\mathcal{C}}(X) \to \mathbf{1}_{\mathcal{C}}$$

is an equivalence; see [30, Proposition A.2.8]. Because F is assumed to be symmetric monoidal and conservative, this is an equivalence if and only if it is so after applying F; ie if and only if $F(X) \otimes_{\mathcal{D}} F(D_{\mathcal{C}}(X)) \to \mathbf{1}_{\mathcal{D}}$ is an equivalence. But $F(D_{\mathcal{C}}(X)) \simeq D_{\mathcal{D}}(F(X))$, as symmetric-monoidal functors preserve dualizable objects, and the result follows. □

Remark 6.2 To our symmetric monoidal category $\mathcal{C}$ we can instead associate the Picard spectrum $\mathfrak{pic}(\mathcal{C})$ [44, Definition 2.2.1]; this is a connective spectrum with the property that

$$\pi_i(\mathfrak{pic}(\mathcal{C})) = \begin{cases} \mathrm{Pic}(\mathcal{C}) & \text{if } i = 0, \\ (\pi_0(\mathrm{End}_{\mathcal{C}}(\mathbf{1}))^{\times} & \text{if } i = 1, \\ \pi_{i-1}(\mathrm{End}_{\mathcal{C}}(\mathbf{1})) & \text{if } i > 1. \end{cases}$$

The key advantage of using the Picard spectrum is that, as a functor from the ∞–category of symmetric monoidal ∞–categories to the ∞–category of connective spectra, $\mathfrak{pic}$ commutes with limits [44, Proposition 2.2.3].

Example 6.3 Let $\mathcal{C}$ be a category and $A \in \mathrm{CAlg}(\mathcal{C})$. Then the Picard spectrum of the category $\mathrm{Mod}_A(\mathcal{C})$ of A–modules internal to $\mathcal{C}$ satisfies

$$\pi_i(\mathfrak{pic}(\mathrm{Mod}_A(\mathcal{C})) = \begin{cases} \mathrm{Pic}(\mathrm{Mod}_A(\mathcal{C})) & \text{if } i = 0, \\ (\pi_0(\mathrm{Hom}_{\mathcal{C}}(\mathbf{1}_{\mathcal{C}}, A))^{\times} & \text{if } i = 1, \\ \pi_{i-1}(\mathrm{Hom}_{\mathcal{C}}(\mathbf{1}_{\mathcal{C}}, A)) & \text{if } i > 1. \end{cases}$$

This follows because A is the tensor unit in $\mathrm{Mod}_A(\mathcal{C})$. Indeed, writing $F : \mathcal{C} \to \mathrm{Mod}_{\mathcal{C}}(A)$ for the extension of scalars functor (so that $A \simeq F(\mathbf{1}_{\mathcal{C}})$), we have

$$\mathrm{End}_{\mathrm{Mod}_A(\mathcal{C})}(A) = \mathrm{Hom}_{\mathrm{Mod}_A(\mathcal{C})}(F(\mathbf{1}_{\mathcal{C}}), A) \simeq \mathrm{Hom}_{\mathcal{C}}(\mathbf{1}_{\mathcal{C}}, A)$$

by adjunction.

6B Invertible objects in the $\mathbf{Sp}_{k,n}$–local category

Our main group of interest is the Picard group of $K(k)\vee\cdots\vee K(n)$–local spectra.

Definition 6.4 Let $\mathrm{Pic}_{k,n}=\mathrm{Pic}(\mathrm{Sp}_{k,n})$, the group of invertible $K(k)\vee\cdots\vee K(n)$–local spectra.

Remark 6.5 By [36, Lemma 2.2] the localization functors induce natural morphisms $\mathrm{Pic}_{0,n}\to\mathrm{Pic}_{1,n}\to\cdots\to\mathrm{Pic}_{n,n}$.

Remark 6.6 The morphism $X\mapsto E_n\,\overline{\wedge}\,X$ induces a functor

$$\mathrm{Pic}_{k,n}\to\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})).$$

We can fully understand the latter Picard group.

Lemma 6.7 *For all* $0\le k\le n$,

$$\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}))\cong\mathrm{Pic}(\mathrm{Mod}_{E_n})\cong\mathrm{Pic}(E_{n*})\cong\mathbb{Z}/2.$$

Proof We always have $\mathrm{Pic}(\mathrm{Mod}_{E_n})\subseteq\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}))$ because any invertible E_n–module is compact, and hence $E(n,J_k)$–local. The other inclusion follows if any $M\in\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}))$ is compact as an E_n–module. Such an M is automatically dualizable in $\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})$, and hence compact in Mod_{E_n} by Proposition 5.9. This gives the first of the above isomorphisms, and the others hold by work of Baker and Richter [2]. $\square$

We now give criteria for when $X\in\mathrm{Pic}_{k,n}$ is invertible. This (partially) extends work of Hopkins, Mahowald and Sadofsky [26], who considered the case $k=n$.

Theorem 6.8 *Let* $X\in\mathrm{Sp}_{k,n}$. *The following are equivalent*:

1. $X\in\mathrm{Pic}_{k,n}$.
2. $E_n\,\overline{\wedge}\,X\in\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}))$.
3. $E_n\,\overline{\wedge}\,X\in\mathrm{Pic}(\mathrm{Mod}_{E_n})$.
4. $(E_{k,n})^\vee_*X\cong(E_n)_*$, *up to suspension.*

Proof The equivalence of (1) and (2) follows from Corollary 4.7 and Lemma 6.1, while the equivalence of (2) and (3) follows from Lemma 6.7, which also shows that

(3) implies (4). Finally, to see that (4) implies (3), we note that if M is any E_n–module whose homotopy groups are free of rank one over $(E_n)_*$, then M is equivalent to a suspension of E_n; for the elementary proof, see [23, Proposition 2.2]. Thus, (4) implies that $E_n \mathbin{\overline{\wedge}} X \simeq E_n$, up to suspension, and hence (3) holds. □

Remark 6.9 When $n = 1$, there are two possibilities, the $K(1)$ and $E(1)$–local Picard groups, both of which are known:

$$\mathrm{Pic}_{0,1} = \begin{cases} \mathbb{Z} \oplus \mathbb{Z}/2 & \text{if } p = 2, \\ \mathbb{Z} & \text{if } p > 2, \end{cases} \qquad \mathrm{Pic}_{1,1} = \begin{cases} \mathbb{Z}_2 \oplus \mathbb{Z}/4 \oplus \mathbb{Z}/2 & \text{if } p = 2, \\ \mathbb{Z}_p \oplus \mathbb{Z}/(p-1) \oplus \mathbb{Z}/2 & \text{if } p > 2. \end{cases}$$

These are due to Hovey and Sadofsky [31] and Hopkins, Mahowald and Sadofsky [26], respectively.

When $n = 2$, we have three possibilities, the $K(2)$, $K(1) \vee K(2)$, and $E(2)$–local Picard groups. The first and last are known for $p > 2$:

$$\mathrm{Pic}_{0,2} = \begin{cases} \mathbb{Z} \oplus \mathbb{Z}/3 \oplus \mathbb{Z}/3 & \text{if } p = 3, \\ \mathbb{Z} & \text{if } p > 3, \end{cases}$$
$$\mathrm{Pic}_{2,2} = \begin{cases} \mathbb{Z}_3 \oplus \mathbb{Z}_3 \oplus \mathbb{Z}/16 \oplus \mathbb{Z}/3 \oplus \mathbb{Z}/3 & \text{if } p = 3, \\ \mathbb{Z}_p \oplus \mathbb{Z}_p \oplus \mathbb{Z}/(2(p^2-1)) & \text{if } p > 3. \end{cases}$$

These are due to a combination of authors: Hovey and Sadofsky [31], Lader [38], Goerss, Henn, Mahowald and Rezk [20], Karamanov [37], and Hopkins (unpublished).

This leaves the remaining case of $\mathrm{Pic}_{1,1}$. We note the following.

Proposition 6.10 *If $p \geq 3$, then the morphism $\mathrm{Pic}_{0,2} \to \mathrm{Pic}_{1,1}$ of Remark 6.5 is an injection.*

Proof The morphism in question factors through the morphism $\mathrm{Pic}_{0,2} \to \mathrm{Pic}_{2,2}$ and so it suffices to show that this is an injection. When $p > 3$ this is clear, and so we focus on the case $p = 3$. In this case, the calculations of Goerss, Henn, Mahowald and Rezk [20] show that this map is an injection. □

Remark 6.11 As noted in the proof, the interesting case in the above proposition is the case $p = 3$. In fact, for all n and $p \gg n$ we have that $\mathrm{Pic}_{0,n} \to \mathrm{Pic}_{i,n}$ is an injection for $i \geq 0$. However, here $\mathrm{Pic}_{0,n} \cong \mathbb{Z}$ (by [31]), so this is not particularly helpful.

6C Descent and Picard groups

In Remark 6.2 we recalled that we can associate a connective Picard spectrum $\mathfrak{pic}(\mathcal{C})$ to a symmetric-monoidal ∞–category $\mathcal{C}$. Using descent for the $E(n, J_k)$–local category, we now construct a spectral sequence whose π_0 computes $\mathrm{Pic}_{k,n}$. We need to introduce another algebraic gadget to describe the spectral sequence.

Definition 6.12 We let $\mathrm{Pic}^{\mathrm{alg}}_{k,n}$ denote the first cohomology of the complex

$$(E_n)_0^\times \rightrightarrows ((E_{k,n})_0^\vee(E_n))^\times \overrightarrow{\rightrightarrows} \cdots$$

induced by taking the units in degree 0 in the cobar complex.

Theorem 6.13 *There exists a spectral sequence with*

$$E_2^{s,t} \cong \begin{cases} \mathbb{Z}/2 & \text{if } s=t=0, \\ \mathrm{Pic}^{\mathrm{alg}}_{k,n} & \text{if } s=t=1, \\ \widehat{\mathrm{Ext}}^{s,t-1}_{(E_{k,n})_*^\vee(E_n)}((E_n)_*,(E_n)_*) & t\geq 2, \end{cases}$$

which converges for $t-s\geq 0$ *to* $\pi_{t-s}\mathfrak{pic}(\mathrm{Sp}_{k,n})$*. In particular, when* $t=s$*, this computes* $\mathrm{Pic}_{k,n}$*. The differentials in the spectral sequence run* $d_r\colon E_r^{s,t}\to E_r^{s+r,t+r-1}$*.*

Proof Because $\mathfrak{pic}$ commutes with limits (Remark 6.2), Theorem 4.6 implies that

$$\mathfrak{pic}(\mathrm{Sp}_{k,n}) \simeq \tau_{\geq 0}\,\mathrm{Tot}\big(\,\mathfrak{pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})) \rightrightarrows \mathfrak{pic}(\mathrm{Mod}_{E_n\barwedge E_n}(\mathrm{Sp}_{k,n})) \overrightarrow{\rightrightarrows}\,\big). \tag{6-1}$$

We have (compare Example 6.3)

$$\pi_t(\mathfrak{pic}(\mathrm{Mod}_{E_n^{\barwedge i}}(\mathrm{Sp}_{k,n})) \cong \begin{cases} \mathrm{Pic}(\mathrm{Mod}_{E_n^{\barwedge i}}) & \text{if } t=0, \\ \pi_0(E_n^{\barwedge i})^\times & \text{if } t=1, \\ \pi_{t-1}(E_n^{\barwedge i}) & \text{if } t\geq 2. \end{cases} \tag{6-2}$$

The Bousfield–Kan spectral sequence associated to (6-1) has the form

$$E_2^{s,t} \cong H^s\big(\,\pi_t\mathfrak{pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n})) \rightrightarrows \pi_t\mathfrak{pic}(\mathrm{Mod}_{E_n\barwedge E_n}(\mathrm{Sp}_{k,n})) \overrightarrow{\rightrightarrows}\,\big).$$

By (6-2) when $t\geq 2$, the spectral sequence is just a shift of the $E(n,J_k)$–local Adams spectral for $X=S^0$ sequence considered in Proposition 4.13.

When $t=0$ and $i=0$, we have $\mathrm{Pic}(\mathrm{Mod}_{E_n}(\mathrm{Sp}_{k,n}))\cong\mathbb{Z}/2$ by Lemma 6.7. We do not know the higher terms, but this does not matter as only the $\mathbb{Z}/2$ is relevant for the $s=t=0$ part of the spectral sequence.

Finally, we consider the $t=1$ part of the spectral sequence. Again using (6-2),

$$E_2^{s,1} \cong H^s\big((E_n)_0^\times \rightrightarrows ((E_{k,n})_0^\vee(E_n))^\times \overrightarrow{\rightrightarrows}\cdots\big).$$

By definition, when $s=1$ this is $\mathrm{Pic}^{\mathrm{alg}}_{k,n}$. □

Remark 6.14 The proof shows that when $t = 1$, we can compute $E_2^{s,1}$ as the s^{th} cohomology of the complex in Definition 6.12. However, unless $k = 0$ or n we do not have a convenient description of this group (for the case $k = n$, see Example 6.18 below).

Definition 6.15 We will say that $\mathrm{Pic}_{k,n}$ is algebraic if the only contributions come from the $s = 0$ and $s = 1$ lines of the spectral sequence.

Remark 6.16 The $E_2^{0,0}$ term of the spectral sequence always survives the spectral sequence, as it is the Picard group of E_n–modules. It is however possible that there is a nontrivial differential in the $E_r^{1,1}$ spot.

Theorem 6.17 *Suppose that $2p-2 \geq n^2+n-k$ and $p-1$ does not divide $k+s$ for $0 \leq s \leq n-k$. Then $\mathrm{Pic}_{k,n}$ is algebraic. For example, this holds if $2p-2 > n^2+n$.*

Proof For all primes p and $t \geq 2$ we have that $E_2^{s,t} = 0$ unless $t-1 \equiv 0 \bmod 2(p-1)$ by Proposition 4.26. In particular, for $s > 2$, $E_2^{s,s} = 0$ unless $s \equiv 1 \bmod 2p-2$, and the lowest possible nonalgebraic term is in filtration degree $2p-1$.

By Theorem 4.24 and the assumption that $p-1$ does not divide $k+s$ for $0 \leq s \leq n-k$ we have that $E_2^{s,s} = 0$ for $s > n^2+n-k$. Therefore, if additionally $2p-2 \geq n^2+n-k$, there can be no nonalgebraic contributions to the spectral sequence.

Finally, if $2p-2 > n^2+n$, then $p > n+1$, and in particular $p-1$ does not divide $k+s$ for $0 \leq s \leq n-k$. ◻

Example 6.18 Let us spell out the details in the case $k = n$. We first claim that the spectral sequence of Theorem 6.13 takes the form

$$E_2^{s,t} \cong \begin{cases} \mathbb{Z}/2 & \text{if } s = t = 0, \\ H_c^s(\mathbb{G}_n, (E_n)_0^\times) & \text{if } t = 1, \\ H_c^s(\mathbb{G}_n, \pi_{t-1}E_n) & \text{if } t \geq 2, \end{cases}$$

and converges for $t-s \geq 0$ to $\pi_{t-s}\mathfrak{pic}(\mathrm{Sp}_{K(n)})$.

The identification is much as in Remark 4.14. For the $t = 1$ term, we note that $\pi_0(E_n^{\overline{\wedge} i})^\times \cong \mathrm{Hom}^c(\mathbb{G}_n^{\times(i-1)}, (E_n)_0)^\times \cong \mathrm{Hom}^c(\mathbb{G}_n^{\times(i-1)}, (E_n)_0^\times)$.

The existence of such a spectral sequence is folklore; see [18, Remark 6.10] or [45, Remark 2.6]. In fact, the latter also proves Theorem 6.17 in the case $k = n$.

7 $E(n, J_k)$–local Brown–Comenetz duality

We recall the classical definition of Brown–Comenetz duality. The group $\mathbb{Q}/\mathbb{Z}$ is an injective abelian group, and so the functor

$$X \mapsto \mathrm{Hom}(\pi_0 X, \mathbb{Q}/\mathbb{Z})$$

defines a cohomology theory on spectra represented by a spectrum $I_{\mathbb{Q}/\mathbb{Z}}$; this is the Brown–Comenetz dual of the sphere. The Brown–Comenetz dual of a spectrum X is then defined as $I_{\mathbb{Q}/\mathbb{Z}} X := F(X, I_{\mathbb{Q}/\mathbb{Z}})$, and satisfies

$$[Y, I_{\mathbb{Q}/\mathbb{Z}} X]_0 \cong \mathrm{Hom}(\pi_0(X \wedge Y), \mathbb{Q}/\mathbb{Z}).$$

It is an insight of Hopkins [24] that there is a good notion of Brown–Comenetz duality (also known as Gross–Hopkins duality) internal to the $K(n)$–local category, given by defining $I_n X = F(M_n X, I_{\mathbb{Q}/\mathbb{Z}})$ for a $K(n)$–local spectrum X. For details on this, see [52]. As we will see, this definition can also naturally be made in the $E(n, J_k)$–local category. We begin with the following generalization of a result of Stojanoska [50, Proposition 2.2]. We recall that, by definition, $M_{0,n} = L_n$. In this case, the following lemma just says that $F(L_n X, Y)$ is already L_n–local.

Proposition 7.1 *For any X and Y, the natural map $F(L_n X, Y) \to F(M_{k,n} X, Y)$ is an $E(n, J_k)$–localization.*

Proof We can repeat Stojanoska's argument. First, we show that $F(M_{k,n} X, Y)$ is $E(n, J_k)$–local. Indeed, let Z be $E(n, J_k)$–acyclic. Then we must show that

$$F(Z, F(M_{k,n} X, Y)) \simeq F(Z \wedge M_{k,n} X, Y) \simeq F(M_{k,n} Z \wedge X, Y)$$

is contractible. Here we have used that $M_{k,n}$ is smashing. But $M_{k,n} Z \simeq M_{k,n} L_{k,n} Z$ by Theorem 2.31 and this is contractible because Z is $E(n, J_k)$–acyclic.

We now show that the fiber $F(L_{k-1} X, Y)$ is $E(n, J_k)$–acyclic. By Proposition 2.15 it suffices by a localizing subcategory argument to show this after smashing with a generalized Moore spectrum $M(k)$ of type k. Then (up to suspension),

$$\begin{aligned} F(L_{k-1} X, Y) \wedge M(k) &\simeq F(L_{k-1} X, Y) \wedge DM(k) \\ &\simeq F(M(k), F(L_{k-1} X, Y)) \\ &\simeq F((L_{k-1} M(k)) \wedge X, Y) \simeq *. \end{aligned}$$

□

Definition 7.2 The $E(n, J_k)$–local Brown–Comenetz dual of X is

$$I_{k,n}X = I_{\mathbb{Q}/\mathbb{Z}}(M_{k,n}X) = F(M_{k,n}X, I_{\mathbb{Q}/\mathbb{Z}}).$$

We let $I_{k,n}$ denote the $E(n, J_k)$–local Brown–Comenetz dual of $L_{k,n}S^0$.

Remark 7.3 It does not matter if we ask that X be $E(n, J_k)$–local in the previous definition, as $I_{k,n}X$ only depends on the $E(n, J_k)$–localization of X. Indeed, we have equivalences

$$I_{k,n}X = F(M_{k,n}X, I_{\mathbb{Q}/\mathbb{Z}}) \simeq F(M_{k,n}L_{k,n}X, I_{\mathbb{Q}/\mathbb{Z}}) = I_{k,n}(L_{k,n}X).$$

In particular, $I_{k,n} = I_{k,n}(L_{k,n}S^0) \simeq I_{k,n}S^0$.

From the definition of $I_{\mathbb{Q}/\mathbb{Z}}$, we deduce the following.

Lemma 7.4 *There is a natural isomorphism*

$$[Y, I_{k,n}X]_0 \cong \mathrm{Hom}(\pi_0(M_{k,n}(X) \wedge Y), \mathbb{Q}/\mathbb{Z}).$$

As a consequence of Proposition 7.1 we deduce the following.

Lemma 7.5 *$I_{k,n}X$ is always $E(n, J_k)$–local. In fact, $I_{k,n}X \simeq L_{k,n}I_{\mathbb{Q}/\mathbb{Z}}(L_nX)$ and moreover, $I_{k,n}X \cong L_{k,n}I_{j,n}X$ for any $j \le k$.*

It follows that we have natural maps given by localization,

$$I_{0,n} \to I_{1,n} \to \cdots \to I_{n,n}.$$

Example 7.6 Let $n = 1$ and $p > 2$. Then $I_{0,1} \simeq L_1(S^2_p)$, the localization of the p–completion of S^2. On the other hand, when $p = 2$ we have $I_{0,1} \simeq \Sigma^2 L_1(DQ^\wedge_2)$ where DQ is the dual question mark complex [15, Remark 1.5]. Similarly, $I_{1,1} \simeq L_{K(1)}S^2$ if $p > 2$, while $I_{1,1} \simeq \Sigma^2 L_{K(1)}DQ$.

Example 7.7 We always have $I_{k,n}(K_n) \simeq K_n$. Indeed, first note that $L_{k,n}K_n \simeq K_n$, so Lemma 7.5 allows us to reduce the case where $k = 0$ (although the proof is no more difficult in the other cases—just note that $M_{k,n}K_n \simeq K_n$). Using the fact that

$$[Y, K_n]_* \cong \mathrm{Hom}_{(K_n)_*}((K_n)_*X, (K_n)_*),$$

we argue as in [32, Theorem 10.2(a)] to see that

$$[Y, K_n]_0 \cong \mathrm{Hom}((K_n)_0X, \mathbb{F}_p) \cong \mathrm{Hom}((K_n)_0X, \mathbb{Q}/\mathbb{Z}) \cong [Y, I_{0,n}(K_n)]_0.$$

This implies that $I_{0,n}(K_n) \simeq K_n$, as claimed.

Theorem 7.8 *Let $X \in \mathrm{Sp}_{k,n}$. Then the natural map $X \to I^2_{k,n}X$ is an isomorphism when $\pi_*(F(k) \wedge X) \cong \pi_*(L_{k,n}F(k) \wedge X)$ is finite in each degree. In particular, this holds for $X = L_{k,n}S^0$.*

Proof Let $\kappa_X \colon X \to I^2_{k,n}X$ denote the natural map. We first note that

$$I^2_{k,n}(F(k) \mathbin{\overline{\wedge}} X) \simeq I^2_{k,n}(F(k) \wedge X) \simeq F(k) \wedge I^2_{k,n}(X),$$

because $F(k)$ is compact (and hence dualizable) in Sp. As in the proof of Theorem 10.2 in [32], this identifies $\kappa_{F(k)\wedge X} \simeq \mathrm{id}_F \wedge \kappa_X$, and so it is enough to show that κ_Y is an equivalence, where $Y = F(k) \wedge X$.

Because $F(k)$ has type k, $L_{k-1}F(k) \simeq *$, and $M_{k,n}F(k) \simeq L_nF(k)$, so

$$M_{k,n}Y = M_{k,n}(F(k) \wedge X) \simeq M_{k,n}F(k) \wedge X \simeq L_nF(k) \wedge X \simeq Y.$$

Likewise, $M_{k,n}(I_{k,n}Y) \simeq M_{k,n}(DF(k) \wedge I_{k,n}X) \simeq DF(k) \wedge I_{k,n}X \simeq I_{k,n}Y$. This implies that $\pi_*I^2_{k,n}Y \cong \mathrm{Hom}(\mathrm{Hom}(\pi_*Y, \mathbb{Q}/\mathbb{Z}), \mathbb{Q}/\mathbb{Z})$, which is the same as π_*Y because π_*Y is finite in each degree. Therefore κ_Y is an equivalence, as required. □

Remark 7.9 The Gross–Hopkins dual $I_{n,n}$ is always an invertible $K(n)$–local spectrum. We do not know what happens for $I_{k,n}$ in general; however we note the following result.

Proposition 7.10 *The following are equivalent:*

(1) $I_{k,n} \in \mathrm{Sp}^{\mathrm{dual}}_{k,n}$;

(2) $I_{k,n} \in \mathrm{Pic}_{k,n}$;

(3) $(E_{k,n})^\vee_*(I_{k,n})$ *is a finitely generated E_*–module;*

(4) $E_n \mathbin{\overline{\wedge}} I_{k,n}$ *is $K(n)$–local.*

Proof Suppose first that (1) holds. Then, $F(I_{k,n}, I_{k,n}) \simeq DI_{k,n} \mathbin{\overline{\wedge}} I_{k,n}$, but on the other hand $F(I_{k,n}, I_{k,n}) \simeq I^2_{k,n}(L_{k,n}S^0) \simeq L_{k,n}S^0$ by Theorem 7.8. It follows that $I_{k,n} \in \mathrm{Pic}_{k,n}$, ie that (2) holds. The converse, (2) $\Longrightarrow$ (1), always holds; see for example [30, Proposition A.2.8].

The equivalence of (1) and (3) is just Theorem 5.11.

Finally to see that (1) $\iff$ (4), we note that it suffices to show that $(K_n)_* I_{k,n}$ is finite. In fact, because $(K_n)_*$ is a graded field, it suffices to see that $(K_n)^* I_{k,n}$ is finite. For this, we compute, using Example 7.7 and Theorem 7.8,

$$\begin{aligned}[I_{k,n}, K_n]_* &\simeq [I_{k,n}, I_{k,n}(K_n)]_* \\ &\simeq [I_{k,n}, F(K_n, I_{k,n})]_* \\ &\simeq [K_n, F(I_{k,n}, I_{k,n})]_* \\ &\simeq [K_n, L_{k,n} S^0]_*.\end{aligned}$$

By [32, Lemma 10.4] if M_n denotes a generalized Moore spectrum of type n, then $[E_n, L_{K(n)} M_n]_* \simeq [E_n, L_{k,n} M_n]_*$ is finite (the last equivalence follows, for example, from the fact that $L_n M_n \simeq L_{K(n)} M_n$ for a generalized Moore spectrum of type n). As in [32, Corollary 10.5] it follows that $[E_n \mathbin{\overline{\wedge}} DM_n, L_{k,n} S^0]$ is finite, and hence so is $[K_n, L_{k,n} S^0]$, as K_n lies in the thick subcategory generalized by $E_n \mathbin{\overline{\wedge}} DM_n$ (note that DM_n is also the localization of a generalized Moore spectrum of type n; see [32, Proposition 4.18]). □

Question 7.11 *For which values of k and n do the conditions of Proposition 7.10 hold?*

Remark 7.12 Condition (4) clearly holds in the case $k = n$. Of course, Proposition 7.10 is precisely Hovey and Strickland's proof in this case. However, due to the p–completion, this does not hold for $n = 1$ and $k = 0$ (Example 7.6). In fact, this fails at all heights when $k = 0$, as we now explain.

Remark 7.13 Fix $n > 1$ and $k = 0$, and take $p \gg n$. Then $\mathrm{Pic}_{0,n} \cong \mathbb{Z}$, generated by $L_n S^1$ [31, Theorem 5.4]. Therefore, if Proposition 7.10 held for $k = 0$, we must have $I_{0,n} \simeq L_n S^k$ for some $k \in \mathbb{Z}$. On the other hand, work of Hopkins and Gross [24], as written up by Strickland [52], and known results about the $K(n)$–local Picard group [26, Proposition 7.5] show that $I_{n,n} \simeq \Sigma^{n^2-n} S\langle\det\rangle$, where $S\langle\det\rangle$ is the determinant sphere spectrum [7]. It cannot then be the case that $L_{K(n)} I_{0,n} \simeq I_{n,n}$; a contradiction to Lemma 7.5. We do not know what occurs in the cases $k \neq 0, n$.

References

[1] **O Antolín-Camarena**, **T Barthel**, *Chromatic fracture cubes*, from "Equivariant topology and derived algebra" (S Balchin, D Barnes, M Kędziorek, M Szymik, editors), London Math. Soc. Lecture Note Ser. 474, Cambridge Univ. Press (2022) 100–118 MR Zbl

[2] **A Baker**, **B Richter**, *Invertible modules for commutative $\mathbb{S}$–algebras with residue fields*, Manuscripta Math. 118 (2005) 99–119 MR Zbl

[3] **P Balmer**, *The spectrum of prime ideals in tensor triangulated categories*, J. Reine Angew. Math. 588 (2005) 149–168 MR Zbl

[4] **P Balmer**, *Spectra, spectra, spectra: tensor triangular spectra versus Zariski spectra of endomorphism rings*, Algebr. Geom. Topol. 10 (2010) 1521–1563 MR Zbl

[5] **P Balmer**, *Separable extensions in tensor-triangular geometry and generalized Quillen stratification*, Ann. Sci. Éc. Norm. Supér. 49 (2016) 907–925 MR Zbl

[6] **P Balmer**, *On the surjectivity of the map of spectra associated to a tensor-triangulated functor*, Bull. Lond. Math. Soc. 50 (2018) 487–495 MR Zbl

[7] **T Barthel**, **A Beaudry**, **P G Goerss**, **V Stojanoska**, *Constructing the determinant sphere using a Tate twist*, Math. Z. 301 (2022) 255–274 MR Zbl

[8] **T Barthel**, **N Castellana**, **D Heard**, **G Valenzuela**, *Stratification and duality for homotopical groups*, Adv. Math. 354 (2019) art. id. 106733 MR Zbl

[9] **T Barthel**, **D Heard**, *The E_2–term of the $K(n)$–local E_n–Adams spectral sequence*, Topology Appl. 206 (2016) 190–214 MR Zbl

[10] **T Barthel**, **D Heard**, **N Naumann**, *On conjectures of Hovey–Strickland and Chai*, Selecta Math. 28 (2022) art. no. 56 MR Zbl

[11] **T Barthel**, **D Heard**, **G Valenzuela**, *Local duality in algebra and topology*, Adv. Math. 335 (2018) 563–663 MR Zbl

[12] **T Barthel**, **N Stapleton**, *Centralizers in good groups are good*, Algebr. Geom. Topol. 16 (2016) 1453–1472 MR Zbl

[13] **T Bauer**, *Bousfield localization and the Hasse square*, from "Topological modular forms" (C L Douglas, J Francis, A G Henriques, M A Hill, editors), Mathematical Surveys and Monographs 201, Amer. Math. Soc., Providence, RI (2014) 79–88 MR Zbl

[14] **D J Benson**, **S B Iyengar**, **H Krause**, *Colocalizing subcategories and cosupport*, J. Reine Angew. Math. 673 (2012) 161–207 MR Zbl

[15] **E S Devinatz**, *K–theory and the generating hypothesis*, Amer. J. Math. 112 (1990) 787–804 MR Zbl

[16] **E S Devinatz**, **M J Hopkins**, *Homotopy fixed point spectra for closed subgroups of the Morava stabilizer groups*, Topology 43 (2004) 1–47 MR Zbl

[17] **A Dold**, **D Puppe**, *Duality*, *trace*, *and transfer*, from "Proceedings of the international conference on geometric topology" (K Borsuk, editor), PWN, Warsaw (1980) 81–102 MR Zbl

[18] **W G Dwyer**, **J P C Greenlees**, **S B Iyengar**, *Gross–Hopkins duality and the Gorenstein condition*, J. K–Theory 8 (2011) 107–133 MR Zbl

[19] **A D Elmendorf**, **I Kriz**, **M A Mandell**, **J P May**, *Rings, modules, and algebras in stable homotopy theory*, Mathematical Surveys and Monographs 47, Amer. Math. Soc., Providence, RI (1997) MR Zbl

[20] **P Goerss**, **H-W Henn**, **M Mahowald**, **C Rezk**, *On Hopkins' Picard groups for the prime 3 and chromatic level 2*, J. Topol. 8 (2015) 267–294 MR Zbl

[21] **P G Goerss**, **M J Hopkins**, *Moduli spaces of commutative ring spectra*, from "Structured ring spectra" (A Baker, B Richter, editors), London Math. Soc. Lecture Note Ser. 315, Cambridge Univ. Press (2004) 151–200 MR Zbl

[22] **J Greenlees**, *Homotopy invariant commutative algebra over fields*, from "Building bridges between algebra and topology" (D Herbera, W Pitsch, S Zarzuela, editors), Springer (2018) 103–169 MR Zbl

[23] **D Heard**, **V Stojanoska**, *K–theory, reality, and duality*, J. K–Theory 14 (2014) 526–555 MR Zbl

[24] **M J Hopkins**, **B H Gross**, *The rigid analytic period mapping, Lubin–Tate space, and stable homotopy theory*, Bull. Amer. Math. Soc. 30 (1994) 76–86 MR Zbl

[25] **M Hopkins**, **J Lurie**, *On Brauer groups of Lubin–Tate spectra, I*, preprint (2017) Available at `https://www.math.ias.edu/~lurie/papers/Brauer.pdf`

[26] **M J Hopkins**, **M Mahowald**, **H Sadofsky**, *Constructions of elements in Picard groups*, from "Topology and representation theory" (E M Friedlander, M E Mahowald, editors), Contemp. Math. 158, Amer. Math. Soc., Providence, RI (1994) 89–126 MR Zbl

[27] **M J Hopkins**, **J H Smith**, *Nilpotence and stable homotopy theory, II*, Ann. of Math. 148 (1998) 1–49 MR Zbl

[28] **M Hovey**, *Bousfield localization functors and Hopkins' chromatic splitting conjecture*, from "The Čech centennial" (M Cenkl, H Miller, editors), Contemp. Math. 181, Amer. Math. Soc., Providence, RI (1995) 225–250 MR Zbl

[29] **M Hovey**, *Operations and co-operations in Morava E–theory*, Homology Homotopy Appl. 6 (2004) 201–236 MR Zbl

[30] **M Hovey**, **J H Palmieri**, **N P Strickland**, *Axiomatic stable homotopy theory*, Mem. Amer. Math. Soc. 610, Amer. Math. Soc., Providence, RI (1997) MR Zbl

[31] **M Hovey**, **H Sadofsky**, *Invertible spectra in the $E(n)$–local stable homotopy category*, J. London Math. Soc. 60 (1999) 284–302 MR Zbl

[32] **M Hovey**, **N P Strickland**, *Morava K–theories and localisation*, Mem. Amer. Math. Soc. 666, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[33] **M Hovey**, **N Strickland**, *Comodules and Landweber exact homology theories*, Adv. Math. 192 (2005) 427–456 MR Zbl

[34] **S B Iyengar**, **H Krause**, *The Bousfield lattice of a triangulated category and stratification*, Math. Z. 273 (2013) 1215–1241 MR Zbl

[35] **D C Johnson**, **Z Yosimura**, *Torsion in Brown–Peterson homology and Hurewicz homomorphisms*, Osaka Math. J. 17 (1980) 117–136 MR Zbl

[36] **Y Kamiya**, **K Shimomura**, *Picard groups of some local categories*, Publ. Res. Inst. Math. Sci. 43 (2007) 303–314 MR Zbl

[37] **N Karamanov**, *On Hopkins' Picard group* Pic_2 *at the prime* 3, Algebr. Geom. Topol. 10 (2010) 275–292 MR Zbl

[38] **O Lader**, *Une résolution projective pour le second groupe de Morava pour* $p \geq 5$ *et applications*, PhD thesis, Université de Strasbourg (2013) Available at `https://tel.archives-ouvertes.fr/tel-00875761`

[39] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl

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

[41] **A Mathew**, *The Galois group of a stable homotopy theory*, Adv. Math. 291 (2016) 403–541 MR Zbl

[42] **A Mathew**, *Examples of descent up to nilpotence*, from "Geometric and topological aspects of the representation theory of finite groups" (J F Carlson, S B Iyengar, J Pevtsova, editors), Springer Proc. Math. Stat. 242, Springer (2018) 269–311 MR Zbl

[43] **A Mathew**, **N Naumann**, **J Noel**, *Nilpotence and descent in equivariant stable homotopy theory*, Adv. Math. 305 (2017) 994–1084 MR Zbl

[44] **A Mathew**, **V Stojanoska**, *The Picard group of topological modular forms via descent theory*, Geom. Topol. 20 (2016) 3133–3217 MR Zbl

[45] **P Pstrągowski**, *Chromatic Picard groups at large primes*, Proc. Amer. Math. Soc. 150 (2022) 4981–4988 MR Zbl

[46] **D C Ravenel**, *Localization with respect to certain periodic homology theories*, Amer. J. Math. 106 (1984) 351–414 MR Zbl

[47] **D C Ravenel**, *Complex cobordism and stable homotopy groups of spheres*, Pure and Applied Mathematics 121, Academic, Orlando, FL (1986) MR Zbl

[48] **D C Ravenel**, *Nilpotence and periodicity in stable homotopy theory*, Annals of Mathematics Studies 128, Princeton Univ. Press (1992) MR Zbl

[49] **C Rezk**, *Notes on the Hopkins–Miller theorem*, from "Homotopy theory via algebraic geometry and group representations" (M Mahowald, S Priddy, editors), Contemp. Math. 220, Amer. Math. Soc., Providence, RI (1998) 313–366 MR Zbl

[50] **V Stojanoska**, *Duality for topological modular forms*, Doc. Math. 17 (2012) 271–311 MR Zbl

[51] **N P Strickland**, *Products on* MU*–modules*, Trans. Amer. Math. Soc. 351 (1999) 2569–2606 MR Zbl

[52] **N P Strickland**, *Gross–Hopkins duality*, Topology 39 (2000) 1021–1033 MR Zbl

[53] **N Strickland**, **N Bellumat**, *Iterated chromatic localisation*, preprint (2019) arXiv 1907.07801

[54] **P Symonds**, **T Weigel**, *Cohomology of p–adic analytic groups*, from "New horizons in pro-p groups" (M du Sautoy, D Segal, A Shalev, editors), Progr. Math. 184, Birkhäuser, Boston, MA (2000) 349–410 MR Zbl

[55] **F L Wolcott**, *Variations of the telescope conjecture and Bousfield lattices for localized categories of spectra*, Pacific J. Math. 276 (2015) 483–509 MR Zbl

[56] **Z Yosimura**, BP *operations and homological properties of* $\mathrm{BP}_*\mathrm{BP}$*–comodules*, Osaka J. Math. 20 (1983) 123–136 MR Zbl

[57] **Z Yosimura**, *Acyclicity of BP–related homologies and cohomologies*, Osaka J. Math. 22 (1985) 875–893 MR Zbl

Department of Mathematical Sciences, Norwegian University of Science and Technology Trondheim, Norway

drew.k.heard@ntnu.no

Received: 13 August 2021 Revised: 15 February 2022

Algebraic & Geometric Topology 23:8 (2023) 3707–3743
DOI: 10.2140/agt.2023.23.3707
Published: 5 November 2023

Smooth one-dimensional topological field theories are vector bundles with connection

DANIEL BERWICK-EVANS
DMITRI PAVLOV

We prove that smooth 1–dimensional topological field theories over a manifold are equivalent to vector bundles with connection. The main novelty is our definition of the smooth 1–dimensional bordism category, which encodes *cutting* laws rather than *gluing* laws. We make this idea precise through a smooth version of Rezk's complete Segal spaces. With such a definition in hand, we analyze the category of field theories using a combination of descent, a smooth version of the 1–dimensional cobordism hypothesis, and standard differential-geometric arguments.

53C05, 53C29, 55R10, 57R56

1 Introduction

The goal of this paper is to give a definition of smooth 1–dimensional field theory that plays well with the differential geometry of manifolds. A technical ingredient in our approach is the notion of *smooth ∞–categories* (developed in Section 2). A smooth ∞–category is a smooth version of a complete Segal space. This framework can be used to encode cutting axioms for the value of a field theory on a cobordism rather than the usual gluing axioms. This perspective on cobordisms is implicit in the work of Galatius, Madsen, Tillman and Weiss [12] and Lurie [20] in the settings of topological and ∞–categories, respectively. Translating into the smooth setting yields field theories that are determined by familiar differential-geometric objects.

Theorem A *The space of smooth* 1*–dimensional oriented topological field theories over a smooth manifold* X *is equivalent to the nerve of the groupoid of* (*finite-dimensional*) *vector bundles with connection over* X *and connection-preserving vector bundle isomorphisms. The equivalence is natural in* X.

In our view, the above characterization of smooth 1–dimensional topological field theories is the only admissible one. As such, the main contribution of this paper is a precise definition of smooth field theory for which Theorem A holds. The definition readily generalizes both to higher dimensions and nontopological smooth field theories, as pursued by Grady and the second author [13, Section 4]. Through its connection to familiar objects, Theorem A gives a concrete idea of what these more complicated field theories seek to generalize. Our methods — particularly the role of descent in Theorem C — are chosen with higher-dimensional generalizations in mind; see [13, Theorem 1.0.1].

The intuition behind Theorem A goes back to Segal [24, Section 6]. The 1–dimensional bordism category over X has objects compact 0–manifolds with a map to X and morphisms compact 1–manifolds with boundary with a map to X. A 1–dimensional topological field theory over X is a symmetric monoidal functor from the 1–dimensional bordism category over X to the category of vector spaces. Hence, to each point in X a topological field theory assigns a finite-dimensional vector space and to each path the field theory assigns a linear map. A vector bundle with connection produces this data using parallel transport. Conversely, given the data of parallel transport one may assign values to arbitrary 1–dimensional bordisms in X by taking tensor products and duals. This gives a functor from the groupoid of vector bundles with connection to the groupoid of 1–dimensional topological field theories. To verify Theorem A, one must show that this functor is an equivalence. This can be thought of as the following smoothly parametrized variant of the 1–dimensional cobordism hypothesis of Baez and Dolan [1]. We note that in dimension 1, orientations are equivalent to framings.

Theorem B *Let* $\mathsf{Vect}^{\otimes}$ *denote the symmetric monoidal smooth* ∞*–category of vector spaces and* Vect *the underlying smooth* ∞*–category without monoidal structure. There is an equivalence between* 1*–dimensional oriented topological field theories over* X *valued in* $\mathsf{Vect}^{\otimes}$ *and* C^{∞}*–functors from the smooth path category of* X *to* Vect.

Theorem A follows from Theorem B by identifying a functor from the path category to Vect with a smooth vector bundle and connection (see Section 4). Such a relationship between parallel transport and representations of path categories goes back to Kobayashi [17], who introduced the group of smooth based loops modulo thin homotopy and established that smooth homomorphisms from this group to a Lie group G are in bijection with isomorphism classes of principal G–bundles with connection. Similar results were proved by Freed [11], and Schreiber and Waldorf [22]. An analogous result for gerbes with connections was established for two-dimensional thin homotopies by Bunke, Turner and Willerton [7]. A statement explicitly relating field theories to vector bundles with connection was loosely formulated by Segal in his early work on geometric models for elliptic cohomology. A precise statement was given by Stolz and Teichner in their language of field theories fibered over manifolds [26, Theorem 1.8], though a proof has yet to appear. Below we draw inspiration from all of these authors; the new ingredient is in our treatment of the smooth bordism category.

1.1 What makes a smooth cobordism category difficult to define?

Composition of cobordisms is more subtle than one might hope: given two d–manifolds and a $(d-1)$–manifold along which one wishes to glue, one only obtains a glued manifold *up to diffeomorphism*. The usual solution is to define morphisms in a cobordism category as smooth d–manifolds up to diffeomorphism, thereby obliterating the problem. However, if we wish to include extra structures on bordisms, two problems arise: (1) gluings may fail to exist and (2) gluing isomorphism classes of geometric structures is typically ill-defined. For example, isomorphism classes of metrics cannot be glued, and gluing smooth maps to a target manifold X along a codimension 1 submanifold may not result in a smooth map to X. Although the focus in this paper is on the topological bordism category over X, a guiding principle is to make definitions for which various flavors of generalization pose no serious technical difficulties.

A reason for pursuing such generalizations comes from Stolz and Teichner's work on a geometric model for elliptic cohomology [25; 26], following Segal [24]. Their program seeks to generalize the relationship between 1–dimensional field theories, vector bundles, and K–theory to provide a model for elliptic cohomology with cocycles coming from (supersymmetric) 2–dimensional Euclidean field theories. They address the problem of composition by equipping bordisms with germs of collars, so bordisms are composable when collars match. This allows one to incorporate geometric structures on paths by simply endowing the collars with geometric structures. However, it introduces a new technical issue when relating 1–dimensional field theories to vector bundles: isomorphism classes of objects in their 1–dimensional bordism category are points of X *together* with the germ of a collar of a path. Such a large space of objects can be rather unwieldy in computations (particularly in the presence of geometric structures on the bordisms). The original motivation for this paper was to find a way around the technical difficulties brought on by the introduction of collars, with an eye towards studying supersymmetric Euclidean field theories. We note that since the writing of this paper, an analog of Theorem A was proved by Ludewig and Stoffel [18] in a framework that incorporates collared bordisms (we comment further on this work in Section 1.3).

One approach that avoids collars follows Kobayashi [17], Caetano and Picken [8], and Schreiber and Waldorf [22] who study paths in a manifold modulo *thin homotopy*; these are smooth paths modulo smooth homotopies whose rank is at most 1. Each equivalence class has a representative given by a path with sitting instant, meaning a path in X for which some neighborhood of the start and end point is mapped constantly to X. These sitting instants allow concatenation of smooth paths in a straightforward way, which simplifies many technical challenges (compare Lemma 5.0.1), and leads to a version of a 1–dimensional bordism category over X. However, endowing an equivalence class of a path with a geometric structure, eg a metric, is hopeless. If one does not pass to equivalence classes but instead works with honest paths with sitting instants, the resulting path category fails to restrict along open covers of X: restricted paths may not have sitting instants. This destroys a type of locality that we find both philosophically desirable and computationally essential (compare Theorem C), related to Mayer–Vietoris sequences in Stolz and Teichner's program; see [26, Conjecture 1.17]. In summary, techniques involving bordisms with sitting instants seem appropriate only for a certain class of topological field theories.

Lurie [20] and Galatius, Madsen, Tillman and Weiss [12] take a different road, considering a topological category (or Segal space) of bordisms, wherein composition need only be defined up to homotopy. This allows one to effectively add or discard the collars with impunity since this data is contractible. Geometric structures can be incorporated as stable tangential structures, ie maps from bordisms to certain classifying spaces. This framework also leads to field theories that are relatively easy to work with: one can obtain a precise relationship between maps from X to $\mathrm{BO}(k)$ and 1–dimensional topological field theories over X. However, the price one pays is that such bundles are not smooth, but merely topological. By this we mean a particular space of field theories is homotopy equivalent to the space of maps from X (viewed as a topological space, not a manifold) to the classifying space of vector bundles; in particular, from this vantage the data of a connection is contractible. Our search for a smooth bordism category is tantamount to asking for a differential refinement of this data. In the case of line bundles, such a refinement is the jumping-off point for the subject of (ordinary) differential cohomology, and one can view our undertaking as a close cousin.

1.2 Why model categories?

Our approach combines aspects of Stolz and Teichner's definition of bordism categories internal to smooth stacks [25; 26] and the Segal space version (in the world of model categories) studied by Lurie. What we obtain is not a bordism *category* strictly speaking, but rather bordisms in X comprise a collection of objects and morphisms with a partially defined composition; this is the categorical translation of the geometric idea to encode cutting laws rather than gluing laws. To make sense of functors out of this bordism "category" we provide ourself with an ambient category of smooth categories with partially defined composition. This is directly analogous to Rezk's category of complete Segal spaces as a model for ∞–categories. However, making the framework precise requires a foray into the world of model categories. This sort of machinery rarely turns up in standard differential geometry, so might seem a little misplaced at first glance; however, in the setting of field theories some basic features of this language seem unavoidable. For example, the bordism category over X ought to be equivalent to the bordism category over an open cover $\{U_i \to X\}$ with appropriate compatibility conditions on intersections; asking for these categories to be *isomorphic* is too strong since, for example, they have different sets of objects. Hence, the appropriate categorical setting for describing bordisms over X must have some native notion of (weak) equivalence of bordism categories, and the language of model categories was built precisely to facilitate computations in such situations.

Smooth ∞–categories are fibrant objects for a model category structure we place on simplicial objects in smooth stacks. This model structure is chosen to satisfy three properties: (1) nerves of categories fibered over manifolds determine fibrant objects, eg the category of (smooth) vector spaces is fibrant; (2) all objects are cofibrant, and in particular bordisms over X define a cofibrant object; and (3) there is a weak equivalence between 1–dimensional bordisms over X and 1–dimensional bordisms over an open cover $\{U_i\}$ of X with compatibility on overlaps. These three properties are the only features of the model structure we actually use. It is important that bordisms over X do not give a fibrant object: this is precisely the failure of composition to be defined in general. However, to compute the (derived mapping) space of smooth functors from bordisms to vector spaces, fibrant replacement of the source is unnecessary.

We refer the reader to Hirschhorn [15], Barwick [2], and Lurie [19, Appendix A] for background on model categories relevant to this paper.

1.3 Subsequent work

Since the first version of this paper appeared in 2015, its basic ideas have been used and expanded by several authors. Some of these developments were alluded to above; we discuss this further presently.

Ludewig and Stoffel [18] constructed a bordism category using model-categorical techniques, incorporating ideas similar to those in this paper. One important difference is that their bordisms are equipped with the germ of a collar, following the ideas of Stolz and Teichner [26]. Ludewig and Stoffel go on to prove a version of Theorem A; see [18, Theorems 1.1 and 1.2]. This shows that incorporating collars in the context of 1–dimensional topological field theories has no effect on the underlying geometric objects of study. Another important result of Ludewig and Stoffel shows that a version of Theorem A holds where the target category consists of (not necessarily locally free) sheaves of vector spaces [18, Theorem 5.2]. These more general sheaves are an important piece of Stolz and Teichner's formalism; see [26, Remark 3.16]. As in the formalism below, a crucial tool in Ludewig and Stoffel's work is the descent property for field theories. We also mention that the main results of Benini, Perin and Schenkel [3] verify and utilize descent for a distinct (though related) category of 1–dimensional *algebraic* quantum field theories.

In [13], Grady and the second author generalize the 1–dimensional topological bordism category defined in this paper to arbitrary d–dimensional bordism d–categories. Their definitions also allow one to incorporate a wide class of geometric structures on d–dimensional bordisms, expanding the ideas of Stolz and Teichner [26] to the fully extended context. The categorical foundations of Grady–Pavlov involve a smooth refinement of d–fold iterated Segal spaces, generalizing the definition of smooth ∞–category developed below. The main result of Grady and Pavlov [13] is that fully extended d–dimensional geometric field theories satisfy descent; their proof generalizes the ideas of Section 2. This descent statement leads to good behavior for groupoids of field theories over a manifold. In particular, concordance classes of d–dimensional geometric field theories are representable [13, Section 7.2]. This fits nicely with Stolz and Teichner's point of view: the main conjectures from [26, Section 1] together with Brown representability require that concordance classes of $1|1$– and $2|1$–dimensional Euclidean field theories be representable. Descent gives a conceptually satisfying mechanism for representability of a wide class of geometric field theories.

Another important development is a proof — by Grady and the second author [14] — of a geometric version of the cobordism hypothesis. The model categorical framework is crucial, as freely adjoining duals to a smooth (∞, d)–category can be realized by Bousfield localization. The formalism and techniques of Section 2 continue to play a central role: generalizations of cutting-and-gluing constructions decompose bordisms into elementary handles that in turn generate the bordism categories of interest.

1.4 Notation and terminology

Definition 1.4.1 Let Cart denote the *cartesian site* whose objects are $\mathbb{R}^n$ for $n \in \mathbb{N}$, morphisms are all smooth maps, and coverings are the usual open coverings, $\coprod \mathbb{R}^n \to \mathbb{R}^n$.

We will use the notation $[k]$ to denote the finite set $\{0, 1, \ldots, k\}$ as an object of the category Δ of simplices. The word "space" will often be used to refer to simplicial sets (ie objects in the category $\mathsf{sSet} = \mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{Set})$), eg a simplicial space is a bisimplicial set.

We will sometimes refer to objects in C^∞–Cat and C^∞–$\mathsf{Cat}^\otimes$ as categories even when they are not, eg we will often refer to 1–$\mathsf{Bord}^{\mathsf{or}}(X)$ as the 1–dimensional oriented bordism category over X.

1.5 Structure of the paper

We have attempted to keep the model-categorical discussion separate from the geometric one, relegating the former to Section 2 and the latter to Sections 3–5.

Acknowledgements

We thank Peter Teichner for many enlightening discussions about field theories, Ralph Cohen for his encouragement, Konrad Waldorf for helpful correspondence about representations of path groupoids, and Gabriel Drummond-Cole for feedback on an earlier draft. Pavlov was partially supported by the SFB 878 grant.

2 From ∞–categories to smooth ∞–categories

Complete Segal spaces provide a model for ∞–categories. A complete Segal space is a simplicial space (ie a functor $\Delta^{\mathrm{op}} \to \mathsf{sSet}$) satisfying a Segal and completeness condition reviewed below. As described by Rezk [21], the value of a Segal space on $[k] \in \Delta$ can be thought of as the classifying space for chains of morphisms of length k in an ordinary category. This has an obvious smooth enhancement via simplicial objects in smooth stacks $\Delta^{\mathrm{op}} \to \mathsf{Stacks}$ wherein the value on $[k]$ is a classifying stack for chains of composable morphisms of length k. This is the approach we take leading to the definition of a smooth ∞–category. By adjunction, one can also view these smooth ∞–categories as sheaves of complete Segal spaces on the smooth site Cart.

For our applications to field theories, we will assemble 1–dimensional bordisms in M into a functor $\Delta^{\mathrm{op}} \to \mathsf{Stacks}$ whose value on $[k] \in \Delta$ is the classifying stack of (smoothly) composable chains of bordisms in M of length k. As we shall see, this functor does *not* satisfy the Segal condition. The problem is geometric and unavoidable: arbitrary chains of bordisms in M compose to a piecewise smooth bordism that need not be smooth. To work with such an object, we require a larger category that includes simplicial objects in stacks that do not satisfy the Segal condition.

A systematic method for dealing with this type of issue is to construct a model category whose fibrant objects satisfy a Segal and completeness condition. This allows one to work with nonfibrant objects precisely when their failure to be fibrant is not homotopically problematic. For example, mapping out of a nonfibrant object usually presents no issues, whereas mapping into a nonfibrant object can be problematic. For complete Segal spaces, such a model structure was constructed by Rezk as a localization of the Reedy model structure on simplicial spaces. It is only a mild elaboration to extend these ideas to simplicial objects in stacks, ie smooth ∞–categories.

2.1 Complete Segal spaces as fibrant objects in a model category

We overview the small part of Rezk's theory of complete Segal spaces that we require; see Rezk [21] for a more thorough treatment.

Definition 2.1.1 A *Segal space* is a functor $C\colon \Delta^{\mathrm{op}} \to \mathsf{sSet}$ that satisfies the Segal condition, meaning the *Segal map*

(2.1.2)
$$C(k) \to C(1) \times^h_{C(0)} \cdots \times^h_{C(0)} C(1), \quad k \geq 1,$$

into the homotopy fibered product is a weak equivalence.

When the spaces are Reedy fibrant, the homotopy fibered product can be computed as the ordinary fibered product.

To obtain a category of Segal spaces into which the category of categories can be naturally embedded, it is important to also enforce a *completeness condition* that we recall presently. For any Segal space C, let hC denote the underlying homotopy category. This is an ordinary category defined by Rezk [21, Section 5], whose objects are 0–simplices of $C(0)$ and whose morphisms from x_0 to x_1 are the connected components of the homotopy fiber of $C(1)$ over (x_0, x_1) for the projection

$$d_0 \times d_1 \colon C(1) \to C(0) \times C(0).$$

For a Segal space C, let $C_{\mathrm{equiv}} \subset C(1)$ be the subspace consisting of connected components of the above fiber that correspond to isomorphisms in the homotopy category. Rezk shows that the degeneracy map $s_0 \colon C(0) \to C(1)$ factors through the subspace C_{equiv}.

Definition 2.1.3 A Segal space (Definition 2.1.1) is *complete* if the map

$$s_0 \colon C(0) \to C_{\mathrm{equiv}} \tag{2.1.4}$$

is a weak equivalence.

Following Rezk [21, Section 7], we can define a model category ∞–Cat whose underlying category is simplicial spaces and whose fibrant objects are complete Segal spaces. We achieve this by localizing a given standard model structure, namely, the Reedy model structure on simplicial spaces, which coincides with the injective model structure; see, for example, Bergner and Rezk [5, Propositions 3.10 and 4.1]. For a brief review of Reedy model structures, see Section 6. The first set of morphisms in this localization come from the morphisms of simplicial presheaves on Δ for each $k \in \mathbb{N}$ given by

$$\varphi_k \colon [1] \sqcup^h_{[0]} [1] \sqcup^h_{[0]} \cdots \sqcup^h_{[0]} [1] \to [k], \tag{2.1.5}$$

where the source is a k–fold iterated homotopy pushout in the category of simplicial presheaves on Δ (alias: simplicial spaces), and we have identified $[m] \in \Delta$ with its associated (representable) presheaf. The maps $[0] \to [1]$ defining the homotopy pushout $[1] \sqcup^h_{[0]} [1]$ are $0 \mapsto 0$ and $0 \mapsto 1$, and the higher pushouts iterate this basic version. The homotopy colimit in (2.1.5) can be computed as the ordinary colimit because the underlying diagram is cofibrant. If we map the source and target of φ_k into a simplicial space C and consider the map that φ_k induces on these mapping spaces, we obtain the

Segal map (2.1.2). (The resulting map φ_k has a cofibrant domain and codomain, which means that derived mapping spaces out of them can be computed as ordinary mapping spaces.)

Let E denote the simplicial space associated to the simplicial set given by the nerve of the groupoid with two objects x and y, and two nonidentity morphisms $x \to y$ and $y \to x$. Consider the canonical map

$$x\colon E \to [0], \tag{2.1.6}$$

where again we have identified $[0] \in \Delta$ with its associated representable simplicial presheaf. Rezk [21, Theorem 6.2] shows that $\mathrm{Map}(E, C) \simeq C_{\mathrm{equiv}}$ and the map

$$C(0) \to \mathrm{Map}(E, C) \simeq C_{\mathrm{equiv}}$$

induced by (2.1.6) is a weak equivalence if and only if C is complete. (In fact, this map is weakly equivalent to the map (2.1.4) defined above.)

Definition 2.1.7 Endow the category $\mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{sSet})$ of simplicial spaces with the Reedy model structure, which coincides with the injective structure. Define the *model category of complete Segal spaces*, denoted by ∞–Cat, as the left Bousfield localization of this model structure along the maps φ_k and x from (2.1.5) and (2.1.6).

Remark 2.1.8 It is immediate from the properties of the localized model structure that fibrant objects in ∞–Cat are simplicial spaces that are fibrant in the Reedy model structure and satisfy the Segal and completeness conditions; see Rezk [21, Theorem 7.2], or compare the proof of Lemma 2.2.7 below. In particular, fibrant objects in ∞–Cat coincide with Reedy fibrant complete Segal spaces in the sense of Definition 2.1.3.

2.2 Smooth ∞–categories

As mentioned at the beginning of the section, we take smooth ∞–categories to be a stack-valued version of Segal spaces. A (*smooth*) *stack* is a functor $F\colon \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}$ satisfying descent for good open covers $\{U_i\}$ of objects $S \in \mathsf{Cart}$, meaning the canonical map

$$F(S) \to \mathrm{holim}\Big(\prod F(U_i) \rightrightarrows \prod F(U_{ij}) \substack{\to\\\to\\\to} \prod F(U_{ijk}) \substack{\to\\\to\\\to\\\to} \cdots\Big) \tag{2.2.1}$$

is a weak equivalence. It will be useful later to observe this comes from mapping the source and target of

$$S \xleftarrow{p} \mathrm{hocolim}\Big(\coprod_i U_i \leftleftarrows \coprod_{i,j} U_i \cap U_j \substack{\leftarrow\\\leftarrow\\\leftarrow} \coprod_{i,j,k} U_i \cap U_j \cap U_k \substack{\leftarrow\\\leftarrow\\\leftarrow\\\leftarrow} \cdots\Big) \tag{2.2.2}$$

to F and considering the map induced by p between the resulting mapping spaces of simplicial presheaves on Cart. In this description, the above is regarded as a morphism of simplicial presheaves and (in particular) we have identified S and the U_i with their representable presheaves and taken the homotopy colimit in simplicial presheaves. The latter homotopy colimit can be computed as the Čech nerve of U in simplicial presheaves (where $U_i \cap U_j \cap U_k$ is placed in simplicial degree 2, and analogously for other intersections) due to the Reedy cofibrancy of the underlying simplicial diagram. The resulting morphism has a projectively cofibrant domain and codomain, which allows us to compute derived mapping spaces out of them as ordinary mapping spaces if the target is projectively fibrant, ie an objectwise Kan complex. We emphasize that in (2.2.2) the coproduct $\coprod$ is taken in the category of presheaves (which is different from the category of sheaves). The original reference for the model structure on stacks is Jardine [16]. A description in terms of left Bousfield localizations can be found in Dugger, Hollander and Isaksen [10].

Definition 2.2.3 The model category $\mathsf{PreStacks}$ is the projective model structure on simplicial presheaves on the cartesian site Cart (Definition 1.4.1), ie $\mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{sSet})$. The model category Stacks is the left Bousfield localization of $\mathsf{PreStacks}$ along the morphisms (2.2.2) for all good open covers $\{U_i\}_{i \in I}$ of any $S \in \mathsf{Cart}$.

Remark 2.2.4 Fibrant objects in $\mathsf{PreStacks}$ (Definition 2.2.3) are precisely presheaves valued in Kan complexes, whereas fibrant objects in Stacks are precisely those fibrant objects in $\mathsf{PreStacks}$ that satisfy the homotopy descent condition (2.2.1).

We now consider the Reedy model structure on the category of simplicial prestacks, ie functors $\Delta^{\mathsf{op}} \to \mathsf{PreStacks}$. The existence and basic properties of this model structure follows from Hirschhorn [15, Theorem 15.3.4], as we review briefly. Weak equivalences are objectwise, meaning a map $F \to G$ of simplicial prestacks is an equivalence if we get an equivalence of prestacks for each fixed $[n] \in \Delta$. Fibrations and cofibrations are described in terms of relative matching and latching maps; see Section 6. By adjunction, the Reedy model structure also gives a model structure on the presheaf category

$$\Delta^{\mathsf{op}} \times \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}.$$

In this description, fibrant objects $C : \Delta^{\mathsf{op}} \times \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}$ are precisely those presheaves that define Reedy fibrant simplicial spaces $C(-, S) \colon \Delta^{\mathsf{op}} \to \mathsf{sSet}$ for each $S \in \mathsf{Cart}$.

We wish to localize this Reedy model structure on simplicial prestacks along the morphisms (2.1.5), (2.1.6) and (2.2.2), but to do so we need to promote these to morphisms in presheaves on $\Delta \times \mathsf{Cart}$. We achieve this in the following way. For each representable presheaf associated to an object $[n] \in \Delta$ or $S \in \mathsf{Cart}$, we take the morphisms of presheaves

$$\mathrm{id}_{[n]} \times p, \quad \varphi_k \times \mathrm{id}_S, \quad x \times \mathrm{id}_S, \tag{2.2.5}$$

where φ_k, x and p are as in (2.1.5), (2.1.6) and (2.2.2), and $\mathrm{id}_{[n]}$ or id_S denotes the identity morphism on the corresponding representable presheaf. Letting $[n] \in \Delta$ and $S \in \mathsf{Cart}$ range over all possible objects, we obtain our localizing morphisms.

Definition 2.2.6 Define the *model category of smooth ∞–categories*, denoted by C^∞–Cat, as the left Bousfield localization of the Reedy model structure on

$$\mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{PreStacks})$$

along the set of morphisms (2.2.5). A *smooth functor* is a morphism in C^∞–Cat.

Existence and basic properties of such a localization is shown by Barwick in [2, Theorem 4.7], since the injective or projective model structure on the category of simplicial presheaves is left proper and combinatorial; see, for example, Lurie [19, Section A.2.7, Proposition A.2.8.2 and Remark A.2.8.4]. We summarize what we require as follows.

Lemma 2.2.7 *Fibrant objects $C \in \mathsf{C}^\infty$–Cat (Definition 2.2.6) are simplicial presheaves on $\Delta \times \mathsf{Cart}$ such that*

(1) *for any $S \in \mathsf{Cart}$, the restriction $C(-, S)\colon \Delta^{\mathsf{op}} \to \mathsf{sSet}$ is a fibrant complete Segal space (Remark 2.1.8);*
(2) *for any $[n] \in \Delta$, the restriction $C([n], -)\colon \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}$ is a fibrant smooth stack (Remark 2.2.4).*

Proof By Barwick [2, Theorem 4.7], an object C is fibrant in the local model structure if it is fibrant in the Reedy model structure (ie before localization) and has the additional property that for all maps $f\colon A \to B$ in (2.2.5), the induced map of derived mapping spaces

$$\mathrm{Map}(B, C) \to \mathrm{Map}(A, C) \tag{2.2.8}$$

is a weak equivalence. As observed above, C being fibrant before localization reduces to $C(-, S)$ being a Reedy fibrant simplicial space for any $S \in \mathsf{Cart}$. We observe that the maps in (2.2.5) have cofibrant source and fibrant target, so the derived mapping spaces in (2.2.8) can be computed as the usual mapping spaces. For the maps $\phi_k \times \mathrm{id}_S$ and $x \times \mathrm{id}_S$, we see that $C(-, S)$ must be a complete Segal space. For the maps $\mathrm{id}_{[n]} \times p$ we see that $C([n], -)$ must be a stack. □

2.3 Symmetric monoidal smooth ∞–categories

For our intended application to field theories, we also require a version of symmetric monoidal smooth ∞–categories. Following ideas of Segal [23], we implement this via the category Γ, the opposite category of finite pointed sets.

Just like for Δ, presheaves on Γ are first equipped with the Reedy model structure, which is then localized with respect to appropriate maps. The category Γ has nontrivial automorphisms, so the usual notion of a Reedy model structure must be generalized to accommodate this new setting, resulting in the *strict model structure* of Bousfield and Friedlander [6, Section 3]. This approach was generalized by Berger and Moerdijk [4], resulting in the notion of a *generalized Reedy category* and the associated model structure.

As explained in Segal [23], given a Γ–object X, we can think of $X_{\langle n\rangle} = X_n$ as the space of (formal) n–tuples of elements of some commutative monoid. Here "formal" means that points of X_n are not actual n–tuples, but rather have certain structure that makes them formally behave like ones. Specifically, given a map of finite pointed sets $f\colon \langle m\rangle \to \langle n\rangle$, the associated map $X_f\colon X_{\langle m\rangle} \to X_{\langle n\rangle}$ should be thought of as multiplying the elements indexed by $f^{-1}\{j\}$ for each $j \in \langle n\rangle$ (the product of an empty family is the identity element), and throwing away elements indexed by $f^{-1}\{*\}$. Given a Γ–object X, its n^{th} latching map $L_n X \to X_n$ can be thought of as the subobject of X_n comprising those (formal) n–tuples where at least one element is the identity. Given a Γ–object X, its n^{th} matching map $X_n \to M_n X$ can be thought of as sending a formal n–tuples in X_n to the compatible family of formal tuples given by multiplying two or more elements, or throwing away one or more elements. A Γ–object X is cofibrant in the strict model structure if for all $n \in \Gamma$, the latching map $L_n X \to X_n$ is a projective cofibration of objects equipped with an action of Σ_n. A Γ–object X is fibrant in the strict model structure if for all $n \in \Gamma$, the matching map $X_n \to M_n X$ is a fibration.

To define the Reedy model structure on presheaves on $\Delta \times \Gamma$, it suffices to observe that generalized Reedy categories are closed under finite products by Berger and Moerdijk [4, Section 1] and both Δ^{op} and Γ^{op} have generalized Reedy category structures by Berger and Moerdijk [4, Examples 1.9(a) and (b)]. Thus, we can consider the (generalized) Reedy model structure on the category of functors

$$\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \to \mathsf{PreStacks}.$$

Next, we turn attention to the morphisms used to define a local model structure. As above, we define these morphisms in their adjoint form in the category of functors

$$\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \times \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}.$$

Denote objects of Γ by $\langle m\rangle = \{*, 1, \dots, m\}$, where $*$ is the basepoint. Then consider the set of morphisms

$$u\colon \varnothing \to \langle 0\rangle, \qquad t^{m,n}\colon \langle m\rangle \sqcup_{\langle 0\rangle} \langle n\rangle \to \langle m+n\rangle,$$

where the maps $t^{m,n}$ are induced by a pair of maps of finite pointed sets (which yield morphisms in Γ in the opposite direction) $\langle m\rangle \leftarrow \langle m+n\rangle$ and $\langle n\rangle \leftarrow \langle m+n\rangle$. The first of these maps is the identity on the subset $\{1, \dots, m\}$ and sends $m+1, \dots, m+n$ to $*$. The second of these maps uses the obvious bijection from $\{m+1, \dots, m+n\}$ to $\{1, \dots, n\}$ and sends the remainder to $*$.

We then consider a set of localizing morphisms similar to (2.2.5), only now we have to fix objects in a pair of the categories Δ, Γ and Cart. Explicitly, we take morphisms

(2.3.1)
$$\begin{gathered} \mathrm{id}_{[n]} \times \mathrm{id}_{\langle m\rangle} \times p, \quad \varphi_k \times \mathrm{id}_{\langle m\rangle} \times \mathrm{id}_S, \quad x \times \mathrm{id}_{\langle m\rangle} \times \mathrm{id}_S, \\ \mathrm{id}_{[n]} \times t^{m,n} \times \mathrm{id}_S, \quad \mathrm{id}_{[n]} \times u \times \mathrm{id}_S, \end{gathered}$$

where $[n] \in \Delta$, $\langle m\rangle \in \Gamma$, and $S \in \mathsf{Cart}$ vary over all possible objects.

Definition 2.3.2 Define the *model category of symmetric monoidal smooth ∞–categories*, denoted by $\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes$, as the left Bousfield localization of the Reedy model structure on the category of functors $\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \to \mathsf{PreStacks}$ with respect to the morphisms (2.3.1).

Again, existence of such a localization is shown by Barwick [2, Theorem 4.7], which also proves that fibrant objects in the localized structure are precisely fibrant and local objects in the original model structure.

Lemma 2.3.3 *Fibrant objects $C \in \mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes$ (Definition 2.3.2) are Reedy fibrant simplicial presheaves $C\colon \Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \times \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}$ (meaning the adjoint map $\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \to \mathsf{PreStacks}$ is Reedy fibrant, where $\mathsf{PreStacks}$ is equipped with the projective model structure of Definition 2.2.3) such that*

(1) *for any* $S \in \mathsf{Cart}$ *and* $\langle m\rangle \in \Gamma$, *the restriction* $C(-, \langle m\rangle, S)\colon \Delta^{\mathsf{op}} \to \mathsf{sSet}$ *is a complete Segal space* (*Remark 2.1.8*);

(2) *for any fixed* $[n] \in \Delta$ *and* $\langle m\rangle \in \Gamma$, *the restriction* $C([n], \langle m\rangle, -)\colon \mathsf{Cart}^{\mathsf{op}} \to \mathsf{sSet}$ *is a smooth stack* (*Remark 2.2.4*);

(3) *for any fixed* $S \in \mathsf{Cart}$ *and* $[n] \in \Delta$, *the restriction* $C([n], -, S)\colon \Gamma^{\mathsf{op}} \to \mathsf{sSet}$ *is a special* Γ*–space* (*Segal* [23, Definition 1.2] *and Bousfield and Friedlander* [6, Section 4]).

Proof Fibrant objects in the local model structure are fibrant object in the Reedy model structure (selected by the first condition) that additionally satisfy the locality condition (2.2.8) with respect to the maps (2.3.1). For the maps $\phi_k \times \mathrm{id}_{\langle m\rangle} \times \mathrm{id}_S$ and $x \times \mathrm{id}_{\langle m\rangle} \times \mathrm{id}_S$ we see using the adjunction property that $C(-, \langle m\rangle, S)$ must be a complete Segal space. For the maps $\mathrm{id}_{[n]} \times \mathrm{id}_{\langle m\rangle} \times p$ we see that $C([n], [m], -)$ must be a stack. For the maps $\mathrm{id}_{[n]} \times t^{m,n} \times \mathrm{id}_S$ and $\mathrm{id}_{[n]} \times u \times \mathrm{id}_S$ we see that $C([n], -, S)$ must be a special Γ–space. □

Definition 2.3.4 We have a functor

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}^{\otimes} \to \mathsf{C}^\infty\text{–}\mathsf{Cat}$$

that restricts a simplicial presheaf on $\Delta \times \Gamma \times \mathsf{Cart}$ to $\Delta \times \langle 1\rangle \times \mathsf{Cart} \cong \Delta \times \mathsf{Cart}$. We call this the *forgetful functor* from symmetric monoidal smooth ∞–categories to smooth ∞–categories.

By virtue of Lemmas 2.2.7 and 2.3.3, the forgetful functor preserves fibrant objects and weak equivalences between them, and in fact is a right Quillen functor, which can be seen as follows. Hirschhorn [15, Theorem 15.5.2] shows that there is no difference between the Reedy model structures on

$$\mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{Fun}(\Gamma^{\mathsf{op}}, \mathsf{PreStacks})),$$

$$\mathsf{Fun}(\Gamma^{\mathsf{op}}, \mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{PreStacks})),$$

$$\mathsf{Fun}(\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}}, \mathsf{PreStacks}),$$

using that $\Delta^{\mathsf{op}} \times \Gamma^{\mathsf{op}}$ is a (generalized) Reedy category, with Hirschhorn's proof still working for generalized Reedy categories. In particular, the forgetful functor can be presented as evaluation at $\langle 1\rangle \in \Gamma$,

$$\mathsf{Fun}(\Gamma^{\mathsf{op}}, \mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{PreStacks})) \to \mathsf{Fun}(\Delta^{\mathsf{op}}, \mathsf{PreStacks}),$$

hence it is a right Quillen functor because Reedy (acyclic) fibrations are projective (acyclic) fibrations.

2.4 Smooth ∞–categories from sheaves of categories

The main example of a Segal space comes from a simplicial space-valued nerve of an ordinary category; see Rezk [21, Section 3.3]. For small categories $\mathcal{C}$ and $\mathcal{D}$, let $\mathrm{Iso}(\mathcal{C}^{\mathcal{D}})$ denote the category whose objects are functors $\mathcal{D} \to \mathcal{C}$ and whose morphisms are natural isomorphisms of functors. Then define

(2.4.1) $$\mathbb{N}^{\infty}\colon \mathsf{Cat} \to \mathrm{Fun}(\Delta^{\mathsf{op}}, \mathsf{sSet}), \quad \mathcal{C} \mapsto ([l] \mapsto \mathrm{N}(\mathrm{Iso}(\mathcal{C}^{[l]}))).$$

Rezk [21, Proposition 6.1] proves that $\mathbb{N}^{\infty}(C)$ is a complete Segal space; we sketch the idea. The set of (k,l)–bisimplices of $\mathbb{N}^{\infty}(\mathcal{C})$ is the set of diagrams in $\mathcal{C}$

(2.4.2) $$\mathrm{Nerve}_k(\mathrm{Iso}(\mathcal{C}^{[l]})) = \left\{ \begin{array}{ccccccc} c_{00} & \xrightarrow{f_{10}} & c_{10} & \xrightarrow{f_{20}} & \cdots & \xrightarrow{f_{l0}} & c_{l0} \\ \downarrow & & \downarrow & & & & \downarrow \\ c_{01} & \xrightarrow{f_{11}} & c_{11} & \xrightarrow{f_{21}} & \cdots & \xrightarrow{f_{l1}} & c_{l1} \\ \downarrow & & \downarrow & & & & \downarrow \\ \vdots & & \vdots & & & & \vdots \\ \downarrow & & \downarrow & & & & \downarrow \\ c_{0k} & \xrightarrow{f_{1k}} & c_{1k} & \xrightarrow{f_{2k}} & \cdots & \xrightarrow{f_{lk}} & c_{lk} \end{array} \right\},$$

where $c_{ij} \in \mathcal{C}$ are objects, the horizontal arrows f_{ij} are morphisms in $\mathcal{C}$, and the vertical arrows are isomorphisms in $\mathcal{C}$. These diagrams stack horizontally, from which one deduces that the resulting simplicial space satisfies the Segal condition. Furthermore, the space we get from setting $l = 0$ consists of chains of invertible morphisms; unraveling the definitions (and with a bit of work), this verifies that the Segal space is complete.

To generalize this construction for a nerve valued in smooth ∞–categories, we consider diagrams like (2.4.2) with each c_{ij} an object in a category-valued presheaf on Cart.

Definition 2.4.3 Given a (strict) presheaf $\mathcal{C}\colon \mathsf{Cart}^{\mathsf{op}} \to \mathsf{Cat}$ of categories, consider the presheaf of groupoids $\mathcal{C}_{\Delta}\colon \mathsf{Cart}^{\mathsf{op}} \times \Delta^{\mathsf{op}} \to \mathsf{Grpd}$ defined by the formula

$$(S,[k]) \mapsto \mathrm{Iso}(\mathcal{C}(S)^{[k]}).$$

Define the *nerve of a category-valued presheaf* as $\mathbb{N}^{\mathrm{C}^{\infty}}(\mathcal{C}) := \mathrm{N}(\mathcal{C}_{\Delta})$.

Lemma 2.4.4 *If $\mathcal{C}$ satisfies descent, then so does $\mathcal{C}_{\Delta}$ and therefore $\mathbb{N}^{\mathrm{C}^{\infty}}(\mathcal{C})$ is fibrant in the model structure of Definition 2.2.6.*

Proof This follows immediately from Lemma 2.2.7. The part with fixed S is verified by Rezk. The part with fixed $[k]$ is satisfied because $(-)^{[k]}$ and $\mathrm{Iso}(-)$ both preserve homotopy limits of categories and therefore preserve the descent property for presheaves of categories. $\square$

We will also need a symmetric monoidal version of the previous lemma. This is the classical construction of a Γ–object from a symmetric monoidal object with a strict monoidal structure.

Definition 2.4.5 Given a presheaf $\mathcal{C}\colon \mathsf{Cart}^{\mathsf{op}} \to \mathsf{SymCat}$ of symmetric monoidal categories with a strict monoidal structure (and symmetric strict monoidal functors as morphisms), consider the presheaf

$$\mathcal{C}_\Gamma\colon \mathsf{Cart}^{\mathsf{op}} \times \Gamma^{\mathsf{op}} \to \mathsf{Cat}$$

defined by the formula

$$\mathcal{C}_\Gamma(S, \langle l\rangle) = \mathcal{C}(S)^l$$

with the naturality in $S \in \mathsf{Cart}$ induced by $\mathcal{C}$ and naturality in $\langle l\rangle \in \Gamma$ induced by the strict monoidal structure on $\mathcal{C}(S)$ and its symmetric braiding. Define the *nerve of a symmetric monoidal category-valued presheaf* by $\mathbb{N}^{\mathrm{C}^\infty_\otimes}(\mathcal{C}) := \mathrm{N}((\mathcal{C}_\Gamma)_\Delta)$.

We say a presheaf of symmetric monoidal categories with strict monoidal structure on Cart satisfies descent if its underlying presheaf of categories (forgetting the symmetric monoidal structure) does. This construction continues to work when Cart is replaced by any other site. In Lemma 2.4.6, we need to use the site $\mathsf{Cart} \times \Gamma$ given by the product of the site Cart and the category Γ equipped with the trivial Grothendieck topology.

Lemma 2.4.6 *If $\mathcal{C}$ satisfies descent, then so does $\mathcal{C}_\Gamma$ and therefore $\mathbb{N}^{\mathrm{C}^\infty_\otimes}(\mathcal{C})$ is fibrant in the model structure of Definition 2.3.2.*

Proof The functor $(-)^l$ preserves homotopy limits and therefore preserves the descent condition. Thus $\mathcal{C}_\Gamma$ satisfies descent and we can invoke the previous lemma. $\square$

3 Bordisms, path categories, and field theories

In this section we present our definition of a smooth 1–dimensional topological field theory over a manifold, as well as a closely related notion of a transport functor. In Section 3.1, we apply the nerve construction of the previous section to the stack of vector bundles on Cart to obtain the symmetric monoidal smooth ∞–category of smooth vector spaces, which is a fibrant object in $\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes$. The 1–dimensional oriented bordism "category" over X, denoted by $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$, is defined in Section 3.2 and is a *non*fibrant object of $\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes$. Together, these define the main ingredients of a smooth field theory.

The model categories $\mathsf{C}^\infty\text{–}\mathsf{Cat}$ and $\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes$ are simplicial model categories, so they are equipped with mapping simplicial set functors of the form

$$C^{\mathsf{op}} \times C \to \mathsf{sSet}.$$

These functors are right Quillen bifunctors, and "derived mapping space" below refers to their right derived functors.

Definition 3.0.1 A 1*–dimensional smooth oriented topological field theory over* X is a point in the derived mapping space $\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X), \mathsf{Vect})$, and the *space of* 1*–dimensional smooth oriented topological field theories over* X is this derived mapping space.

We will also define a closely related (nonfibrant) smooth ∞–category of smooth paths in X, denoted by $\mathcal{P}(X)$, which is a nonfibrant object of $\mathsf{C}^\infty\text{–}\mathsf{Cat}$.

Definition 3.0.2 A *transport functor on* X is a point in the derived mapping space $\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(X), \mathsf{Vect})$, and the *space of transport functors on* X is the derived mapping space $\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(X), \mathsf{Vect})$.

There is a restriction functor

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}^\otimes(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X), \mathsf{Vect}) \to \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(X), \mathsf{Vect})$$

from field theories to transport functors.

Before jumping into detailed definitions of the objects in $\mathsf{C}^\infty\text{–}\mathsf{Cat}$, we overview some of the ideas that go into in Lurie's definition of the bordism category [20] as a Segal space; see also Calaque and Scheimbauer [9]. The standard way of chopping up a manifold M into k pieces is a Morse function on M with a choice of k nondegenerate critical values. One can consider the group of diffeomorphisms of M that preserve the inverse images of these critical values. The classifying space of this diffeomorphism group is roughly the value of Lurie's Segal space on $[k] \in \Delta$. This is rough because one also wants the classifying space to allow for varying Morse functions and varying critical values. In total, the result is a classifying space of the ways of cutting a manifold M along k

codimension 1 submanifolds. In terms of the usual description of the bordism category, this corresponds to a composable chain of bordisms of length $k-1$. Forgetting a codimension 1 submanifold or adding extra multiplicity gives maps between these classifying spaces, producing the requisite simplicial maps. In the 1–dimensional case, these classifying spaces are particularly easy owing to the simplicity of Morse decompositions of 1–manifolds. Our approach to $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ is the same idea, but we replace the classifying *space* with a (smooth) classifying *stack*.

3.1 Smooth vector spaces

Define the sheaf of symmetric monoidal categories, $\mathcal{V}\colon \mathsf{Cart}^{\mathrm{op}} \to \mathsf{SymCat}$, as follows. The objects of $\mathcal{V}(S)$ are elements of $\mathbb{N}$, corresponding to the dimension of a trivial bundle on $S \cong \mathbb{R}^m$, and morphisms $m \to n$ are smooth maps $S \to \mathsf{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ into the space of linear maps. The morphism of groupoids associated with maps $S \to S'$ of objects in Cart is the identity map on objects, and on morphisms we precompose. The strict monoidal structure is determined by multiplication in $\mathbb{N}$ and tensor products of linear maps. The (nontrivial) braiding is induced by the obvious block matrix. This satisfies descent because smooth functions do.

Definition 3.1.1 Let $\mathsf{Vect} \in \mathsf{C}^{\infty}\text{–}\mathsf{Cat}^{\otimes}$ (Definition 2.3.2) be the fibrant object obtained by applying Lemma 2.4.6 to the sheaf of symmetric monoidal categories $\mathcal{V}$ defined above.

To explain this a bit more concretely, the vertices of the simplicial set associated to $S \in \mathsf{Cart}$, $[k] \in \Delta$ and $\langle 1 \rangle \in \Gamma$ are chains of length k of morphisms of vector bundles over S,

(3.1.2)
$$\{V_0 \xrightarrow{\phi_1} V_1 \xrightarrow{\phi_2} \cdots \xrightarrow{\phi_k} V_k \mid V_i \to S\},$$

where the dimensions of the V_i correspond to the natural numbers in the formal definition. The 1–simplices of this simplicial set are commutative diagrams of vector bundles,

(3.1.3)
$$\left\{\begin{matrix} V_0 & \xrightarrow{\phi_1} & V_1 & \xrightarrow{\phi_2} & \cdots & \xrightarrow{\phi_k} & V_k \\ \downarrow & & \downarrow & & & & \downarrow \\ V_0' & \xrightarrow{\phi_1'} & V_1' & \xrightarrow{\phi_2'} & \cdots & \xrightarrow{\phi_k'} & V_k' \end{matrix}\right\},$$

where the vertical arrows are vector bundle isomorphisms. These 0- and 1–simplices vary with $[k] \in \Delta$ by composing horizontal morphisms of vector bundles or by inserting

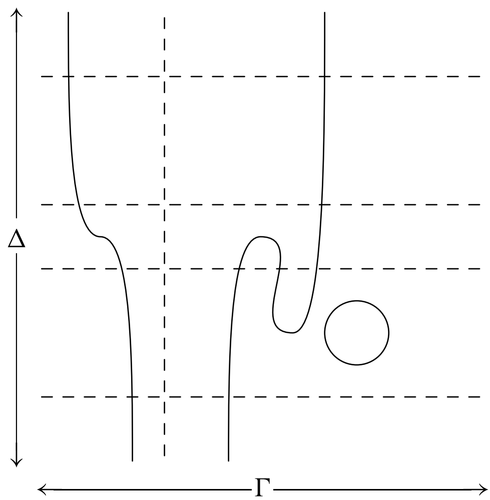


Figure 1: A vertex (ie an object) in the simplicial set $1\text{–}\mathsf{Bord}_{\mathrm{pt}}(\mathrm{pt})(2,3)$. The bordism is drawn in solid black, the height function is given by the height in the picture, and the cut functions are represented by the dotted horizontal lines. The map to $\{0,1,2\} \in \Gamma$ maps the left of the vertical dotted line to 1 and the right of the vertical line to 2. The fiber over zero is empty. Regularity at the cut values means that the intersections of the bordism with the dotted lines are transverse. Restricting attention to the bordism confined within an adjacent pair of horizontal dotted lines gives the three Segal Δ–maps and similar restrictions corresponding to the vertical dotted line gives the two Segal Γ–maps. The action by Σ_2 interchanges the bordisms on the left and right sides of the vertical dotted line.

a horizontal identity morphism of vector bundles. We can also pull this data back along smooth maps $S' \to S$.

3.2 The definition of the 1–dimensional bordism category

Definition 3.2.1 (the 1–dimensional oriented bordism category over X) Given $X \in \mathsf{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ (the most important example of X being the presheaf induced by a smooth manifold), the nonfibrant smooth symmetric monoidal ∞–category $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ (Definition 2.3.2) is defined by taking the nerve of the following presheaf of groupoids on $\Delta \times \mathsf{Cart} \times \Gamma$: send $[l] \in \Gamma$, $[k] \in \Delta$, and $S \in \mathsf{Cart}$ to the groupoid whose objects are given by data

(1) M^1, an oriented 1–manifold that defines a trivial bundle $M^1 \times S \to S$,

(2) a map $\gamma\colon M^1 \times S \to X \times \{*, 1, \dots, l\}$,

(3) cut functions $t_0 \le t_1 \le t_2 \le \cdots \le t_k \in \mathrm{C}^\infty(S)$,

(4) a proper map $h\colon M^1 \times S \to \mathbb{R} \times S$ over S called the *height function* such that for each $s \in S$ the fiber of h over $s \in S$ has $t_i(s)$ as a regular value for all i.

Isomorphisms in the groupoid are certain equivalence classes defined as follows. First consider the set of smooth functions $\epsilon_0, \epsilon_k, \epsilon'_0, \epsilon'_k : S \to (0,\infty)$ together with orientation-preserving diffeomorphisms

$$\phi\colon h^{-1}((t_0 - \epsilon_0, t_k + \epsilon_k) \times S) \to (h')^{-1}((t'_0 - \epsilon'_0, t'_k + \epsilon'_k) \times S)$$

over $S \times X$ such that ϕ restricts to a fiberwise diffeomorphism over $S \times X$ of the form

$$h^{-1}([t_i, t_j] \times S) \to (h')^{-1}([t'_i, t'_j] \times S)$$

for any $0 \le i \le j \le k$. Two such elements are equivalent if their restrictions to $h^{-1}([t_0, t_k] \times S)$ coincide. The quotient sets admit a well-defined composition operation, given by pointwise composition.

Functoriality in S is given by the composition of γ, t, h, and ϕ with the given map $S' \to S$. Functoriality in Γ is given by postcomposing γ with the given map of finite sets $\{*, 1, \ldots, l\} \to \{*, 1, \ldots, l'\}$. Functoriality in Δ with respect to a morphism of simplices $[k'] \to [k]$ is given by dropping those t_i for which i is not in the image of $[k']$ and duplicating those t_i that are in the image of more than one element of $[k']$, and restricting ϕ accordingly.

We observe that $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ is covariant in X: a smooth map $X \to Y$ induces a smooth symmetric monoidal functor $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X) \to 1\text{–}\mathsf{Bord}^{\mathrm{or}}(Y)$.

Remark 3.2.2 The properness assumption on 0–simplices in property (1) guarantees that the 1–dimensional bordism "between" $S \times \{t_0\}$ and $S \times \{t_k\}$ is compact in each fiber over S.

Remark 3.2.3 The second piece of data, $M^1 \times S \to X \times \{*, 1, \ldots, l\}$, encodes both the map from the bordism to X, and the partition of connected components of this bordism associated with the monoidal structure. Our definition of $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ does not satisfy the Segal Γ–condition, but this failure of fibrancy is not a problem for computing field theories.

Remark 3.2.4 Below, we will find it convenient to replace $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ with a weakly equivalent object $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)$, which coincides with $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ for all $[n] \in \Delta$ except for $n = 0$, where we replace the Γ–object $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)(0)$ with the Γ–object

$\langle m \rangle \mapsto (X \sqcup X)^m$, where $\langle m \rangle = \{*, 1, \dots, m\}$ and X^m denotes the representable presheaf of the m–fold cartesian product of $X \sqcup X$, corresponding to the two orientations of points. There are canonical homotopy equivalences $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X) \to 1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ and $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X) \to 1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)$ that identify points in X with constant paths in X equipped with the cut function $t_0 = 0$. The advantage of this replacement is that $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)(0)(m)$ is a representable presheaf for all $\langle m \rangle \in \Gamma$, hence a cofibrant object in the projective model structure on Stacks.

3.3 The category of smooth paths in a smooth manifold

Similar to the intuition behind cutting bordisms, we can also consider an analogous structure for paths in X. In this case a Morse function is afforded by the parametrization of the path itself.

Definition 3.3.1 Given $X \in \mathrm{Fun}(\mathsf{Cart}^{\mathrm{op}}, \mathsf{Set})$ (the most important example of X being the presheaf induced by a smooth manifold), define the *smooth path category* $\mathcal{P}X \in \mathrm{C}^\infty\text{–}\mathsf{Cat}$ (*Definition 2.2.6*) *of* X as the nerve of the presheaf of groupoids on $\Delta \times \mathsf{Cart}$ constructed as follows. A pair $[k] \in \Delta$, $S \in \mathsf{Cart}$ is sent to the groupoid whose objects consist of a map $\gamma\colon S \times \mathbb{R} \to X$ and cut functions $t_0 \le t_1 \le \cdots \le t_k \in \mathrm{C}^\infty(S)$.

Isomorphisms in the groupoid are equivalence classes of a certain equivalence relation. Elements in the underlying set of this equivalence relation are smooth functions $\epsilon_0, \epsilon_k, \epsilon_0', \epsilon_k'\colon S \to (0, \infty)$ together with orientation-preserving diffeomorphisms

$$\phi\colon (t_0 - \epsilon_0, t_k + \epsilon_k) \times S \to (t_0' - \epsilon_0', t_k' + \epsilon_k') \times S$$

over $S \times X$ such that ϕ restricts to a fiberwise diffeomorphism over $S \times X$ of the form

$$[t_i, t_j] \times S \to [t_i', t_j'] \times S$$

for any $0 \le i \le j \le k$. Two such elements are equivalent if their restrictions to $[t_0, t_k] \times S$ coincide. The quotient sets admit a well-defined composition operation, given by pointwise composition.

Functoriality in S is given by the appropriate composition of γ, t, and ϕ with the given map $S' \to S$. Functoriality in Δ with respect to a morphism of simplices $[k'] \to [k]$ is given by dropping those t_i for which i is not in the image of $[k']$ and duplicating those t_i that are in the image of more than one element of $[k']$, and restricting ϕ accordingly.

We recall that a manifold X defines a presheaf (of sets) on Cart, and the fiber of $\mathcal{P}X$ over $[0] \in \Delta$ is homotopy equivalent to this presheaf via the map $(\gamma, t_0) \mapsto \gamma(t_0)$, with an inverse that sends a map $f\colon S \to X$ to an S–family of constant paths, $S \times \mathbb{R} \to S \xrightarrow{f} X$, with cut function the zero function. The fiber of $\mathcal{P}X$ over $S \in \mathsf{Cart}$ and $[k] \in \Delta$ is the nerve of the groupoid of S–families of paths in X with $k+1$ marked points and diffeomorphisms of these paths. The object $\mathcal{P}X$ is covariant in X, meaning a smooth map $X \to Y$ induces a smooth functor $\mathcal{P}X \to \mathcal{P}Y$; hence $\mathcal{P}$ is a functor from Mfld to C^∞–Cat.

There is a smooth functor $\mathcal{P}X \to U(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X))$ in C^∞–Cat (where U denotes the forgetful functor of Definition 2.3.4) we get by viewing a family of paths as the family of bordisms $S \times M^1 = S \times \mathbb{R}$ and the height function h the projection to $\mathbb{R}$. This has an induced restriction map

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}^{\otimes}(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X), \mathsf{Vect}) \to \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}X, \mathsf{Vect})$$

from 1–dimensional field theories over X to representations to the smooth path category of X. Here C^∞–$\mathsf{Cat}^{\otimes}(-,-)$ and C^∞–$\mathsf{Cat}(-,-)$ denote the corresponding *derived* mapping simplicial sets.

Remark 3.3.2 In analogy to Remark 3.2.4, we will find it convenient to replace $\mathcal{P}X$ with a weakly equivalent object $\mathcal{P}'X$, which coincides with $\mathcal{P}X$ for all $[n] \in \Delta$ except for $n = 0$, where we replace $\mathcal{P}X(0)$ with X itself. There are canonical homotopy equivalences $X \to \mathcal{P}X(0)$ and $\mathcal{P}X(0) \to X$ that identify X with constant paths in X equipped with the cut function $t_0 = 0$. The advantage of this replacement is that $\mathcal{P}'X(0)$ is a representable presheaf, hence a cofibrant object in the projective model structure on Stacks.

3.4 Descent for field theories and representations of paths

A key step to verifying the main theorem is that field theories over X and representations of smooth paths in X can be computed locally in the following sense.

Theorem C *Let $\{U_i\}$ be an open cover of a smooth manifold X. The canonical maps*

$$\operatorname{hocolim} \mathcal{P}(U_k) \to \mathcal{P}(X), \quad \operatorname{hocolim} 1\text{–}\mathsf{Bord}^{\mathrm{or}}(U_k) \to 1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$$

are equivalences of smooth ∞–categories and symmetric monoidal smooth ∞–categories, respectively. Here k runs over all finite tuples of elements in I and U_k denotes the intersection of U_i for all $i \in k$. This immediately implies that the assignments

$$X \mapsto \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(X), \mathsf{Vect}), \quad X \mapsto \mathsf{C}^\infty\text{–}\mathsf{Cat}^{\otimes}(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X), \mathsf{Vect})$$

are stacks on the site of smooth manifolds.

Proof See Grady and Pavlov [13, Theorem 1.0.1] for the case of bordism categories. The case of path categories then follows formally. We remark that the most technical part of the cited proof — Section 6.6 in [op. cit.] — becomes completely trivial in the 1–dimensional case, since the nerves of relevant categories are contractible for trivial reasons. □

We apply the above result to reduce our main theorems to the case $X \in \mathsf{Cart}$. In particular, this simplifies the construction of a transport functor from a vector bundle with connection, since the general case $X \in \mathsf{Mfld}$ would require us to work with arbitrary cocycles for vector bundles, bringing considerable technicalities, whereas for $X \in \mathsf{Cart}$ all vector bundles over X are trivial, and the problem reduces to manipulating connection 1–forms.

Definition 3.4.1 We define $\mathsf{Vect}^\nabla \in \mathsf{Stacks}$ as follows. Given $X \in \mathsf{Cart}$, we send it to the nerve of groupoid whose objects are pairs (n,ω), where $n \geq 0$ specifies a finite-dimensional vector space $V = \mathbb{R}^n$ and $\omega \in \Omega^1(X, \mathrm{End}(V))$. Morphisms $(n,\omega_0) \to (n,\omega_1)$ are smooth maps $f \in C^\infty(X, \mathrm{GL}(V))$ such that $\omega_1 = \mathrm{Ad}_{f^{-1}}\omega_0 + f^{-1}df$, where Ad denotes the adjoint action.

The groupoid $\mathsf{Vect}^\nabla(X)$ is equivalent to the groupoid of trivial vector bundles with connection over the cartesian space X.

Corollary 3.4.2 *Consider the functors*

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}(-, \mathsf{Vect}^\nabla), \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(-), \mathsf{Vect})\colon \mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})^{\mathsf{op}} \to \mathsf{sSet}$$

that send $X \in \mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ *to* $\mathsf{C}^\infty\text{–}\mathsf{Cat}(X, \mathsf{Vect}^\nabla)$ *and* $\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}X, \mathsf{Vect})$, *respectively. The space of natural weak equivalences*

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}(-, \mathsf{Vect}^\nabla) \to \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(-), \mathsf{Vect})$$

is weakly equivalent to the space of natural weak equivalences between the same functors, restricted along the Yoneda embedding $\mathsf{Cart} \to \mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ *to the category* Cart. *In particular, any natural weak equivalence*

$$\mathsf{Vect}^\nabla \to \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(-), \mathsf{Vect})$$

on Cart *can be extended to a natural weak equivalence on* $\mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ *in a unique way up to a contractible choice.*

Proof Cart generates $\mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ under homotopy colimits. Both functors,

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}(-, \mathsf{Vect}^\nabla) \quad \text{and} \quad \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}(-), \mathsf{Vect}),$$

send homotopy colimits in $\mathsf{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$ to homotopy limits in sSet. For the former, it boils down to the classical fact that vector bundles with connection satisfy descent, whereas for the latter it follows from Theorem C. □

In the next section, we construct a specific weak equivalence

$$\mathsf{Vect}^{\nabla} \to \mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}(-), \mathsf{Vect})$$

in $\mathsf{PreStacks}$, which will prove that representations of the path category of any smooth manifold X are precisely vector bundles with connection over X.

4 Representations of the smooth path category of a manifold

In this section we prove the following.

Proposition 4.0.1 *Given* $X \in \mathsf{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$, *the derived mapping space*

$$\mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}X, \mathsf{Vect})$$

is naturally weakly equivalent to $\mathsf{Vect}^{\nabla}(X)$ *of Definition 3.4.1. (In particular, we can take* $X \in \mathsf{Mfld}$.)

Proof By Corollary 3.4.2, it suffices to construct such a natural weak equivalence for $X \in \mathsf{Cart}$. Replace $\mathcal{P}X$ with weakly equivalent $\mathcal{P}'X$ from Remark 3.3.2. Recall that Vect is fibrant in $\mathsf{C}^{\infty}\text{–}\mathsf{Cat}$ (Lemma 2.2.7). Furthermore, the stack of objects $(\mathcal{P}'X)_{[0]} = \mathcal{P}'X(0) = X$ is a representable (hence cofibrant) presheaf in $\mathsf{PreStacks}$ and Vect and $\mathcal{P}'X$ are constructed as objectwise nerves of groupoids. Hence, the derived mapping space $\mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect})$ can be computed using the nonderived hom $\mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect})$. The map

$$\mathsf{Vect}^{\nabla}(X) \to \mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect})$$

is constructed in Definition 4.0.2 and is shown to be an isomorphism in Lemma 4.1.2 and Proposition 4.2.10. □

The following construction codifies the parallel transport data of a connection on a vector bundle as a functor.

Definition 4.0.2 Given $X \in \mathsf{Cart}$, we construct a map (natural in X)

$$\mathsf{Vect}^{\nabla}(X) \to \mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect})$$

as follows. A (trivial) vector bundle with connection (n,ω) defines a smooth functor $R\colon \mathcal{P}'X \to \mathsf{Vect}$ via parallel transport: to $f: S \to X$, an S–family of points in X, we assign the object n over S, defining a functor $X = \mathcal{P}'X(0) \to \mathsf{Vect}(0)$. To a family of oriented paths $S \times \mathbb{R} \to X$, we apply the fiberwise parallel transport with respect to the connection 1–form ω, yielding a morphism of (trivial) vector bundles over S. These maps are invariant under families of diffeomorphisms of 1–manifolds, so we obtain a functor $\mathcal{P}'_S X(1) \to \mathsf{Vect}_S(1)$, which is again natural in S so defines a fibered functor $\mathcal{P}'X(1) \to \mathsf{Vect}(1)$ by extension from individual fibers in the usual fashion. We extend in the obvious way to $\mathcal{P}'X(k) \to \mathsf{Vect}(k)$, where naturality with respect to maps in Δ follows from compatibility of parallel transport with concatenation of paths. Hence, we have constructed a functor from the path category of X to smooth vector spaces. Lastly, we observe that an isomorphism of vector bundles with connection leads to a natural isomorphism of functors of such functors, ie an edge in the simplicial mapping space. An n–simplex comes from a composable n–tuple of isomorphisms of vector bundles with connection.

4.1 Reduction to parallel transport data

In this section we whittle the proof of Proposition 4.0.1 down to a statement about parallel transport data, by which we shall mean smooth endomorphism-valued functions on paths that compose under concatenation of paths and are compatible with restrictions to intersections of the cover. The next definition and lemma describe the precise manner in which a representation of the path category determines a transport functor on X.

Definition 4.1.1 We define the stack $\mathrm{Tran} \in \mathsf{PreStacks}$ of transport data as follows. Given $X \in \mathsf{Cart}$, we send it to the nerve of groupoid whose objects are pairs (n, F), where $n \geq 0$ determines a vector space $V = \mathbb{R}^n$ and F is a morphism

$$F\colon \mathbb{R} \times \mathsf{Hom}(\mathbb{R}, X) \to \mathsf{End}(V)$$

in $\mathsf{PreStacks}$ such that the following three properties are satisfied: (1) for any $p\colon \mathbb{R} \to X$ and any $L_1, L_2 \in \mathbb{R}$ the following functoriality property holds:

$$F(L_2, p \circ S_{L_1}) \circ F(L_1, p) = F(L_1 + L_2, p),$$

where $S_{L_1}(t) = t - L_1$; (2) for any $p\colon \mathbb{R} \to X$ we have $F(0, p) = \mathrm{id}_V$; (3) F is invariant under diffeomorphisms of paths: if $g\colon \mathbb{R} \to \mathbb{R}$ is an orientation-preserving diffeomorphism such that $g(0) = 0$, then

$$F(L, p) = F(g^{-1}(L), p \circ g).$$

(The first argument of F specifies the length L of a smooth path p in X. The path itself is given by the second argument of F and is parametrized by $[0, L] \subset \mathbb{R}$.) Morphisms $F_1 \to F_2$ are smooth maps $h\colon X \to \mathrm{GL}(V)$ such that

$$h(p(L)) \circ F_1(L, p) = F_2(L, p) \circ h(p(0)).$$

Lemma 4.1.2 *There is an isomorphism in* $\mathsf{PreStacks}$,

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}'(-), \mathsf{Vect}) \to \mathrm{Tran}.$$

($\mathcal{P}'X$ is constructed in Remark 3.3.2.)

Proof Fix some $X \in \mathsf{Cart}$; we need to define a morphism

$$\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect}) \to \mathrm{Tran}(X).$$

Both sides are nerves of groupoids, so we define the map first on objects, then on morphisms. Pick a functor $R\colon \mathcal{P}'X \to \mathsf{Vect}$. The presheaf of objects $\mathcal{P}X(0) = X$ maps via R to a fixed object of Vect given by some dimension $n \geq 0$. The data of F is obtained by evaluating on $[1] \in \Delta$ with cut function $t_0 = 0$. On objects, we map $R \mapsto (n, F)$. Property (1) boils down to functoriality with respect to the three coface maps $[1] \to [2]$ in Δ, where the middle face map computes the composition. Property (2) boils down to functoriality with respect to the codegeneracy map $[1] \to [0]$ in Δ. Property (3) boils down to the fact that isomorphisms in $\mathcal{P}X([n], S)$ are endpoint-preserving diffeomorphisms between n–chains of paths in X, whose individual $(n+1)$ vertices are identity maps (this holds for $\mathcal{P}'X$, not for $\mathcal{P}X$). Finally, a morphism $R \to R'$ is given by a map $h\colon X = \mathcal{P}'X(0) \to \mathrm{GL}(n)$, which yields a morphism $h\colon (n, F) \to (n', F')$ (where $n = n'$).

Conversely, the inverse map

$$\mathrm{Tran}(X) \to \mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect})$$

sends (n, F) to the functor $R\colon \mathcal{P}'X \to \mathsf{Vect}$ that maps the presheaf of objects $\mathcal{P}'X(0)$ to n. The presheaf of k–simplices for $k \geq 1$ is mapped to the corresponding transport maps between endpoints. □

4.2 From parallel transport data to vector bundles with connection

From the above discussion, we have shown that a point in the (derived) mapping space $\mathsf{C}^\infty\text{–}\mathsf{Cat}(\mathcal{P}X, \mathsf{Vect})$ for $X \in \mathsf{Cart}$ defines a parallel transport data for a vector bundle on X, ie an object of $\mathrm{Tran}(X)$. In this section we explain how parallel transport data defines a vector bundle with connection. More precisely, we construct an equivalence $\mathrm{Tran}(X) \to \mathsf{Vect}^\nabla(X)$. Most of the ideas below are present in Freed [11] and Schreiber and Waldorf [22], and we have adapted them to our situation with some minor modifications.

Lemma 4.2.1 *Given $X \in \mathsf{Cart}$ and $(n, F) \in \mathrm{Tran}(X)$, the map F assigns the identity map on $V = \mathbb{R}^n$ to constant paths in X.*

Proof Since a constant path γ can be factored as the concatenation $\gamma * \gamma$, the value of F on γ must be a projection in V, denoted by P_γ. Furthermore, there is a family of constant paths parametrized by $[0,t]$ coming from the restriction of γ to $[0,t'] \subset [0,t]$. Over $t' = 0$, the constant path is the identity morphism in the path category and therefore is assigned the identity linear map. Smoothness gives a family of projections connecting P_γ on V that is the identity projection at an endpoint. Since the rank of the projection is discrete, it must be constant along this family. Therefore, P_γ is the identity. □

Lemma 4.2.2 *Given $X \in \mathsf{Cart}$ and $(n, F) \in \mathrm{Tran}(X)$, the map F lands in the invertible morphisms, ie the morphism $F(\gamma)$ for an family of paths γ is an* ***isomorphism*** *on $V = \mathbb{R}^n$.*

Proof Since a path of length zero is assigned the identity linear map on V, by continuity there is an $\epsilon > 0$ such that the restriction of any path γ to $[0, \epsilon]$ is assigned an invertible morphism. Observe that this holds for any point on a given path (though possibly with variable ϵ). Choosing a finite subcover and factoring the value on a path into the value on pieces of the path subordinate to the subcover, we see that the value on a path is a composition of vector space isomorphisms, and therefore an isomorphism. □

The remaining work is in the construction of an inverse to the map $\mathsf{Vect}^\nabla(X) \to \mathrm{Tran}(X)$ (see Definition 4.0.2 and Lemma 4.1.2). For this we use the following two lemmas of Schreiber and Waldorf [22, Lemmas 4.1 and 4.2], reproduced here for convenience.

Lemma 4.2.3 *For a finite-dimensional vector space V, smooth functions*

$$F \colon \mathbb{R} \times \mathbb{R} \to \mathrm{Aut}(V)$$

satisfying the cocycle condition $F(y,z) \cdot F(x,y) = F(x,z)$ and $F(x,x) = \mathrm{id}$ are in bijection with 1–forms, $\Omega^1(\mathbb{R}; \mathrm{End}(V))$.

Proof Given such a 1–form A, consider the initial value problem

$$(\partial_t \alpha)(t) = A_t(\partial_t)(\alpha(t)), \quad \alpha(s) = \mathrm{id}, \tag{4.2.4}$$

where $\alpha : \mathbb{R} \to \mathrm{Aut}(V)$ and $s \in \mathbb{R}$. We obtain a unique solution $\alpha(t)$ depending on s, and define $F(s,t) = \alpha(t)$. The function F is smooth in s because the original coefficients were smooth in s, and is globally defined because the equation is linear. To verify that $F(s,t)$ satisfies the cocycle condition, we calculate

$$\partial_t(F(y,t)F(x,y)) = (\partial_t F(y,t))F(x,y) = A_t(\partial_t)F(y,t)F(x,y),$$

and since $F(y,y)F(x,y) = F(x,y)$, uniqueness dictates that $F(y,t)F(x,y) = F(x,t)$. Conversely, for $F : \mathbb{R} \times \mathbb{R} \to \mathrm{Aut}(V)$, let $\alpha(t) = F(s,t)$ for some $s \in \mathbb{R}$ and let

$$A_t(\partial_t) = (\partial_t \alpha(t))\alpha(t)^{-1}.$$

When F satisfies the cocycle condition, $A_t(\partial_t)$ is independent of the choice of s:

$$F(s_0,t) = F(s_1,t)F(s_0,s_1) \implies (\partial_t F(s_0,t))F(s_0,t)^{-1} = (\partial_t F(s_1,t))F(s_1,t)^{-1}.$$

This gives the desired bijection. □

Lemma 4.2.5 *Let $A, A' \in \Omega^1(\mathbb{R}; \mathrm{End}(V))$ be endomorphism valued 1–forms on $\mathbb{R}$ and $g : \mathbb{R} \to \mathrm{Aut}(V)$ be a smooth function. Let F_A and $F_{A'}$ be the smooth functions corresponding to A and A' by Lemma 4.2.3. Then*

$$g(y) \cdot F_A(x,y) = F_{A'}(x,y) \cdot g(x)$$

if and only if $A' = \mathrm{Ad}_{g^{-1}} A + g^{-1} dg$.

Proof The function $g(y)F_A(x,y)g(x)^{-1}$ solves the initial value problem (4.2.4) for A',

$$\begin{aligned}\partial_y(g(y)F(x,y)g(x)^{-1}) &= (\partial_y g(y))F(x,y)g(x)^{-1} + g(y)\partial_y F(x,y)g(x)^{-1} \\ &= (\partial_y g(y)g(y)^{-1})(g(y)F(x,y)g(x)^{-1}) \\ &\qquad + (g(y)A_y(\partial_y)g(y)^{-1})(g(y)F(x,y)g(x)^{-1}),\end{aligned}$$

so by uniqueness we obtain the desired equality. □

Now we construct a differential form from the parallel transport data that will give rise to a connection. Throughout, $(d,F) \in \mathrm{Tran}(X)$ is a transport data on $X \in \mathsf{Cart}$ with typical fiber $V = \mathbb{R}^d$. Let $\gamma : \mathbb{R} \to X$ be a path such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$; restrictions of γ to intervals (as a family over $\mathbb{R}^2$) will give a family of paths in $\mathcal{P}X$, ie a 0–simplex. Define

$$F_\gamma(x,y) = F(\gamma : [x,y] \to X), \qquad F : \mathbb{R} \times \mathbb{R} \to \mathrm{End}(V).$$

By the above lemma, F_γ gives us a 1–form A_γ with values in $\mathrm{End}(V)$. By varying γ, we want to promote this to a 1–form on X whose value at (p,v) is $A_\gamma(\partial_t)$.

Definition 4.2.6 The morphism $\mathrm{Tran} \to \mathsf{Vect}^{\nabla}$ is defined as follows. Fix $X \in \mathsf{Cart}$ and $(d, F) \in \mathrm{Tran}(X)$; we need to produce $A \in \Omega^1(X, \mathrm{End}(V))$, where $V = \mathbb{R}^d$. Fix a point $p \in X$ and a tangent vector $v \in X$; we need to define $A_p(v) \in \mathrm{End}(V)$. We use the linear structure on $X \cong \mathbb{R}^n$ to define $A_p(v) = (A_{tv})_0(1)$, where tv is the path through $p \in X$ with velocity vector $v \in X$, A_{tv} is a 1–form on $\mathbb{R}$, and $(A_{tv})_0(1)$ is the value of this 1–form at 0 evaluated at $1 \in T_0\mathbb{R}$. This defines a functor $\mathrm{Tran}(X) \to \mathsf{Vect}^{\nabla}(X)$ on objects, and on morphisms we send $h\colon X \to \mathrm{GL}(V)$ to itself, now viewed as a morphism in $\mathsf{Vect}^{\nabla}(X)$.

Lemma 4.2.7 *Definition 4.2.6 is well defined: objects in* Tran *are sent to smooth differential* 1*–forms* $A \in \Omega^1(X, \mathrm{End}(V))$, *and isomorphisms are sent to gauge transformations* $A' \mapsto \mathrm{Ad}_{g^{-1}} A + g^{-1} dg$.

Proof The claim on isomorphisms follows from Lemma 4.2.5. To verify the claim on objects, we first observe that A is smooth: choose families of affine paths in a neighborhood of p and invoke smoothness of the representation. Furthermore, we claim that A satisfies $A(\lambda v) = \lambda A(v)$ for all $\lambda > 0$. To see this, define $\gamma_\lambda(t) = \gamma(\lambda t)$ for $\lambda > 0$. We compute

$$A(\lambda v) = \partial_t F_{\gamma_\lambda}(0, t)|_{t=0} = \partial_t F_{\gamma}(0, \lambda t)|_{t=0} = \lambda A(v).$$

Lemma 4.2.8 shows that this property implies A is linear. □

Lemma 4.2.8 *A smooth function* $A\colon V \to W$ *between vector spaces that satisfies* $A(\lambda v) = \lambda A(v)$ *for* $\lambda > 0$ *is linear.*

Proof It suffices to show that A is equal to its derivative at zero. From the assumptions it follows that $A(0) = 0$. Smoothness of A implies that $dA(0)$ exists, and we compute its value on v by the one-sided limit

$$(dA(0))(v) = \lim_{\lambda \to 0^+} A(\lambda v)/\lambda = \lim_{\lambda \to 0^+} \lambda A(v)/\lambda = A(v),$$

completing the proof. □

The next lemma shows that A determines the given representation. Our techniques are in the spirit of D Freed's [11, Appendix B], though benefited from K Waldorf pointing out to us the utility of Hadamard's lemma in this context.

Lemma 4.2.9 *For* $X \in \mathsf{Cart}$ *and* $(d, F) \in \mathrm{Tran}(X)$, *the value of* F *on a path* γ *is the path-ordered exponential associated to the* $\mathrm{End}(V)$*–valued* 1*–form* A *constructed in Definition 4.2.6, where* $V = \mathbb{R}^d$.

Proof Let $X \cong \mathbb{R}^n$ and $\gamma\colon [0,T] \to \mathbb{R}^n$ be a path. Fix N a large integer, and let γ_i denote the restriction of γ to $[T(i-1)/N, Ti/N]$ for $1 \le i \le N$. By definition of $\mathrm{Tran}(X)$,

$$F(\gamma) = F(\gamma_N) \circ \cdots \circ F(\gamma_2) \circ F(\gamma_1).$$

Reparametrize γ_i by $\tilde{\gamma}_i(t) := \gamma_i(T(t+i-1)/N)$ and let $\ell_i\colon [0,T/N] \to \mathbb{R}^n$ denote the affine path of length 1 starting at $\gamma_i(0)$ with velocity $\dot{\gamma}_i(0) = v_i$. By Hadamard's lemma there is a smooth function g_i with $\gamma_i(s) - \ell_i(s) = s^2 g_i(s)$. Define $G\colon [0,1] \to \mathrm{End}(V)$ by $G(t) := F(\tilde{\gamma}_i|_{[0,t]})$. Using that $\tilde{\gamma}_i(s) = \ell_i(s(T/N)) + s^2(T^2/N^2) g_i(s(T^2/N^2))$ and applying Hadamard's lemma to G we obtain

$$G(t) = G(0) + tG'(0) + t^2 G_2(t) = \mathrm{id} + t(T/N)A_{\ell_i}(v_i) + O(N^{-2})$$

for some function $G_2\colon [0,1] \to \mathrm{End}(V)$. The $O(N^{-2})$ estimate comes from Taylor's formula and the fact that the original domain of definition $[0,T]$ is compact, so a uniform estimate can be given for the coefficient before $(T/N)^2$. The claimed form of the derivative $G'(0)$ follows from Lemma 4.2.8 and an argument in Schreiber and Waldorf [22, Lemma B.2] (reproduced in the next paragraph) to show that $A_{\gamma_i}(v_i) = A_{\ell_i}(v_i)$.

First we consider the family of paths $\Gamma(t,\alpha) := \ell_i(t) + \alpha g_i(t)$ depending on the parameter α for $0 \le \alpha \le 1$. Define $q\colon [0,1]^2 \to [0,1]^2$ by $(t,\alpha) \mapsto (t, t^2\alpha)$. The composition

$$(\Gamma \circ q)(t,\alpha) = \ell_i(t) + \alpha t^2 g_i(t)$$

defines a smooth homotopy running from ℓ_i (when $\alpha = 0$) to γ_i (when $\alpha = 1$). For a fixed α, we evaluate F on the family of paths $\Gamma \circ q$ obtained from restriction to $[0,t] \times \{\alpha\} \subset [0,1]^2$ and differentiate with respect to t using the chain rule,

$$\frac{d}{dt} F((\Gamma \circ q)|_{[0,t]\times\{\alpha\}})|_{t=0} = d(F(\Gamma))|_{q(0,\alpha)} \circ \frac{dq}{dt}\Big|_{t=0} = d(F(\Gamma))|_{(0,0)} \circ (1,0).$$

The right-hand side is independent of α, whereas the left-hand side is $A_{\ell_i}(v_i)$ for $\alpha = 0$ and $A_{\gamma_i}(v_i)$ when $\alpha = 1$, so the claim follows.

Putting this together, we have $F(\gamma_i) = \mathrm{id} + (T/N)A_{\ell_i}(v_i) + O(N^{-2})$, and taking $N \to \infty$,

$$\begin{aligned} F(\gamma) &= \lim_{N\to\infty} (\mathrm{id} + (T/N)A_{\ell_1}(v_1))(\mathrm{id} + (T/N)A_{\ell_2}(v_2)) \cdots (\mathrm{id} + (T/N)A_{\ell_N}(v_N)) \\ &= \lim_{N\to\infty} \exp((T/N)A_{\ell_1}(v_1)) \exp((T/N)A_{\ell_2}(v_2)) \cdots \exp((T/N)A_{\ell_N}(v_N)) \\ &= \mathcal{P}\exp(A(\dot{\gamma})), \end{aligned}$$

since the limit is the definition of the path-ordered exponential of A along γ. □

Proposition 4.2.10 *The morphism*

$$\mathsf{Tran} \to \mathsf{Vect}^{\nabla}$$

in PreStacks *constructed in Definition 4.2.6 is an isomorphism.*

Proof The inverse isomorphism is the composition

$$\mathsf{Vect}^{\nabla}(X) \to \mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}'X, \mathsf{Vect}) \to \mathsf{Tran}(X),$$

where the two maps are constructed in Definition 4.0.2 and Lemma 4.1.2. As shown in Lemma 4.2.9, for every $X \in \mathsf{Cart}$, the composition

$$\mathsf{Tran}(X) \to \mathsf{Vect}^{\nabla}(X) \to \mathsf{Tran}(X)$$

is an isomorphism on objects. Morphisms in $\mathsf{Tran}(X)$ and $\mathsf{Vect}^{\nabla}(X)$ were defined as smooth maps $X \to \mathrm{GL}(V)$ satisfying certain respective properties, and we have shown in Lemma 4.2.7 that these properties are preserved by these functors. □

5 Smooth 1–dimensional field theories and the cobordism hypothesis

We keep the notation of the previous section: $X \in \mathsf{Cart}$ is a cartesian space, on which we consider a field theory.

By construction, there is a map $\mathcal{P}X \to U(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X))$ (in the category C^{∞}–Cat of Definition 2.2.6, where U is the forgetful functor of Definition 2.3.4) that views a path as a bordism. This will allow us to apply arguments from the preceding section to the bordism category.

Lemma 5.0.1 *The value of a field theory — see Definition 3.0.1 — on a family of bordisms $\gamma\colon S \times M \to X$ as a vertex in $1\text{–}\mathsf{Bord}^{\mathrm{or}}_S(X)$ is equal to the value on a bordism $\gamma_{\mathrm{sit}}\colon S \times M \to X$ that has the same image in X as γ but has sitting instants, meaning that the map $\gamma_{\mathrm{sit}}\colon M \to X$ is constant near t_0 and t_1.*

Proof Using the Γ–structure, it suffices to prove the lemma for arcs in X, ie S–families $\gamma\colon S \times I \to X$ for I an interval. Choose $b\colon \mathbb{R} \to \mathbb{R}$ to be a smooth bump function such that $b|_{(-\infty,1/3]} = 0$, $b|_{[2/3,\infty)} = 1$, and $b|_{(1/3,2/3)} \subset (0,1)$. Consider a new $(S\times\mathbb{R})$–family of 1–manifolds that, for $t \in \mathbb{R}$, is given by $\gamma_{\mathrm{sit}} := \gamma \circ \Gamma(x,t)$, where $\Gamma(x,t) = tx + (1-t)b(x)$ for $x \in I$. To this family a field theory assigns a smooth family of linear maps. We observe that for all $t \in (0,1]$, the fibers in this family are isomorphic as morphisms in the fiber of $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ over $S = \mathrm{pt} \in \mathsf{Cart}$. Therefore, a field theory assigns the *same* linear maps for all $t \neq 0$. By smoothness, we obtain the same linear map at $t = 0$ and the resulting path has sitting instants around 0 and 1 by construction. □

We are now ready to prove Theorem B.

Theorem D *Evaluating at* $\langle 1\rangle \in \Gamma$ *and restricting along* $\mathcal{P}X \to U(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X))$ *yields a weak equivalence of derived mapping simplicial sets, natural in* $X \in \mathsf{Mfld}$ (*and, more generally*, $X \in \mathrm{Fun}(\mathsf{Cart}^{\mathsf{op}}, \mathsf{Set})$),

$$\mathsf{C}^{\infty}\text{–}\mathsf{Cat}^{\otimes}(1\text{–}\mathsf{Bord}^{\mathrm{or}}(X), \mathsf{Vect}^{\otimes}) \to \mathsf{C}^{\infty}\text{–}\mathsf{Cat}(\mathcal{P}X, \mathsf{Vect}).$$

Thus, there is an equivalence between 1*–dimensional oriented topological field theories over* X *valued in* $\mathsf{Vect}^{\otimes}$ *and* C^{∞}*–functors from the smooth path category of* X *to* Vect.

Proof Applying Theorem C, we reduce the problem to the case $X \in \mathsf{Cart}$. Applying Remark 3.2.4, we replace $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ with $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)$, for which we have $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)([0], \langle 1\rangle) = X \sqcup X$, corresponding to two possible orientations of points in X. These two copies of X a priori map to some $d, d' \in \mathbb{N} = \mathsf{Vect}(X)([0], \langle 1\rangle)$, which uniquely determines the maps on $1\text{–}\mathsf{Bord}'^{\mathrm{or}}(X)([0], \langle m\rangle)$ for $m \neq 1$. As shown below, we necessarily have $d = d'$, which corresponds to the dimension of the vector bundle determined by the field theory.

To understand the value of a field theory on morphisms, since the target category $\mathsf{Vect}^{\otimes}$ is fibrant (Lemma 2.3.3), in particular, satisfies the Segal Δ–condition, a functor $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X) \to \mathsf{Vect}$ is determined (up to a contractible choice) by its value over the fiber $[1] \in \Delta$. Furthermore, since any bordism can be expressed as a disjoint union of connected bordisms, we can restrict attention to S–families of connected 1–manifolds in $1\text{–}\mathsf{Bord}^{\mathrm{or}}_S(X)(1)$.

In the case that cut functions satisfy $t_0 < t_1$, Morse theory of 1–manifolds cuts a given connected bordism into elementary pieces that are of three types: (1) bordism from a point to a point (all points of $M^1 \times \{s\}$ are regular values for h), (2) bordisms from the empty set to a pair of points (0–handles), and (3) bordisms from a pair of points to the empty set (1–handles). For a given bordism, ie 0–simplex of $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$, this reduction comes from a choice of (new) height function that is Morse with regular values at the prescribed cut values, which defines a 1–simplex in $1\text{–}\mathsf{Bord}^{\mathrm{or}}(X)$ connecting the original bordism to one with a Morse height function. Then we can impose additional cut points using the Morse height function to reduce to the cases above. The relations among these generators are precisely the familiar birth-death diagrams from 1–dimensional Morse theory.

When cut functions satisfy $t_0 = t_1$, since t_0 is a regular value and the bordism is connected, this bordism is in the image of the degeneracy map, ie is an identity path in the bordism category. For S connected and $t_0 \leq t_1$ with $t_0 = t_1$ somewhere on S, then this is necessarily a bordism of type (1) above.

In the case that the above types of generating bordisms are mapped constantly to X, meaning the map $x\colon M^1 \times S \to X$ factors through the projection to S, the standard dualizable object argument shows that the value of the field theory on the $(+)$–point must be a vector space $(V_+)_x = \mathbb{R}^d$, and the value on the $(-)$–point is the dual space, $(V_-)^*_x = \mathbb{R}^{d'}$, which in our formulation amounts to showing $d = d'$.

Now we need to show that the value on a generating bordism with an arbitrary map to X is determined by the value of the field theory on the path category. For generating bordisms of type (1) this is clear, since such a bordism can be identified with a morphism in the path category.

For bordisms of type (2) and (3) we use Lemma 5.0.1 to identify the value of a field theory on a 0– or 1–handle with the value on a handle that has a sitting instant at its Morse critical point. Then we can factor the handle into 3 pieces: one given by a subset of the sitting instant of the Morse critical points (ie a handle that is mapped constantly to X) and two paths given by the closure of the complement of this subset in the original handle. Hence, the value of the original bordism is determined by previously computed dualizing data at the sitting instant together with the value on paths between points. □

Proof of Theorem A The result follows from Theorem B and Proposition 4.0.1. □

6 Reedy model structures

In this auxiliary section we review the necessary facts from the theory of Reedy model structures.

Let C be a model category. Following Hirschhorn [15, Definition 15.3.3], we review the Reedy model structure on the category of simplicial objects in C, ie the category of functors $\Delta^{\mathrm{op}} \to C$.

First, define the n^{th} *latching functor*, $L_n \colon \mathrm{Fun}(\Delta^{\mathrm{op}}, C) \to C$ as

$$L_n X = \operatorname*{colim}_{[m]\to[n]} X_m,$$

where the colimit is indexed by surjections of finite ordered sets $[m] \leftarrow [n]$ that are not isomorphisms (ie the union of degenerate simplices). Similarly, define the n^{th} *matching functor* $M_n \colon \mathrm{Fun}(\Delta^{\mathrm{op}}, C) \to C$ as

$$M_n X = \lim_{[m]\leftarrow[n]} X_m,$$

where the limit is indexed by injections of finite ordered sets $[m] \to [n]$ that are not isomorphisms (ie the defining data of a boundary of an n–simplex).

Now, in the Reedy model structure on simplicial objects in C, a map $X \to Y$ is a cofibration if

$$X_n \sqcup_{L_n X} L_n Y \to Y_n$$

is a cofibration in C for any $[n] \in \Delta$. Similarly, a map $X \to Y$ is a fibration if

$$X_n \to Y_n \times_{M_n Y} M_n X$$

is a fibration for any $[n] \in \Delta$. In particular, an object X is cofibrant if the latching map $L_n X \to X$ is a cofibration in C for any $[n]$ and fibrant if the matching map $X \to M_n X$ is a fibration in C for any $[n] \in \Delta$.

By Hirschhorn [15, Theorem 15.6.27] the Reedy model is cofibrantly generated, with generating (acyclic) cofibrations as in Hirschhorn [15, Definition 15.6.23]: if $A \to B$ is a generating (acyclic) cofibration in C, then $A \otimes [n] \sqcup_{A\otimes\partial[n]} B \otimes \partial[n] \to B \otimes [n]$ is a generating (acyclic) cofibration of the Reedy model structure.

References

[1] **J C Baez**, **J Dolan**, *Higher-dimensional algebra and topological quantum field theory*, J. Math. Phys. 36 (1995) 6073–6105 MR Zbl

[2] **C Barwick**, *On left and right model categories and left and right Bousfield localizations*, Homology Homotopy Appl. 12 (2010) 245–320 MR Zbl

[3] **M Benini**, **M Perin**, **A Schenkel**, *Smooth 1–dimensional algebraic quantum field theories*, Ann. Henri Poincaré 23 (2022) 2069–2111 MR Zbl

[4] **C Berger**, **I Moerdijk**, *On an extension of the notion of Reedy category*, Math. Z. 269 (2011) 977–1004 MR Zbl

[5] **J E Bergner**, **C Rezk**, *Reedy categories and the Θ–construction*, Math. Z. 274 (2013) 499–514 MR Zbl

[6] **A K Bousfield**, **E M Friedlander**, *Homotopy theory of Γ–spaces, spectra, and bisimplicial sets*, from "Geometric applications of homotopy theory, II" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 658, Springer (1978) 80–130 MR Zbl

[7] **U Bunke**, **P Turner**, **S Willerton**, *Gerbes and homotopy quantum field theories*, Algebr. Geom. Topol. 4 (2004) 407–437 MR Zbl

[8] **A Caetano**, **R F Picken**, *An axiomatic definition of holonomy*, Internat. J. Math. 5 (1994) 835–848 MR Zbl

[9] **D Calaque**, **C Scheimbauer**, *A note on the (∞, n)–category of cobordisms*, Algebr. Geom. Topol. 19 (2019) 533–655 MR Zbl

[10] **D Dugger**, **S Hollander**, **D C Isaksen**, *Hypercovers and simplicial presheaves*, Math. Proc. Cambridge Philos. Soc. 136 (2004) 9–51 MR Zbl

[11] **D S Freed**, *Classical Chern–Simons theory, I*, Adv. Math. 113 (1995) 237–303 MR Zbl

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

[13] **D Grady**, **D Pavlov**, *Extended field theories are local and have classifying spaces*, preprint (2020) arXiv 2011.01208v2

[14] **D Grady**, **D Pavlov**, *The geometric cobordism hypothesis*, preprint (2021) arXiv 2111.01095v1

[15] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[16] **J F Jardine**, *Simplicial presheaves*, J. Pure Appl. Algebra 47 (1987) 35–87 MR Zbl

[17] **S Kobayashi**, *La connexion des variétés fibrées*, *II*, C. R. Acad. Sci. Paris 238 (1954) 443–444 MR Zbl

[18] **M Ludewig**, **A Stoffel**, *A framework for geometric field theories and their classification in dimension one*, Symmetry Integrability Geom. Methods Appl. 17 (2021) art. no. 072 MR Zbl

[19] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl

[20] **J Lurie**, *On the classification of topological field theories*, from "Current developments in mathematics, 2008" (D Jerison, B Mazur, T Mrowka, W Schmid, R Stanley, S-T Yau, editors), International, Somerville, MA (2009) 129–280 MR Zbl

[21] **C Rezk**, *A model for the homotopy theory of homotopy theory*, Trans. Amer. Math. Soc. 353 (2001) 973–1007 MR Zbl

[22] **U Schreiber**, **K Waldorf**, *Parallel transport and functors*, J. Homotopy Relat. Struct. 4 (2009) 187–244 MR Zbl See the updated version in arXiv 0705.0452v5

[23] **G Segal**, *Categories and cohomology theories*, Topology 13 (1974) 293–312 MR Zbl

[24] **G Segal**, *Elliptic cohomology (after Landweber–Stong, Ochanine, Witten, and others)*, from "Séminaire Bourbaki, 1987/88", Astérisque 161–162, Soc. Mat. France, Paris (1988) Exposé 695, 187–201 MR Zbl

[25] **S Stolz**, **P Teichner**, *What is an elliptic object?*, from "Topology, geometry and quantum field theory" (U Tillman, editor), London Math. Soc. Lecture Note Ser. 308, Cambridge Univ. Press (2004) 247–343 MR Zbl

[26] **S Stolz**, **P Teichner**, *Supersymmetric field theories and generalized cohomology*, from "Mathematical foundations of quantum field theory and perturbative string theory" (H Sati, U Schreiber, editors), Proc. Sympos. Pure Math. 83, Amer. Math. Soc., Providence, RI (2011) 279–340 MR Zbl

Department of Mathematics, University of Illinois at Urbana-Champaign
Urbana, IL, United States

Department of Mathematics and Statistics, Texas Tech University
Lubbock, TX, United States

danbe@illinois.edu

https://danbe.web.illinois.edu/, https://dmitripavlov.org/

Received: 16 August 2021 Revised: 23 December 2021

Algebraic & Geometric Topology **23**:8 (2023) 3745–3762
DOI: 10.2140/agt.2023.23.3745
Published: 5 November 2023

Round fold maps on 3–manifolds

NAOKI KITAZAWA
OSAMU SAEKI

We show that a closed orientable 3–dimensional manifold admits a round fold map into the plane, ie a fold map whose critical value set consists of disjoint simple closed curves isotopic to concentric circles, if and only if it is a graph manifold, generalizing the characterization for simple stable maps into the plane. Furthermore, we also give a characterization of closed orientable graph manifolds that admit directed round fold maps into the plane, ie round fold maps such that the number of regular fiber components of a regular value increases toward the central region in the plane.

57R45; 57K30, 58K30

Dedicated to Professor Kazuhiro Sakuma on the occasion of his 60th birthday

1 Introduction

Let M be a smooth closed manifold of dimension ≥ 2. It is known that if a smooth map $f\colon M \to \mathbb{R}^2$ is generic enough, then it has only fold and cusps as its singularities; see Levine [9; 10] and Whitney [16]. Furthermore, if M has even Euler characteristic (eg if $\dim M$ is odd), then the cusps can be eliminated by homotopy. In particular, every smooth closed orientable 3–dimensional manifold admits a smooth map into $\mathbb{R}^2$ with only fold singularities, ie a *fold map*.

In [13; 14], the second author considered the following smaller class of generic smooth maps. A fold map $f\colon M \to \mathbb{R}^2$ on a smooth closed orientable 3–dimensional manifold M is a *simple stable map* if for every $q \in \mathbb{R}^2$, each component of $f^{-1}(q)$ contains at most one singular point and $f|_{S(f)}$ is an immersion with normal crossings, where $S(f)(\subset M)$ denotes the set of singular points of f. Note that if f is a fold map, then $S(f)$ is a regular closed 1–dimensional submanifold of M. In particular, if $f|_{S(f)}$ is an embedding, then f is a simple stable map. In [14], it has been proved that for a smooth closed orientable 3–dimensional manifold M, the following three properties are equivalent to each other:

(1) M admits a fold map $f\colon M \to \mathbb{R}^2$ such that $f|_{S(f)}$ is an embedding.

(2) M admits a simple stable map into $\mathbb{R}^2$.

(3) M is a graph manifold, ie it is a finite union of S^1–bundles over compact surfaces attached along their torus boundaries.

Thus, for example, if M is hyperbolic, then M never admits such a fold map.

On the other hand, the first author introduced the notion of a round fold map [7; 6; 5]: a smooth map $f\colon M \to \mathbb{R}^2$ is a *round fold map* if it is a fold map and $f|_{S(f)}$ is an embedding onto the disjoint union of some concentric circles in $\mathbb{R}^2$; for details, see Section 2. As has been studied by the first author, round fold maps have various nice properties.

The first main result of this paper is Theorem 3.1, which states that every graph 3–manifold admits a round fold map into $\mathbb{R}^2$. This generalizes the characterization result obtained in [14] for simple stable maps mentioned above.

It is not difficult to observe that if $f\colon M \to \mathbb{R}^2$ is a round fold map of a closed orientable 3–dimensional manifold, then the number of components of the fiber over a regular value changes exactly by one when the regular value crosses the critical value set transversely once. We can thus put a normal orientation to each component of the critical value set in such a way that the orientation points in the direction that increases the number of components of a regular fiber. Then, a round fold map is said to be *directed* if all the circles in the critical value set are directed inward. The second main result of this paper (Theorem 3.2) characterizes those graph 3–manifolds which admit directed round fold maps. It will turn out that the class is strictly smaller than that of closed orientable graph 3–manifolds.

The paper is organized as follows. In Section 2, we prepare several definitions and a lemma concerning round fold maps and graph 3–manifolds necessary for our purposes. We also give an observation on fibered links or open book structures associated with round fold maps and give some examples. In Section 3, we state and prove the main theorems. Basically, we will follow the proof given in [14, Theorem 3.1]: however, in some steps we need to modify the strategy for the constructions of round fold maps. In Section 4, we give some corollaries and show that the class of 3–manifolds that admit directed round fold maps is strictly smaller than that of all graph 3–manifolds, using results obtained in [2; 12]. Finally, we give some open problems related to our results.

Throughout the paper, all manifolds and maps between them are smooth of class C^∞ unless otherwise specified. For a space X, id_X denotes the identity map of X. The symbol $\cong$ denotes a diffeomorphism between smooth manifolds.

2 Preliminaries

2.1 Round fold maps

In this subsection, we recall some notions related to round fold maps and give some examples.

Let M be a closed orientable 3–dimensional manifold and $f\colon M \to \mathbb{R}^2$ a smooth map. We denote by $S(f)$ $(\subset M)$ the set of all singular points of f.

Definition 2.1 A point $p \in S(f)$ is a *definite fold point* (resp. an *indefinite fold point*, or a *cusp point*) if f is represented by the map

$$(u,x,y) \mapsto (u, x^2+y^2) \quad \text{(resp. } (u, x^2-y^2)\text{, or } (u, y^2+ux-x^3)\text{)}$$

around the origin with respect to certain local coordinates around p and $f(p)$. We call a point $p \in S(f)$ a *fold point* if it is a definite or an indefinite fold point. A smooth map $f\colon M \to \mathbb{R}^2$ is called a *fold map* if it has only fold points as its singular points. Note that then $S(f)$ is a closed 1–dimensional submanifold of M and that $f|_{S(f)}$ is an immersion.

Definition 2.2 Let $C^\infty(M,\mathbb{R}^2)$ denote the space of all smooth maps of M into $\mathbb{R}^2$, endowed with the Whitney C^∞–topology. A smooth map $f\colon M \to \mathbb{R}^2$ is a *stable map* if there exists a neighborhood $N(f)$ of f in $C^\infty(M,\mathbb{R}^2)$ such that for every $g \in N(f)$, there are diffeomorphisms $\Phi\colon M \to M$ and $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ such that $g = \varphi \circ f \circ \Phi^{-1}$.

It is known that a smooth map $f\colon M \to \mathbb{R}^2$ is a stable map if and only if the following three conditions hold; for example, see Levine [10].

(1) It has only fold and cusp points as its singularities.

(2) We have $f^{-1}(f(q)) \cap S(f) = \{q\}$ for every cusp point $q \in S(f)$.

(3) The restriction of f to the set of fold points is an immersion with normal crossings.

Definition 2.3 A stable map $f\colon M \to \mathbb{R}^2$ is *simple* if it has no cusp points and for every $q \in \mathbb{R}^2$, each component of $f^{-1}(q)$ contains at most one singular point; for details, see [13; 14].

In the following, for $r > 0$, C_r denotes the circle of radius r centered at the origin in $\mathbb{R}^2$.

Definition 2.4 A finite disjoint union of simple closed curves in $\mathbb{R}^2$ is said to be *concentric* if it is isotopic to a set of concentric circles

$$\bigcup_{i=1}^{m} C_i$$

for some positive integer m.

Definition 2.5 We say that a smooth map $f\colon M \to \mathbb{R}^2$ is a *round fold map* if it is a fold map and $f|_{S(f)}$ is an embedding onto a concentric family of simple closed curves. Note that a round fold map is a simple stable map. Note also that the outermost circle component of $f(S(f))$ consists of the images of definite fold points.

By composing a diffeomorphism of $\mathbb{R}^2$ if necessary, we always assume that a round fold map $f\colon M \to \mathbb{R}^2$ satisfies

$$f(S(f)) = \bigcup_{i=1}^{m} C_i \tag{2-1}$$

for some positive integer m.

In the following, A denotes the annulus $S^1 \times [-1,1]$, and P denotes the compact surface obtained from the 2–sphere by removing three open disks: in other words, P is a *pair of pants*.

Let $f\colon M \to \mathbb{R}^2$ be a round fold map of a closed *orientable* 3–dimensional manifold M. For a component C of $f(S(f))$, take a small arc $\alpha \cong [-1,1]$ in $\mathbb{R}^2$ that intersects $f(S(f))$ exactly at one point in C transversely. Then, $f^{-1}(\alpha)$ is a compact surface with boundary $f^{-1}(a) \cup f^{-1}(b)$, which is diffeomorphic to a finite disjoint union of circles, where a and b are the end points of α. Furthermore, $f|_{f^{-1}(\alpha)}\colon f^{-1}(\alpha) \to \alpha$ can be regarded as a Morse function with exactly one critical point. As M is orientable, we see that $f^{-1}(\alpha)$ is diffeomorphic to the union of D^2 (or P) and a finite number of copies of A; see [15], for example. Therefore, the number of components of $f^{-1}(a)$ differs from that of $f^{-1}(b)$ exactly by one. If $f^{-1}(a)$ has more components than $f^{-1}(b)$, then we normally orient C from b to a; otherwise, we orient C from a to b. It is easily shown that this normal orientation is independent of the choice of α. In this way, each component of $f(S(f))$ is normally oriented. If the normal orientation points inward, then the component is said to be *inward-directed*; otherwise, it is *outward-directed*.

Definition 2.6 Let $f\colon M \to \mathbb{R}^2$ be a round fold map. We say that f is *directed* if all the components of $f(S(f))$ are inward-directed. It is easy to see that a round fold map f is directed if and only if the number of components of a regular fiber over a point in the innermost component of $\mathbb{R}^2 \setminus f(S(f))$ coincides with the number of components of $S(f)$.

Let $f\colon M \to \mathbb{R}^2$ be a round fold map satisfying (2-1). Set $L = f^{-1}(0)$, which is an oriented link in M if it is not empty. Let D be the closed disk centered at the origin with radius $\frac{1}{2}$. Then, $f^{-1}(D)$ is diffeomorphic to $L \times D$, which can be identified with a tubular neighborhood $N(L)$ of L in M. Furthermore, the composition $\varphi = \pi \circ f\colon M \setminus \operatorname{Int} N(L) \to S^1$ is a submersion, where $\pi\colon \mathbb{R}^2 \setminus \operatorname{Int} D \to S^1$ is the standard radial projection and $\varphi|_{\partial N(L)}\colon \partial N(L) = L \times \partial D \to S^1$ corresponds to the projection to the second factor followed by a scalar multiplication. Hence, φ is a smooth fiber bundle and L is a fibered link. (In other words, M admits an open book structure with binding L.) The fiber (or the page) is identified with $F = f^{-1}(J)$, where

$$J = \left[\tfrac{1}{2}, m+1\right] \times \{0\} \subset \mathbb{R}^2,$$

and it is a compact oriented surface. Note that $g = f|_F\colon F \to J$ is a Morse function with exactly m critical points, and that a monodromy diffeomorphism of the fibration over S^1 can be chosen so that it preserves g.

Note that all these arguments work even when $L = \varnothing$. In this case, F is a closed orientable surface and M is the total space of an F–bundle over S^1.

Conversely, if we have a compact orientable surface F, a Morse function $g\colon F \to \left[\frac{1}{2}, m+1\right]$ such that $g(\partial F) = \frac{1}{2}$ and that g has no critical point near the boundary, and a diffeomorphism $h\colon F \to F$ which is the identity on the boundary and which satisfies $g \circ h = g$, then we can construct a round fold map $f\colon M \to \mathbb{R}^2$ in such a way that M is the union of $\partial F \times D^2$ and the total space of the F–bundle over S^1 with geometric monodromy h.

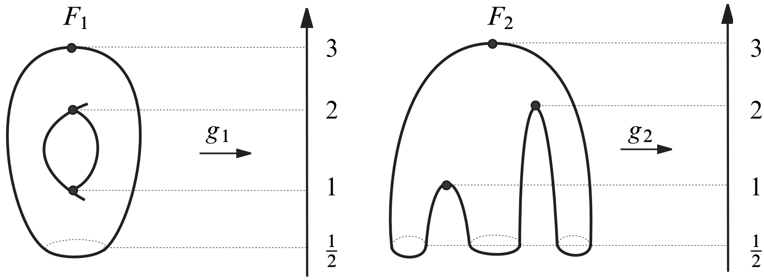


Figure 1: Morse functions on surfaces with Euler characteristic -1.

Example 2.7 Let F be a compact connected orientable surface with $\partial F \neq \varnothing$. Let us consider the identity diffeomorphism as the geometric monodromy in the above construction. Then, we see that the source 3–manifold M of the round fold map is diffeomorphic to $(\partial F \times D^2) \cup (F \times S^1) \cong \partial(F \times D^2)$. By using a handle decomposition argument, we can easily see that $F \times D^2$ is diffeomorphic to D^4 or a boundary connected sum of a finite number of copies of $S^1 \times D^3$. Therefore, M is diffeomorphic either to S^3 or to the connected sum of a finite number of copies of $S^1 \times S^2$.

For example, if we start with the Morse function $g_1 \colon F_1 \to \left[\frac{1}{2}, 4\right]$ as depicted in Figure 1, left, then the singular point set $S(f_1)$ of the resulting round fold map $f_1 \colon M_1 \to \mathbb{R}^2$ has three components and their images coincide with C_1, C_2 and C_3. The first one is outward directed, while the other two are inward directed. Therefore, the fold map f_1 is not directed. In this example, M_1 is diffeomorphic to $(S^1 \times S^2) \sharp (S^1 \times S^2)$.

On the other hand, if we start with the Morse function $g_2 \colon F_2 \to \left[\frac{1}{2}, 4\right]$ as depicted in Figure 1, right, then we get a round fold map $f_2 \colon M_2 \to \mathbb{R}^2$ with the same singular values: however, this round fold map is directed. We can also show that M_2 is again diffeomorphic to $(S^1 \times S^2) \sharp (S^1 \times S^2)$.

2.2 Graph 3–manifolds

In this subsection, we recall the notion of graph 3–manifolds and related results necessary to state our main theorems of the paper, and to get further results.

Definition 2.8 Let M be a compact orientable 3–dimensional manifold possibly with torus boundaries. It is called a *graph manifold* if it is diffeomorphic to a union of S^1–bundles over compact surfaces attached along their torus boundaries.

Let M be a graph manifold. For a boundary component of each S^1–bundle piece, we have a pair of distinguished simple closed curves: an S^1–fiber and a cross-section over the corresponding boundary component of the base surface with respect to a fixed trivialization. Note that such a pair of simple closed curves are unique up to isotopy once we fix a trivialization of the S^1–bundle over the boundary of the base surface. More precisely, the manifold M is oriented, and when the base surface is orientable, we orient the simple closed curves in a way consistent with the orientations of the base surface and the ambient 3–manifold. A decomposition of M as in Definition 2.8 is said to be of *plumbing type* if for each gluing of a pair of torus boundaries, the attaching diffeomorphism, which is orientation reversing, interchanges the S^1–fiber and the cross-section over the corresponding boundary component of the base surface. It is well-known that every graph manifold admits a decomposition of plumbing type; for example, see Hirzebruch, Neumann and Koh [4].

Each decomposition of plumbing type can be represented by a weighted graph: each vertex corresponds to an S^1–bundle piece over a connected surface and each edge corresponds to a gluing of the corresponding torus boundaries. A vertex is weighted with the genus of the base surface together with its orientability and the Euler number of the S^1–bundle. Furthermore, an edge is weighted by a sign $+$ or $-$ corresponding to the orientation preserving/reversing property of the gluing map on the pair of an S^1–fiber and a cross-section. Then, Neumann [12] listed up certain operations to weighted graphs in such a way that the two graph 3–manifolds corresponding to two weighted graphs are diffeomorphic if and only if the graphs are related by a finite iteration of the operations. Neumann also established the notion of normal form plumbing graphs as a special class of weighted graphs as above and showed that every graph 3–manifold has a unique normal form plumbing graph.

Now, in order to state one of our main theorems, we need the following.

Lemma 2.9 *Every closed orientable graph 3–manifold is diffeomorphic to a union of finite numbers of copies of $P \times S^1$ and the solid torus $D^2 \times S^1$ attached along their torus boundaries.*

Proof It is known that every such 3–manifold is diffeomorphic to a union of a finite number of S^1–bundles over compact connected orientable surfaces of genus zero attached along their torus boundaries; for example, see [14, Lemma 3.3]. In fact, if a base surface is nonorientable, then we can decompose the surface into the union of a compact orientable surface of genus zero and some copies of the Möbius band attached along their boundaries, and we see that the S^1–bundle over the Möbius band can be further decomposed into the union of S^1–bundles over compact orientable surfaces of genus zero. If a base surface is orientable of positive genus, then we can decompose it into the union of compact orientable surfaces of genus zero attached along their boundaries. Then, we can decompose the 3–manifold accordingly so that we obtain a desired decomposition.

Now, consider a base surface B, which is orientable of genus zero. If the number of boundary components is greater than or equal to 4, then we can decompose B into a union of a finite number of copies of P attached along their circle boundaries. If the number of boundary components is equal to two, then B is diffeomorphic to the union of P and a disk. If the surface B has no boundary, then we can decompose it into two disks. As orientable S^1–bundles over P or a disk are always trivial, the result follows. □

As a consequence, a graph manifold can be represented by a (multi-)graph, where each vertex corresponds to $P \times S^1$ or a solid torus and each edge corresponds to the gluing along a pair of boundary components. Note that each gluing corresponds to an element of the (orientation preserving) mapping class group of the torus, identified with $SL(2, \mathbb{Z})$.

3 Main theorems and proofs

In this section, we state our main results, Theorems 3.1 and 3.2, of this paper and give their proofs.

Theorem 3.1 *Let M be a closed orientable 3–dimensional manifold. Then, it admits a round fold map into $\mathbb{R}^2$ if and only if it is a graph manifold.*

Theorem 3.1 generalizes the characterization for simple stable maps obtained in [14].

In particular, every closed orientable graph 3–manifold admits a fibered link which is also a graph link. Compare this with Myers [11]. Here, a link in a graph 3–manifold is a *graph link* if its exterior is a graph manifold.

Theorem 3.2 *Let M be a closed connected orientable graph 3–manifold. Then, it admits a directed round fold map into $\mathbb{R}^2$ if and only if it can be decomposed into a union of finite numbers of copies of $P \times S^1$ and $D^2 \times S^1$ such that the corresponding graph is a tree.*

Proof of Theorem 3.1 As noted above, a round fold map is a simple stable map. Therefore, if a closed orientable 3–dimensional manifold admits such a map, then it is necessarily a graph manifold by [14].

Now, suppose M is a graph manifold. We will follow the proof of [14, Theorem 3.1] in order to construct a round fold map $f: M \to \mathbb{R}^2$, except for the first step, in which a nonsingular map is constructed for each S^1–bundle piece in [14] while we construct a fold map for each piece, as explained below.

By virtue of Lemma 2.9, we have disjointly embedded tori $T_1, T_2, \dots, T_\ell$ in M such that each of the components $X_1, X_2, \dots, X_k$ of $M \setminus \bigsqcup_{i=1}^{\ell} \operatorname{Int} N(T_i)$ is diffeomorphic either to $P \times S^1$ or to $D^2 \times S^1$, where $N(T_i)$ denotes a small tubular neighborhood of T_i in M, $1 \le i \le \ell$. By inserting pieces diffeomorphic to $A \times S^1 \cong T^2 \times [-1, 1]$ if necessary, we may assume that the decomposition is of plumbing type (for details, see [4]), where $T^2 = S^1 \times S^1$ denotes the torus. Now, each X_i is diffeomorphic either to $P \times S^1$, $D^2 \times S^1$ or $A \times S^1$.

Take a component X_j, for some $1 \le j \le k$. Suppose it is diffeomorphic to $D^2 \times S^1$. Let $\delta: D^2 \to [-1, 1]$ be the Morse function defined by $\delta(x, y) = -x^2 - y^2$, where D^2 is identified with the unit 2–disk in $\mathbb{R}^2$; see Figure 2, left. Then, define $f|_{X_j}$ to be the composition

$$\eta_j \circ (\delta \times \mathrm{id}_{S^1}) \circ \varphi_j : X_j \xrightarrow{\varphi_j} D^2 \times S^1 \xrightarrow{\delta \times \mathrm{id}_{S^1}} [-1, 1] \times S^1 \xrightarrow{\eta_j} \mathbb{R}^2,$$

where φ_j is a diffeomorphism and η_j is an embedding whose image is a small tubular neighborhood of the circle of radius j centered at the origin. We also arrange η_j in such a way that $\eta_j(\{\pm 1\} \times S^1)$ coincides with the circle of radius $j \pm \frac{1}{3}$.

Suppose X_j is diffeomorphic to $P \times S^1$. We define $f|_{X_j}$ by the composition

$$\eta_j \circ (\iota \times \mathrm{id}_{S^1}) \circ \varphi_j : X_j \xrightarrow{\varphi_j} P \times S^1 \xrightarrow{\iota \times \mathrm{id}_{S^1}} [-1, 1] \times S^1 \xrightarrow{\eta_j} \mathbb{R}^2,$$

where φ_j is a diffeomorphism, $\iota: P \to [-1, 1]$ is the standard Morse function with exactly one saddle point (as depicted in Figure 2, right) and η_j is an embedding as described in the previous paragraph.

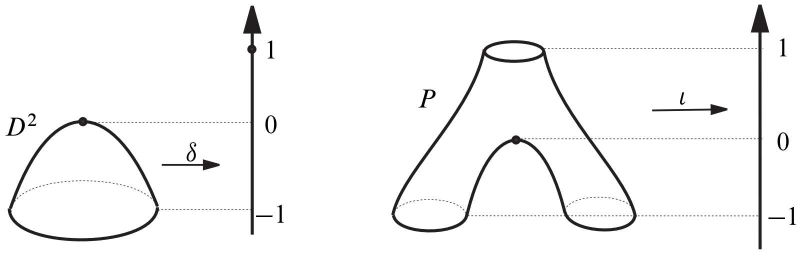


Figure 2: Morse functions δ and ι.

Finally, suppose X_j is diffeomorphic to $A \times S^1$. In this case, we define $f|_{X_j}$ by the composition

$$\eta_j \circ (\rho \times \mathrm{id}_{S^1}) \circ \varphi_j : X_j \xrightarrow{\varphi_j} A \times S^1 \xrightarrow{\rho \times \mathrm{id}_{S^1}} [-1,1] \times S^1 \xrightarrow{\eta_j} \mathbb{R}^2,$$

where φ_j is a diffeomorphism, $\rho\colon A \cong S^1 \times [-1,1] \to [-1,1]$ is the projection to the second factor and η_j is an embedding as described above.

Now, the map $f|_{\bigsqcup_{j=1}^k X_j}$ has only fold singular points, and its restriction to the singular point set is an embedding onto a concentric family of circles in $\mathbb{R}^2$. Then, we can extend the map to get a round fold map $f : M \to \mathbb{R}^2$ as follows, by a method similar to that used in [14, Proof of Theorem 3.1].

In the following, we set $I = [0,1]$. For the construction, we need two smooth maps h_1 and $h_2\colon T^2 \times I \to \mathbb{R}^2$ as follows. (For details, see [14, Section 3].) The map h_1 is constructed as the composition

$$T^2 \times I \xrightarrow{\cong} A \times S^1 \xrightarrow{\upsilon \times \mathrm{id}_{S^1}} [-1,1] \times S^1 \xrightarrow{\eta} \mathbb{R}^2,$$

where $\upsilon\colon A \to [-1,1]$ is a Morse function with exactly one saddle point and one maximum point such that $\upsilon^{-1}(-1) = \partial A$ (see [14, Figure 8]), and η is an embedding.

On the other hand, the smooth map $h_2\colon T^2 \times I \to \mathbb{R}^2$ enjoys the following properties.

(1) The image of h_2 is the disk of radius 3 centered at the origin.

(2) The singular point set $S(h_2)$ is a circle and consists of indefinite fold points.

(3) The map $h_2|_{S(h_2)}$ is an embedding onto the circle C_2.

(4) The inverse image $(h_2)^{-1}(C_3)$ coincides with a boundary component of $T^2 \times I$, and $(h_2)^{-1}(C_1)$ consists of two components one of which coincides with the other boundary component.

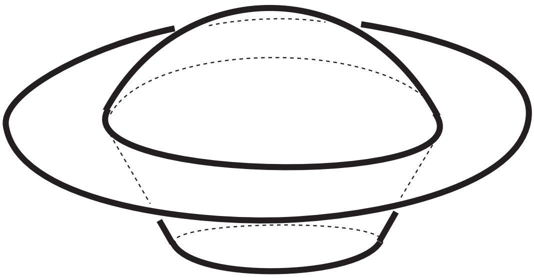


Figure 3: Quotient space of h_2.

(5) The quotient space in the Stein factorization of h_2 is a 2–dimensional polyhedron as depicted in Figure 3. (For the definition of the quotient space in the Stein factorization, see [14, Section 2], for example.)

Here, we omit the detailed construction of h_2, as it is fully explained in [14]. The idea is to use a Dehn surgery on the exterior of a 2–component trivial link in S^3.

Note that on the boundary components of $T^2 \times I$, the maps h_1 and h_2 are S^1–bundles over their images. Therefore, on the boundary components, we have distinguished pairs of simple closed curves: pairs of an S^1–fiber and a cross-section. Another important property of h_1 and h_2 is that for h_1 the canonical diffeomorphism between the components of $\partial(T^2 \times I)$ keeps the S^1–fiber and the cross-section, while for h_2 it interchanges them.

Now, let us proceed as in [14, Proof of Theorem 3.1]. Recall that in the proof there, a simple stable map into S^2 is first constructed: however, in our case, we can directly construct a map into $\mathbb{R}^2$ as the singular value set of $f|_{\bigsqcup_{j=1}^k X_j}$ is a concentric family of circles in $\mathbb{R}^2$. In order to extend the map, we need to arrange an appropriate map on each $N(T_i) \cong T^2 \times I$. Depending on the location of the images by the map $f|_{\bigsqcup_{j=1}^k X_j}$ of the small collar neighborhoods of the boundary tori for gluing, we have 4 cases.[1] Depending on the cases, we may need to decompose $N(T_i)$ into two or three parts, each of which is diffeomorphic to $T^2 \times I$, in order to glue the parts as prescribed by the weighted plumbing graph. The key ideas are to use h_2 in order to interchange the S^1–fiber and the cross-section for gluing, and to use h_1 in order to adjust the direction of the gluing. In order to use h_2, we need to use a disk region in the target: in such a case, we can choose the region that does not contain the point $\infty \in S^2 = \mathbb{R}^2 \cup \{\infty\}$.

[1] In [14, Proof of Theorem 3.1], we had only three cases, since S^2 was considered as the target. Here, as the target is $\mathbb{R}^2$, we have one more case for Case 2 there: however, the argument that we use is the same.

We can also arrange f in such a way that $f|_{S(f)}$ is an embedding by appropriately modifying f near $S(f)$ if necessary. This completes the proof of Theorem 3.1. □

Let us go on to the proof of the second theorem. In the following, we put, for $0 < a < b$,

$$A_{[a,b]} = \{(x,y) \in \mathbb{R}^2 \mid a \le \sqrt{x^2+y^2} \le b\}.$$

We can observe that $f^{-1}(C_{i-\frac{1}{2}})$ is a finite disjoint union of tori for each $i = 1, 2, \dots, m$, since M is orientable. Let K be the closure of a component of

$$M \setminus \Bigl(\bigsqcup_{i=1}^{m} f^{-1}(C_{i-1/2})\Bigr)$$

such that $f(K) \subset A_{[i-1/2,i+1/2]}$. Let $p_K : K \to S^1$ be the composition of $f|_K : K \to A_{[i-1/2,i+1/2]}$ and the radial projection $A_{[i-1/2,i+1/2]} \to S^1$. We can easily see that p_K and its restriction to the boundary are submersions and hence p_K is a locally trivial fibration. The fiber is a disjoint union of copies of D^2, A and P. Since $f|_{S(f)}$ is an embedding and K is connected, the fiber is diffeomorphic to D^2, P, or a finite disjoint union of copies of A. If the fiber is diffeomorphic to D^2, then K is diffeomorphic to $D^2 \times S^1$, since K is an orientable 3–dimensional manifold. If the fiber is diffeomorphic to P, then K is diffeomorphic either to $P \times S^1$ or a nontrivial P–bundle over S^1; see the proof of [14, Lemma 2.4].

Suppose that K is a nontrivial P–bundle over S^1 and that $C_i \subset f(S(f))$ is inward-directed. If $i = 1$, then this leads to a contradiction, since f is a trivial fiber bundle over the innermost region of $\mathbb{R}^2 \setminus f(S(f))$. If $i > 1$, then a component of $f^{-1}(A_{[i-3/2,i-1/2]})$ adjacent to K is either a nontrivial P–bundle over S^1, or a nontrivial $(A \sqcup A)$–bundle over S^1, where $A \sqcup A$ is the disjoint union of two copies of A and the monodromy for the latter bundle interchanges the two components of $A \sqcup A$. In the former case, $C_{i-1} \subset f(S(f))$ is outward-directed. In the latter case, we can repeat the argument toward inner components to find an outward-directed component.

Thus we have proved the following.

Lemma 3.3 *Let $f : M \to \mathbb{R}^2$ be a round fold map of a closed orientable 3–dimensional manifold such that $f(S(f)) = \bigcup_{i=1}^{m} C_i$. If f is directed, then the closure of a component of*

$$M \setminus \Bigl(\bigsqcup_{i=1}^{m} f^{-1}(C_{i-1/2})\Bigr)$$

is never diffeomorphic to the nontrivial P–bundle over S^1.

Proof of Theorem 3.2 First, suppose that there exists a directed round fold map $f: M \to \mathbb{R}^2$. We may assume that it satisfies (2-1). Then the disjoint union of tori $\bigsqcup_{i=1}^{m} f^{-1}(C_{i-1/2})$ decomposes M into a union of copies of $P \times S^1$, $A \times S^1$ and $D^2 \times S^1$ attached along their torus boundaries. Note that by Lemma 3.3, a nontrivial P–bundle over S^1 does not appear, since f is directed. Furthermore, we can ignore the components diffeomorphic to $A \times S^1 \cong T^2 \times [-1, 1]$ for obtaining a decomposition of M.

Since f is directed and M is connected, we see that the components diffeomorphic to $D^2 \times S^1$ are the outermost component $f^{-1}(A_{[m-1/2, m+1/2]})$ together with the components of the innermost part $f^{-1}(A_{[0,1/2]})$: no other components are diffeomorphic to $D^2 \times S^1$. Then, we can easily see that the corresponding graph describing this decomposition of M into copies of $D^2 \times S^1$ and $P \times S^1$ is a tree, as the number of components of regular fibers strictly increases toward the central region.

Conversely, suppose that the graph describing the decomposition of M into copies of $P \times S^1$ and $D^2 \times S^1$ is a tree. By inserting pieces diffeomorphic to $A \times S^1$ if necessary, we may assume that the decomposition is of plumbing type. Then, the graph Γ describing this new decomposition is also a tree. Note that Γ has at least one vertex of degree one. Let k denote the number of vertices of Γ. We label the vertices by $\{1, 2, \ldots, k\}$ in such a way that

(1) the labeling gives a one-to-one correspondence between the set of vertices and the set $\{1, 2, \ldots, k\}$,

(2) the degree of the vertex labeled k is equal to one,

(3) for each $j \in \{1, 2, \ldots, k\}$, the vertices of labels $\geq j$ together with the edges connecting them constitute a connected subgraph of Γ.

This is possible, since Γ is a tree with only vertices of degrees one, two or three.

Then, we follow the procedure as in the proof of Theorem 3.1 for constructing a round fold map on M, except for the components corresponding to vertices of degree one whose label is different from k. Note that in the process described in the proof of [14, Theorem 3.1], we do not need to use $h_1 : T^2 \times I \to \mathbb{R}^2$ in our situation. Furthermore, when we use h_2, we make sure that the corresponding image is contained in $A_{[0,k]}$. Finally, for the components corresponding to vertices of degree one with label $< k$, we just consider the projection $D^2 \times S^1 \to D^2$, where the target D^2 should be enlarged depending on the label. This matches with the construction for the adjacent components.

Now, it is not difficult to see that the resulting map $f\colon M \to \mathbb{R}^2$ is a directed round fold map. This completes the proof. □

4 Further results and open problems

4.1 Corollaries and examples

In this subsection, we give some corollaries of our main theorems. We also show that the class of 3–manifolds that admit directed round fold maps is strictly smaller than that of all graph 3–manifolds.

Corollary 4.1 *Suppose that M is a closed connected orientable graph 3–manifold. If $H_1(M;\mathbb{Q}) = 0$, then it admits a directed round fold map into $\mathbb{R}^2$.*

Proof Let G be the graph corresponding to a decomposition of M into $P \times S^1$ and $D^2 \times S^1$ as described in Lemma 2.9. Then, we can naturally construct a continuous map $\gamma\colon M \to G$ in such a way that for each piece, the complement of a small collar neighborhood of the boundary is mapped to the corresponding vertex. Then, we can show that γ induces a surjection $\gamma_*\colon \pi_1(M) \to \pi_1(G)$. Since $H_1(M;\mathbb{Q}) = 0$, we see that G is a tree. Then, the result follows from Theorem 3.2. □

Since every closed orientable Seifert 3–manifold over the 2–sphere admits a decomposition into a union of a finite number of copies of $P \times S^1$ and $D^2 \times S^1$ such that the corresponding graph is a tree, we have the following.

Corollary 4.2 *Every closed orientable Seifert 3–manifold over S^2 admits a directed round fold map into $\mathbb{R}^2$.*

By virtue of the realization result due to [1], as a corollary, we see that every linking form can be realized as that of a 3–manifold admitting a directed round fold map into $\mathbb{R}^2$. Thus, the linking form cannot detect the nonexistence of a directed round fold map.

On the other hand, as to the cohomology ring, we have the following.

Corollary 4.3 *If a closed orientable 3–manifold M admits a directed round fold map into $\mathbb{R}^2$, then for every pair $\xi, \eta \in H^1(M;\mathbb{Q})$, their cup product $\xi \smile \eta$ vanishes in $H^2(M;\mathbb{Q})$.*

The above corollary follows from [2, Theorem 5.2]. More precisely, let us consider the decomposition of M into the union of copies of $P \times S^1$ and $D^2 \times S^1$ attached along their torus boundaries such that the corresponding graph is a tree. As P and D^2 are of genus 0, and as the cohomology ring of $S^2 \times S^1$ satisfies the property described as in the corollary, we see that the cohomology ring of M also satisfies the same property.

Thus, for example, for every closed orientable surface Σ of genus ≥ 1, the 3–manifold $\Sigma \times S^1$ never admits a directed round fold map into $\mathbb{R}^2$, although it is a graph manifold.

Note that if we use coefficients other than $\mathbb{Q}$, the result might not hold. For example, $\mathbb{R}P^3$ admits a directed round fold map into $\mathbb{R}^2$, as it is the union of two copies of $D^2 \times S^1$ attached along their boundaries; however, for the generator of $H^1(\mathbb{R}P^3; \mathbb{Z}_2) \cong \mathbb{Z}_2$, its square does not vanish in $H^2(\mathbb{R}P^3; \mathbb{Z}_2)$. On the other hand, we do not know if the result in Corollary 4.3 holds for $\mathbb{Z}$–coefficients.

Now, let us consider the normal form plumbing graphs as explained in Section 2.2. The following lemma can be proved by following the proof of [12, Theorem 4.1].

Lemma 4.4 *If a closed connected orientable graph 3–manifold is decomposed into a union of finite numbers of copies of $P \times S^1$ and $D^2 \times S^1$ in such a way that the corresponding graph is a tree, then its normal form plumbing graph is a finite disjoint union of trees.*

Proof Let us consider the tree that represents a given decomposition into copies of $P \times S^1$ and $D^2 \times S^1$. This may not be of plumbing type; however, by inserting copies of $T^2 \times I$ if necessary, we may assume that the tree is of plumbing type.

On the other hand, as shown in the proof of [12, Theorem 4.1], there is an algorithm that turns a given plumbing graph into a normal form. It is not difficult to see that if we start with a tree, then each operation in the algorithm keeps the property that it is a finite disjoint union of trees. Then the result follows. □

As a corollary, we have the following.

Corollary 4.5 *Let M be a closed connected orientable graph 3–manifold whose normal-form plumbing graph contains a cycle. Then, M admits a round fold map into $\mathbb{R}^2$ but does not admit a directed round fold map into $\mathbb{R}^2$.*

For example, some torus bundles over S^1 as described in [12, Theorem 6.1] satisfy the assumption of the above corollary. (More precisely, those torus bundles over S^1 whose monodromy matrix has trace ≥ 3 or ≤ -3 give such examples.)

4.2 Open problems

Finally, we list some related open problems which may interest the reader.

Problem 4.6 (1) Generalize Theorems 3.1 and 3.2 for nonorientable 3–manifolds.

(2) The notion of round fold maps of 3–dimensional manifolds into $\mathbb{R}^2$ as in Definition 2.5 can naturally be generalized to that of round fold maps of n–dimensional manifolds into $\mathbb{R}^p$ for $n \geq p \geq 2$; for details, see [7; 6; 5]. For such a fixed pair (n, p) of dimensions, characterize those closed n–dimensional manifolds which admit round fold maps into $\mathbb{R}^p$.

For the dimension pair $(n, n-1)$, $n \geq 4$, such a generalization has been obtained in [8].

Problem 4.7 Classify the right–left equivalence classes of (directed) round fold maps on a given 3–manifold.

Refer to a certain classification result for simple stable maps given in [14]. For round fold maps of n–dimensional manifolds into $\mathbb{R}^{n-1}$, a classification result has been obtained in [8].

Recall that as explained in Section 2.1, a round fold map corresponds naturally to an open book structure. On the other hand, open book structures are closely related to contact structures on 3–manifolds; for example, see [3]. Therefore, the following problem seems to be reasonable.

Problem 4.8 Clarify the relationship between round fold maps and contact structures on 3–dimensional manifolds through open book decompositions.

One of the main motivations of Neumann's work [12] on plumbing graphs is to analyze the topology of the links of normal surface singularities. The following questions have been addressed by quite a few topologists to the authors.

Problem 4.9 Is there any relation between singularity links and round fold maps? Is it possible to construct explicit round fold maps on the singularity links in a natural way?

Acknowledgements We thank Professor Yuya Koda for stimulating discussions, which motivated the theme of this paper. We also thank the referee for various important comments which drastically improved the presentation of the paper. This work has been supported by JSPS KAKENHI grant JP17H06128.

References

[1] **J Bryden**, **F Deloup**, *A linking form conjecture for* 3*–manifolds*, from "Advances in topological quantum field theory" (J M Bryden, editor), NATO Sci. Ser. II Math. Phys. Chem. 179, Kluwer, Dordrecht (2004) 253–265 MR Zbl

[2] **M I Doig**, **P D Horn**, *On the intersection ring of graph manifolds*, Trans. Amer. Math. Soc. 369 (2017) 1185–1203 MR Zbl

[3] **E Giroux**, *Géométrie de contact: de la dimension trois vers les dimensions supérieures*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed., Beijing (2002) 405–414 MR Zbl

[4] **F Hirzebruch**, **W D Neumann**, **S S Koh**, *Differentiable manifolds and quadratic forms*, Lecture Notes in Pure and Applied Mathematics 4, Marcel Dekker, New York (1971) MR Zbl

[5] **N Kitazawa**, *Constructions of round fold maps on smooth bundles*, Tokyo J. Math. 37 (2014) 385–403 MR Zbl

[6] **N Kitazawa**, *Fold maps with singular value sets of concentric spheres*, Hokkaido Math. J. 43 (2014) 327–359 MR Zbl

[7] **N Kitazawa**, *On manifolds admitting fold maps with singular value sets of concentric spheres*, PhD thesis, Tokyo Institute of Technology (2014)

[8] **N Kitazawa**, **O Saeki**, *Round fold maps of n–dimensional manifolds into $(n-1)$–dimensional Euclidean space*, J. Singul. 26 (2023) 1–12 MR Zbl

[9] **H I Levine**, *Elimination of cusps*, Topology 3 (1965) 263–296 MR Zbl

[10] **H Levine**, *Classifying immersions into $\mathbb{R}^4$ over stable maps of* 3*–manifolds into $\mathbb{R}^2$*, Lecture Notes in Math. 1157, Springer (1985) MR Zbl

[11] **R Myers**, *Open book decompositions of* 3*–manifolds*, Proc. Amer. Math. Soc. 72 (1978) 397–402 MR Zbl

[12] **W D Neumann**, *A calculus for plumbing applied to the topology of complex surface singularities and degenerating complex curves*, Trans. Amer. Math. Soc. 268 (1981) 299–344 MR Zbl

[13] **O Saeki**, *Simple stable maps of* 3*–manifolds into surfaces*, *II*, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 40 (1993) 73–124 MR Zbl

[14] **O Saeki**, *Simple stable maps of* 3*–manifolds into surfaces*, Topology 35 (1996) 671–698 MR Zbl

[15] **O Saeki**, *Topology of singular fibers of differentiable maps*, Lecture Notes in Math. 1854, Springer (2004) MR Zbl

[16] **H Whitney**, *On singularities of mappings of euclidean spaces, I: Mappings of the plane into the plane*, Ann. of Math. 62 (1955) 374–410 MR Zbl

Institute of Mathematics for Industry, Kyushu University
Fukuoka, Japan

Institute of Mathematics for Industry, Kyushu University
Fukuoka, Japan

n-kitazawa@imi.kyushu-u.ac.jp, saeki@imi.kyushu-u.ac.jp

Received: 7 September 2021 Revised: 29 March 2022

Algebraic & Geometric Topology **23**:8 (2023) 3763–3804
DOI: 10.2140/agt.2023.23.3763
Published: 5 November 2023

The upsilon invariant at 1 of 3–braid knots

PAULA TRUÖL

We provide explicit formulas for the integer-valued smooth concordance invariant $\upsilon(K) = \Upsilon_K(1)$ for every 3–braid knot K. We determine this invariant, which was defined by Ozsváth, Stipsicz and Szabó (2017), by constructing cobordisms between 3–braid knots and (connected sums of) torus knots. As an application, we show that for positive 3–braid knots K several alternating distances all equal the sum $g(K) + \upsilon(K)$, where $g(K)$ denotes the 3–genus of K. In particular, we compute the alternation number, the dealternating number and the Turaev genus for all positive 3–braid knots. We also provide upper and lower bounds on the alternation number and dealternating number of every 3–braid knot which differ by 1.

57K10; 20F36, 57K18

1 Introduction

We study *knots* in the 3–sphere S^3, ie nonempty, connected, oriented, closed, smooth 1–dimensional submanifolds of S^3, considered up to ambient isotopy. Two knots K and J are called *concordant* if there exists an annulus $A \cong S^1 \times [0, 1]$ smoothly and properly embedded in $S^3 \times [0, 1]$ such that $\partial A = K \times \{0\} \cup J \times \{1\}$ and such that the induced orientation on the boundary of the annulus agrees with the orientation of K, but is the opposite one on J. Knots up to concordance form a group, the *concordance group* $\mathcal{C}$, with the group operation induced by connected sum.

In [46], Ozsváth, Stipsicz and Szabó used the Heegaard Floer knot complex to define the invariant Υ_K of a knot K, which induces a homomorphism from the knot concordance group to the group of real-valued piecewise-linear functions on the interval $[0, 2]$. The function Υ_K evaluated at $t = 1$, $\upsilon(K) := \Upsilon_K(1)$, induces a homomorphism $\mathcal{C} \to \mathbb{Z}$. In this article, we will call $\upsilon(K)$ *upsilon of* K.

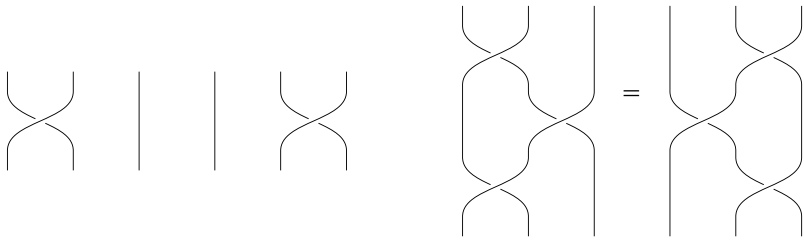


Figure 1: Generators and relation in the braid group B_3. Left: the two generators a and b. Right: the braid relation $aba = bab$.

A 3–*braid* is an element of the *braid group on three strands*, denoted B_3. The classical presentation of B_3 with generators a and b and relation $aba = bab$, the *braid relation*, was introduced by Artin [5]. A *braid word* γ — a word in the generators of B_3 and their inverses — defines a diagram for a (geometric) 3–braid; the generators a and b correspond to the geometric 3–braids given by braid diagrams as in Figure 1. In our figures, braid diagrams will always be oriented from bottom to top. We denote by Δ the braid $aba = bab$, and note that its square $\Delta^2 = (ab)^3$ (the positive full twist on three strands) generates the center of B_3; see Chow [14, Theorem 3]. A 3–*braid knot* is a knot that arises as the closure $\hat{\gamma}$ of a 3–braid γ.

As our main result, we determine the upsilon invariant for all 3–braid knots. More precisely, we show the following.

Theorem 1.1 *Let $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} a^{-p_2} b^{q_2} \cdots a^{-p_r} b^{q_r}$ be a braid word in the generators a and b of B_3 for some integers $\ell \in \mathbb{Z}$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$, where $\Delta^2 = (ab)^3$. Suppose that the closure $K = \hat{\gamma}$ of γ is a knot. Then its upsilon invariant is*

$$\upsilon(K) = \frac{1}{2}\left(\sum_{i=1}^{r}(p_i - q_i)\right) - 2\ell.$$

By Murasugi's classification of the conjugacy classes of 3–braids [45, Proposition 2.1], indeed all 3–braid knots — except for the torus knots that are closures of 3–braids — are covered by Theorem 1.1. However, for torus knots the invariant υ can be calculated explicitly by a combinatorial, inductive formula in terms of their Alexander polynomial [46, Theorem 1.15]; see (12) below. Hence, we have indeed determined $\upsilon(K)$ for all 3–braid knots K.

As an application of Theorem 1.1, we show that the following invariants coincide for *positive* 3–*braid knots* — knots that are the closure of positive 3–braids.

Corollary 1.2 *Let K be a knot that is the closure of a* **positive 3–braid**, *ie an element of B_3 that can be written as a word in the generators a and b only (no inverses). Then*

$$\operatorname{alt}(K) = \operatorname{dalt}(K) = g_T(K) = \mathcal{A}_s(K) = g(K) + \upsilon(K).$$

Here, the *alternation number* $\operatorname{alt}(K)$, *dealternating number* $\operatorname{dalt}(K)$ and *Turaev genus* $g_T(K)$ are different ways of measuring how far the knot K is from being alternating. The best known among them is certainly the first one: the alternation number $\operatorname{alt}(K)$ of a knot K was first defined by Kawauchi [31] as the minimal *Gordian distance* of K to the set of alternating knots. In Section 5, we will review the precise definition and prove Corollary 1.2. The invariant $\mathcal{A}_s(K)$ introduced by Friedl, Livingston and Zentner [23] is defined as the minimal number of double point singularities in a generically immersed concordance from a knot K to an alternating knot. Lastly, $g(K)$ denotes the 3*–genus* of K, the minimal genus of a compact, connected, oriented, smooth surface in S^3 with oriented boundary the knot K.

Two other corollaries of Theorem 1.1 for positive 3–braid knots are the following.

Corollary 1.3 *Let K be a positive 3–braid knot. Then the minimal r such that K is the closure of $a^{p_1}b^{q_1}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}$ for positive integers p_i and q_i, for $i \in \{1,\dots,r\}$, is $r = g(K) + \upsilon(K) + 1$.*

Corollary 1.4 *If K and J are concordant knots that are both closures of positive 3–braids, then the minimal r from Corollary 1.3 is the same for both K and J.*

Proposition 3.2 provides a normal form for 3–braids, the *Garside normal form*, which is different from Murasugi's normal form mentioned above (see Definition 4.15). The Garside normal form allows us to read off from a braid word whether it is conjugate to a positive braid word. In Section 6, we provide formulas for the fractional Dehn twist coefficient for all 3–braids in Garside normal form; see Corollary 6.1.

Proof strategy for Theorem 1.1 A crucial property of the invariant υ is that it provides a lower bound on the 4*–genus* $g_4(K)$ of a knot K, the minimal genus of a compact, connected, oriented surface smoothly embedded in the 4–ball B^4 with oriented boundary the knot K in $S^3 = \partial B^4$: we have

$$|\upsilon(K)| \leq g_4(K) \tag{1}$$

for any knot K [46, Theorem 1.11]. Our general strategy to find $\upsilon(K)$ for any 3–braid knot K will be to construct a cobordism between K and another knot J for which the value of υ is known. A *cobordism* between K and J is a smoothly and properly embedded oriented surface C in $S^3 \times [0,1]$ with boundary $K \times \{0\} \cup J \times \{1\}$ such that the induced orientation on the boundary of C agrees with the orientation of K and disagrees with the orientation of J. We have

$$|\upsilon(K) - \upsilon(J)| \leq g(C) \tag{2}$$

for any cobordism C between K and J, where $g(C)$ denotes the genus of the cobordism; see inequality (15) in Section 4.1. This provides bounds on $\upsilon(K)$ in terms of $\upsilon(J)$ and $g(C)$.

We will find such cobordisms for example by algebraic modifications of a braid word representing K and by saddle moves corresponding to the addition or deletion of generators from such braid words. We will also repeatedly make use of the trick described in Example 4.1 in Section 4.1 of looking at cobordisms of genus 1 between $\hat{\gamma} \# T_{2,2n+1}$ and $\widehat{\gamma b^{2n}}$ for 3–braid words γ and $n \geq 1$.

To prove Theorem 1.1, we will first determine υ for all positive 3–braid knots and then generalize our computations to all 3–braid knots. This extension was somewhat unexpected for the author since, in contrast, the same method would not work to determine slice-torus invariants — see Lewark [33] — like the invariant τ defined by Ozsváth and Szabó [48] or Rasmussen's invariant s [50] for all 3–braid knots. We will elaborate on this in Section 4.4.2.

Remark 1.5 As we will only use properties of the upsilon invariant (see Section 2.2) and not its definition, we can similarly determine any concordance homomorphism $\mathcal{C} \to \mathbb{Z}$ whose absolute value bounds the 4–genus of a knot from below and which takes the same value as υ on torus knots of braid index 2 and 3. An example is $-\frac{1}{2}t$ for the concordance invariant t constructed by Ballinger [8] from the $E(-1)$ spectral sequence on Khovanov homology. The invariant t defines a concordance homomorphism valued in the even integers which satisfies $\left|\frac{1}{2}t(K)\right| \leq g_4(K)$ for any knot K [8, Theorem 1.1]. Moreover, it fulfills $t(T_{p,q}) = -2\upsilon(T_{p,q})$ for the torus knots $T_{p,q}$ for any coprime positive integers p and q [8, page 22]. The same method we use for the proof of Theorem 1.1 shows that $t(K) = -2\upsilon(K)$ for any 3–braid knot K.

Remark 1.6 Theorem 1.1 and a result of Erle [17] imply that $\sigma(K) = 2\upsilon(K)$ for all 3–braid knots K except when $K = \pm T_{3,3\ell+k}$ for odd $\ell > 0$ and $k \in \{1,2\}$. Here $\sigma(K)$

denotes the classical signature of the knot K; see Trotter [54].[1] In the exceptional cases, $\sigma(K) = 2\upsilon(K) - 2$. This observation improves a result by Feller and Krcatovich who showed that $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right| \leq 2$ for all 3–braid knots K [20, Proposition 4.4]; see also Section 4.4.1.

Organization The remainder of this article is organized as follows. In Section 2, we will provide the necessary background on (positive) braids and the upsilon invariant before providing a normal form for 3–braids (Proposition 3.2) that we call the Garside normal form in Section 3. Then in Section 4, after a more detailed outline of our proof strategy (Section 4.1), we will prove Theorem 1.1 first for positive 3–braid knots (Section 4.2) and afterwards in the general 3–braid case (Section 4.3). We will prove Corollaries 1.3 and 1.4 in Section 4.2. Section 4.4 will provide further context on our results. Section 5 is concerned with the proof of Corollary 1.2 (Section 5.1) and the application of our result about the upsilon invariant to alternating distances of general 3–braid knots (Section 5.2). In particular, we determine the alternation number of any 3–braid knot up to an additive error of at most 1. Finally, in Section 6, we determine the fractional Dehn twist coefficient for all 3–braids in Garside normal form.

Acknowledgements I would like to thank Peter Feller for introducing me to the topic and for all the helpful discussions. Thanks also to Lukas Lewark for lots of useful comments, including during my stay in Regensburg in September 2020, and to Xenia Flamm for her feedback. Finally, I thank the referee for many valuable remarks and improvements. This project is supported by the Swiss National Science Foundation grant 181199.

2 Preliminaries

We recall important concepts about knots and braids, and also the necessary properties of the upsilon invariant and the knot invariant τ coming from Heegaard Floer homology.

2.1 Knots and braids

By a fundamental theorem of Alexander [4], every knot in S^3 can be represented as the closure of a geometric n–braid for some positive integer n. An *n–braid* is an element

[1] We use the standard signature convention that the positive torus knots have negative signatures, eg $\sigma(T_{3,2}) = -2$.

of the *braid group on n strands*, denoted by B_n, which is presented by $n-1$ generators $\sigma_1, \dots, \sigma_{n-1}$ and relations

$$\sigma_i \sigma_j = \sigma_j \sigma_i \quad \text{if } |i-j| \geq 2,$$
$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1};$$

see [5]. We call a word in the generators of B_n and their inverses a *braid word*. A braid word defines a diagram for a (geometric) n–braid where the generators σ_i of the braid group correspond to the geometric n–braids given by the braid diagrams in which the i^{th} and $(i+1)^{\text{st}}$ strands cross once positively. In the following, we will always identify braid words with the corresponding geometric braids, and we suppress n if the context is clear.

By gluing the top ends of the (oriented) strands of a geometric braid $\gamma \in B_n$ to the corresponding bottom ends, we get a knot (or link) $\hat{\gamma}$, called the *closure* of γ. If γ induces a permutation with only one cycle on the ends of its n strands, then its closure $\hat{\gamma}$ is a knot and we call it an *n–braid knot*. Note that conjugate braids $\gamma_0, \gamma_1 \in B_n$, denoted by $\gamma_0 \sim \gamma_1$, have isotopic closures $\hat{\gamma}_0 = \hat{\gamma}_1$. For a more detailed account on braids, we refer the reader to [10].

A *positive braid* is an element of the braid group B_n for some n that can be written as a positive braid word $\sigma_{s_1} \sigma_{s_2} \cdots \sigma_{s_l}$ with $s_i \in \{1, \dots, n-1\}$. A knot is called a *positive braid knot* if it can be represented as the closure of a positive braid. The set of positive braid knots contains the sets of (positive) torus knots and algebraic knots, while itself being a subset of the set of positive knots or, more generally, the frequently studied set of (strongly) quasipositive knots.

Let $\operatorname{wr}(\gamma)$ denote the *writhe* of a braid word $\gamma \in B_n$, ie the exponent sum of the word γ. If γ is a positive n–braid such that $K = \hat{\gamma}$ is a knot, then, by work of Bennequin [9] and Rudolph [51] — the latter building on Kronheimer and Mrowka's proof of the local Thom conjecture [32] — we have

$$g_4(K) = g(K) = \tfrac{1}{2}(\operatorname{wr}(\gamma) - n + 1). \tag{3}$$

2.2 The concordance invariants τ and Υ

In [48], Ozsváth and Szabó constructed the knot invariant τ via the knot filtration on the Heegaard Floer chain complex of S^3; the latter was also defined independently by Rasmussen [49]. The invariant τ induces a group homomorphism $\mathcal{C} \to \mathbb{Z}$ from the (smooth) knot concordance group $\mathcal{C}$ to the group of integers $\mathbb{Z}$ and gives a lower bound on the 4–ball genus $g_4(K)$: we have $|\tau(K)| \leq g_4(K)$ for any knot K. For the torus knots $T_{p,q}$, where p and q are coprime positive integers, the invariant τ recovers the 3–genus [48, Corollary 1.7]; namely,

$$\tau(T_{p,q}) = g(T_{p,q}) = \tfrac{1}{2}(p-1)(q-1). \tag{4}$$

Moreover, it follows from [34, Theorem 4 and Corollary 7] together with (3) above that, for any knot K that is the closure of a positive n–braid γ,

$$\tau(K) = \tfrac{1}{2}(\mathrm{wr}(\gamma) - n + 1) = g_4(K) = g(K). \tag{5}$$

The invariant Υ was defined by Ozsváth, Stipsicz and Szabó in [46]. We will not recall the definition of Υ via the knot Floer complex $CFK^\infty(K)$ since the properties of Υ mentioned below will be enough for our later computations and we will not explicitly use the Heegaard Floer theory behind it. For an overview on the properties of Υ, see the original article [46] or Livingston's notes on Υ [35]; see [28] for a survey on Heegaard Floer homology and knot concordance.

For every knot K, the knot invariant $\Upsilon_K : [0, 1] \to \mathbb{R}$ is a continuous, piecewise linear function with the following properties [46]:

$$\Upsilon_K(0) = 0, \tag{6}$$

$$\text{the slope of } \Upsilon_K(t) \text{ at } t = 0 \text{ is given by } -\tau(K), \tag{7}$$

$$\Upsilon_{K_1 \# K_2}(t) = \Upsilon_{K_1}(t) + \Upsilon_{K_2}(t) \text{ for all } 0 \le t \le 1 \text{ and all knots } K_1 \text{ and } K_2, \tag{8}$$

$$\Upsilon_{-K}(t) = -\Upsilon_K(t) \text{ for all } 0 \le t \le 1, \tag{9}$$

$$|\Upsilon_K(t)| \le g_4(K)t \text{ for all } 0 \le t \le 1. \tag{10}$$

Here, $-K$ is the knot obtained by mirroring K and reversing its orientation. Its concordance class is the inverse of the class of K in the knot concordance group $\mathcal{C}$. It follows from (8)–(10) that Υ induces a homomorphism from the concordance group to the group of real-valued piecewise-linear functions on the interval [0, 1].

For some classes of knots, the invariant Υ can be explicitly computed in terms of classical knot invariants like the signature and the Alexander polynomial.

Proposition 2.1 [46, Theorem 1.14] *We have $\Upsilon_K(t) = \frac{1}{2}\sigma(K)t$ for all alternating or quasialternating knots K and all $0 \le t \le 1$.*

For positive torus knots, $\Upsilon_K(t)$ is completely determined by a combinatorial formula in terms of their Alexander polynomial [46, Theorem 1.15]. For torus knots of braid index 2 or 3, the following holds; see eg [18]. For $\ell \ge 0$,

$$\Upsilon_{T_{2,2\ell+1}}(t) = -\tau(T_{2,2\ell+1}) \cdot t = -\ell \cdot t \quad \text{for all } 0 \le t \le 1. \tag{11}$$

For $\ell \geq 0$ and $k \in \{1, 2\}$,

$$(12)\qquad \begin{aligned} &\Upsilon_{T_{3,3\ell+1}}(1) = \Upsilon_{T_{3,3\ell+2}}(1) + 1 = -2\ell, \\ &\Upsilon_{T_{3,3\ell+k}}(t) = -\tau(T_{3,3\ell+k})t = -(3\ell + k - 1)t \quad \text{for all } 0 \leq t \leq \tfrac{2}{3} \\ &\Upsilon_{T_{3,3\ell+k}}(t) \text{ is linear on } \left[\tfrac{2}{3}, 1\right]. \end{aligned}$$

3 The Garside normal form for 3–braids

In this section, we provide a classification result on the conjugacy classes of 3–braids; see Proposition 3.2. This result is basically due to work of Garside [25] who gave the first solution to the conjugacy problem for all braid groups B_n with $n \geq 3$ in 1965. Proposition 3.2 might be known to the experts, but since the explicit formulas appear to be missing from the literature, we will provide them here.

Throughout, we denote the two generators of the braid group B_3 by $a := \sigma_1$ and $b := \sigma_2$ which are subject to the braid relation $aba = bab$. Recall that the braid $\Delta^2 = (aba)^2 = (ab)^3$ generates the center of B_3.

Remark 3.1 Any 3–braid is conjugate to the same braid with generators a and b interchanged. More precisely, let $\gamma = a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $r \geq 1$ and integers p_i and q_i for $i \in \{1, \dots, r\}$ be a 3–braid. Then using $\Delta a = b\Delta$ and $\Delta b = a\Delta$, we have

$$\gamma = \Delta^{-1} \Delta a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r} = \Delta^{-1} b^{p_1} a^{q_1} \cdots b^{p_r} a^{q_r} \Delta \sim b^{p_1} a^{q_1} \cdots b^{p_r} a^{q_r}.$$

In Proposition 3.2, we will provide a certain standard form for the conjugacy classes of 3–braids.

Proposition 3.2 *Let γ be a 3–braid. Then γ is conjugate to one of the 3–braids*

(A) $\Delta^{2\ell} a^p$, $\quad \ell \in \mathbb{Z},\ p \geq 0$,

(B) $\Delta^{2\ell} a^p b$, $\quad \ell \in \mathbb{Z},\ p \in \{1, 2, 3\}$,

(C) $\Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$, $\quad \ell \in \mathbb{Z},\ r \geq 1,\ p_i, q_i \geq 2,\ i \in \{1, \dots, r\}$,

(D) $\Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}$, $\quad \ell \in \mathbb{Z},\ r \geq 1,\ p_r, p_i, q_i \geq 2,\ i \in \{1, \dots, r-1\}$.

If γ is a positive 3–braid, then $\ell \geq 0$. If $\hat{\gamma}$ is a knot, then only cases (B)–(D) can occur and p must be odd in case (B), at least one of the p_i and one of the q_i must be odd in case (C), and at least one of the p_i or q_i must be odd in case (D).

While we will never use it in this article, we note — without proof — the following uniqueness result related to Proposition 3.2.

Remark 3.3 Up to cyclic permutation of the powers $p_1, q_1, \dots, p_r, q_r$ in (C) and $p_1, q_1, \dots, p_{r-1}, q_{r-1}, p_r$ in (D), each 3–braid is conjugate to exactly one of the 3–braids listed in Proposition 3.2. This follows from Garside's work [25]. In his notation, each of the 3–braids listed in (A)–(D) in Proposition 3.2 is the standard form of a certain element in the (so-called) summit set of γ. For 3–braids of the form (C) or (D), the summit set consists of those 3–braids obtained by cyclic permutation of the powers $p_1, q_1, \dots, p_r, q_r$ in (C) and $p_1, q_1, \dots, p_{r-1}, q_{r-1}, p_r$ in (D), respectively.

Definition 3.4 We call a braid word of the form in (A)–(D) a 3*–braid in Garside normal form.*

Remark 3.5 The advantage of the Garside normal form over Murasugi's normal form for 3–braids used later in Section 4.3 (see Definition 4.15) is that positive 3–braids are easier to detect in this normal form: if γ is a positive 3–braid, then γ is conjugate to one of the braids in (A)–(D) with $\ell \geq 0$. Since Garside's solution to the conjugacy problem works for any n–braid with $n \geq 3$, one might hope to generalize an explicit standard form as in Proposition 3.2 to n–braids for any $n \geq 3$.

Remark 3.6 For odd p, case (B) of Proposition 3.2 covers the torus knots of braid index 3. More precisely, if $\gamma \sim \Delta^{2\ell}ab = (ab)^{3\ell+1}$, then its closure is $\hat{\gamma} = T_{3,3\ell+1}$ for $\ell \geq 0$ and $\hat{\gamma} = -T_{3,3(-\ell-1)+2}$ for $\ell < 0$, and if $\gamma \sim \Delta^{2\ell}a^3b \sim (ab)^{3\ell+2}$, then $\hat{\gamma} = T_{3,3\ell+2}$ for $\ell \geq 0$ and $\hat{\gamma} = -T_{3,3(-\ell-1)+1}$ for $\ell < 0$.

Proof of Proposition 3.2 The proof will follow from the following claim.

Claim 1 *Let γ be a positive* 3*–braid. Then γ is conjugate to one of the* 3*–braids in* (A)–(D) *with $\ell \geq 0$.*

We first deduce Proposition 3.2 from this claim. To that end, let γ be any 3–braid. If γ is a positive braid, we are done by Claim 1. If not, then γ can be written in the form $\gamma = \Delta^m\alpha$ where m is a negative integer and α a positive 3–braid [25, Theorem 5]. In fact, inserting $\Delta^{-1}\Delta$ if m is odd, we can assume γ to be of the form $\Delta^{-2n}\alpha$ for some $n \geq 1$ and a positive 3–braid α. The proposition then easily follows using the claim for α. It remains to prove Claim 1.

Proof of Claim 1 A positive 3–braid γ has the form $\gamma = a^{P_1}b^{Q_1} \cdots a^{P_R}b^{Q_R}$ for integers $R \geq 1$ and $P_i, Q_i \geq 0$ for $i \in \{1, \dots, R\}$. If all the P_i or all the Q_i are 0, then (possibly using Remark 3.1) γ is conjugate to a^p for some $p \geq 0$ and we are in case (A) for $\ell = 0$. Possibly after conjugation and reduction of R, we can thus assume that all of the integers P_i and Q_i are nonzero. If $P_1, Q_1 \geq 2$ applies for all $i \in \{1, \dots, R\}$, then γ is of the form in (C) for $\ell = 0$. If $R = 1$, ie $\gamma = a^{P_1} b^{Q_1}$ for integers $P_1, Q_1 \geq 1$, and $P_1 = 1$ or $Q_1 = 1$, then (possibly using Remark 3.1) γ is conjugate to a braid of the form in (B).

It remains to consider the case where $R \geq 2$ and at least one of the P_i or Q_i is 1. In that case — if necessary after conjugation — γ contains $\Delta = aba = bab$ as a subword and is thus conjugate to $\Delta\alpha$ for some positive 3–braid α. Now, let $n \geq 1$ be maximal with the property that γ is conjugate to $\Delta^n \alpha$ for some positive 3–braid α. Then, possibly after conjugation of γ, the braid word α must be one of the following:

$$\begin{aligned}
&a^p, && p \geq 0,\\
&a^p b, && p \geq 1,\\
&a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}, && r \geq 1,\ p_i, q_i \geq 2,\ i \in \{1, \dots, r\},\\
&a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}, && r \geq 1,\ p_r \geq 2,\ p_i, q_i \geq 2,\ i \in \{1, \dots, r-1\}.
\end{aligned} \tag{13}$$

Indeed, using Remark 3.1, up to conjugation these are the only possible words such that $\Delta^n \alpha$ does not contain any additional Δ as a subword. Note that α can be the empty word, which is covered by the first case in (13) for $p = 0$. Further, note that

$$\begin{aligned}
\Delta^{2\ell} a^p b &\sim \Delta^{2\ell+1} a^{p-2},\\
\Delta^{2\ell+1} a &\sim \Delta^{2\ell} a^3 b,\\
\Delta^{2\ell+1} a^p b &\sim \Delta^{2\ell+1} a^{p+1},\\
\Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r} &\sim \Delta^{2\ell+1} a^{p_1+q_r} b^{q_1} a^{p_2} \cdots b^{q_{r-1}} a^{p_r},\\
\Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r} &\sim \Delta^{2\ell} a^{p_1+p_r} b^{q_1} a^{p_2} \cdots a^{p_{r-1}} b^{p_{r-1}}
\end{aligned} \tag{14}$$

for any $\ell \geq 0$, $p \geq 1$ and $p_i, q_i \geq 2$ for $i \in \{1, \dots, r\}$. It follows from a case by case analysis of the cases in (13), using (14) and taking the parity of n into account, that any positive 3–braid is conjugate to one of the 3–braids in (A)–(D) with $\ell \geq 0$. □

4 The upsilon invariant of 3–braid knots

In this section, we prove Theorem 1.1. Along the way, we compute the invariant υ for positive 3–braid knots in Garside normal form (Proposition 4.2) and prove Corollaries 1.3 and 1.4.

4.1 Methodology

We first recall inequality (2) from the introduction — which will be repeatedly used in Section 4 — in more generality.

The *cobordism distance* $d(K,J)$ between two knots K and J is defined as the 4–genus $g_4(K\#-J)$ of the connected sum of K and the inverse of J. Equivalently, the cobordism distance $d(K,J)$ could be defined as the minimal genus of a smoothly and properly embedded oriented surface C in $S^3\times[0,1]$ with boundary $K\times\{0\}\cup J\times\{1\}$ such that the induced orientation on the boundary of C agrees with the orientation of K and disagrees with the orientation of J. Suppose the genus of a cobordism C between two knots K and J is $g(C)$. We then have $d(K,J)\le g(C)$, so by the properties (8)–(10) of Υ from Section 2.2 we get

$$|\Upsilon_K(t)-\Upsilon_J(t)|=|\Upsilon_{K\#-J}(t)|\le g_4(K\#-T)t=d(K,T)t\le g(C)t \tag{15}$$

for all $0\le t\le 1$. This provides bounds on $\Upsilon_K(t)$ in terms of $\Upsilon_J(t)$ and $g(C)$.

We now give an example for the cobordisms we will use later on.

Example 4.1 Among other things, we will frequently use the following trick the author first saw in [20, Example 4.5]. Let γ be a 3–braid such that $K=\hat{\gamma}$ is a knot. Consider the 3–braid $\alpha:=\gamma b^{2n}$ for some $n\ge 1$. Then $\hat{\alpha}$ is also a knot and there is a cobordism between $\hat{\alpha}$ and the connected sum $K\#T_{2,2n+1}$ of genus 1. This cobordism can be realized by two saddle moves (1–handle attachments) of the form shown in Figure 2, right, performed in the two circled regions of Figure 2, left. One of them is used to add a generator b to the braid α to obtain the braid word γb^{2n+1} and the other is used to transform the closure of this new braid word into a connected sum of K and $T_{2,2n+1}$. Recall that our braid diagrams are oriented from bottom to top.

Using $\upsilon(T_{2,2n+1})=-n$ by (11) and that the genus of the cobordism is 1, by (15) for $t=1$ we have

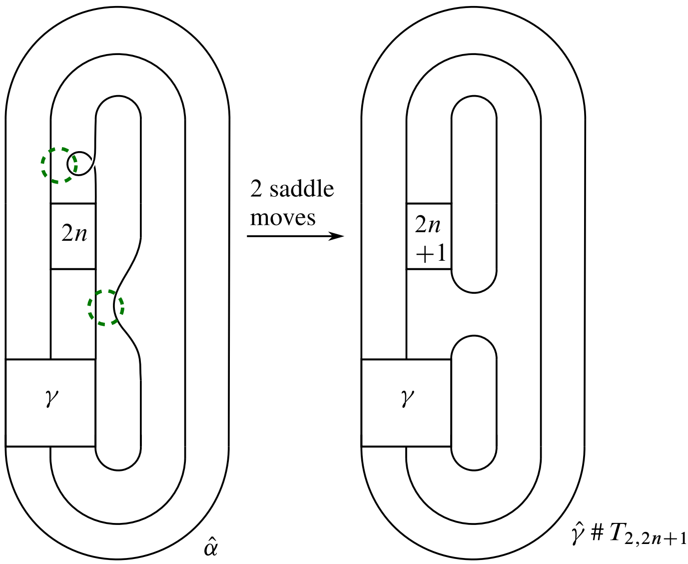
$$|\upsilon(\hat{\alpha})-\upsilon(K\#T_{2,2n+1})|\le 1 \iff |\upsilon(\hat{\alpha})-\upsilon(K)+n|\le 1, \tag{16}$$

which provides the lower bound $\upsilon(K)\ge\upsilon\big(\hat{\alpha}\big)+n-1$ on $\upsilon(K)$.

4.2 The upsilon invariant of positive 3–braid knots

In this section, we determine the invariant υ for all positive 3–braid knots.

By Proposition 3.2 and Remark 3.6, positive 3–braid knots are either the torus knots $T_{3,3\ell+k}$ for $\ell\ge 0$ and $k\in\{1,2\}$ which have braid representatives of Garside normal



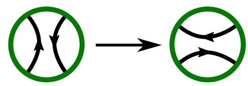


Figure 2: An example illustrating our proof strategy. Left: a schematic of a cobordism between the knots $\hat{\alpha}$ and $\hat{\gamma} \# T_{2,2n+1}$ realized by two saddle moves. Right: a saddle move.

form (B), or closures of positive 3–braids of Garside normal form (C) or (D) (see Definition 3.4). The following proposition thus proves Theorem 1.1 for all positive 3–braid knots.

Proposition 4.2 *Let γ be a positive 3–braid such that $K=\hat{\gamma}$ is a knot. Then*

$$\upsilon(K)=\begin{cases}-2\ell-\frac{1}{2}(p-1) & \text{if } \gamma \text{ is conjugate to a braid in (B)},\\ -\frac{1}{2}\bigl(\sum_{i=1}^{r}(p_i+q_i)\bigr)+r-2\ell & \text{if } \gamma \text{ is conjugate to a braid in (C)},\\ -\frac{1}{2}\bigl(\sum_{i=1}^{r-1}(p_i+q_i)+p_r\bigr)+r-2\ell-\frac{3}{2} & \text{if } \gamma \text{ is conjugate to a braid in (D)}.\end{cases}$$

Remark 4.3 In fact, the formulas from Proposition 4.2 also give the correct upsilon invariant in terms of the Garside normal form of a 3–braid representative of a knot K if K is the closure of any 3–braid in Garside normal form (C) or (D), not necessarily a positive one. This follows from Theorem 1.1 (proved in the next section) and the observations of Section 4.4.3.

Recall that for the torus knots of braid index 3, we know the invariant υ by (12). In the following, we will determine the invariant υ for all knots that are closures of positive 3–braids of Garside normal form (C) or (D).

We first provide an upper bound on $\Upsilon_K(t)$ for positive 3–braid knots K and $0 \le t \le 1$. The following inequality (17) in Lemma 4.4 could also be shown using the dealternating number and a result of Abe and Kishimoto [2, Lemma 2.2], whereas the main work for the upper bound on υ for the knots in the second and third case in Proposition 4.2 will be to rewrite the braid words representing these knots. We use the approach below since it will also give bounds on the minimal cobordism distance between any positive 3–braid knot and an alternating knot; see Remark 4.14.

Lemma 4.4 *Let $\gamma = a^{p_1}b^{q_1}\cdots a^{p_r}b^{q_r}$ be a positive* 3*–braid, where $r \ge 1$ and $p_i, q_i \ge 1$ for $i \in \{1,\dots,r\}$ are integers such that $K = \hat{\gamma}$ is a knot. Then*

$$\Upsilon_K(t) \le (-g(K) + r - 1)t \quad \textit{for all } 0 \le t \le 1. \tag{17}$$

Proof We claim that there is a cobordism C of genus

$$g(C) = \tfrac{1}{2}(r - 1 + \varepsilon) \tag{18}$$

between K and the connected sum

$$J_\varepsilon = T_{2,\sum_{i=1}^r p_i + \varepsilon_p} \# T_{2,q_1+\varepsilon_1} \# T_{2,q_2+\varepsilon_2} \# \cdots \# T_{2,q_r+\varepsilon_r},$$

where $\varepsilon_1,\dots,\varepsilon_r,\varepsilon_p \in \{0,1\}$ are chosen such that J_ε is a connected sum of torus knots (rather than links), ie such that $\sum_{i=1}^r p_i + \varepsilon_p, q_1+\varepsilon_1, q_2+\varepsilon_2,\dots,q_r+\varepsilon_r$ are all odd, and $\varepsilon := \varepsilon_p + \sum_{i=1}^r \varepsilon_i$. This cobordism C can be realized by $r-1+\varepsilon$ saddle moves as follows. Following the schematic in Figure 3, we add ε generators b by ε saddle moves and additionally perform $r-1$ saddle moves of the form shown in Figure 2, right, in the circled regions of Figure 3. In Figure 3, a box on the left labeled p_i or q_i stands for the positive braid a^{p_i} or b^{q_i}, respectively. The Euler characteristic of the cobordism C is $\chi(C) = -r + 1 - \varepsilon$. Since C is connected and — as J_ε and K are knots — has two boundary components, the genus of C is $g(C) = -\frac{1}{2}\chi(C) = \frac{1}{2}(r-1+\varepsilon)$ as claimed. By (15), we get $|\Upsilon_K(t) - \Upsilon_{J_\varepsilon}(t)| \le g(C)t$ for all $0 \le t \le 1$; hence

$$\Upsilon_K(t) \le \Upsilon_{J_\varepsilon}(t) + g(C)t \quad \text{for all } 0 \le t \le 1. \tag{19}$$

By (8) and (11) from Section 2.2,

$$\begin{aligned}\Upsilon_{J_\varepsilon}(t) &= \left(-\frac{1}{2}\left(\sum_{i=1}^r p_i + \varepsilon_p - 1\right) - \frac{1}{2}(q_1+\varepsilon_1-1) - \frac{1}{2}(q_2+\varepsilon_2-1) - \cdots - \frac{1}{2}(q_r+\varepsilon_r-1)\right)t \\ &= -\frac{1}{2}\left(\sum_{i=1}^r (p_i+q_i) - (r+1) + \varepsilon\right)t,\end{aligned}$$

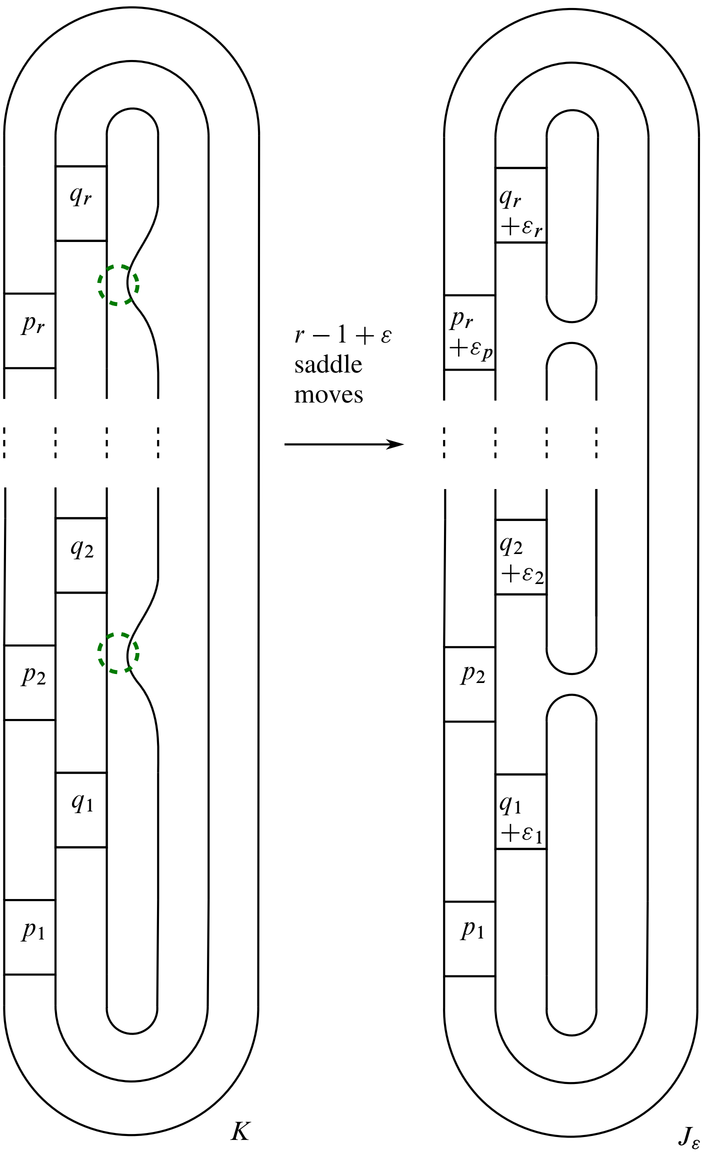


Figure 3: A schematic of a cobordism between $K=\hat{\gamma}$ and the connected sum of torus knots $J_\varepsilon = T_{2,\sum_{i=1}^r p_i+\varepsilon_p} \# T_{2,q_1+\varepsilon_1} \# T_{2,q_2+\varepsilon_2} \# \cdots \# T_{2,q_r+\varepsilon_r}$ realized by $r-1+\varepsilon$ saddle moves.

so (18) and (19) imply

$$\Upsilon_K(t) \leq \left(-\frac{1}{2}\left(\sum_{i=1}^{r}(p_i+q_i)\right)+r\right)t \quad \text{for all } 0 \leq t \leq 1.$$

The claim follows, since by (3),

$$g(K) = \frac{\mathrm{wr}(\gamma) - 2}{2} = \frac{1}{2}\left(\sum_{i=1}^{r}(p_i + q_i) - 2\right). \qquad \square$$

The following two lemmas improve the bound from Lemma 4.4 for knots that are closures of positive 3–braids of Garside normal form (C) or (D), respectively.

Lemma 4.5 *Let $\gamma = \Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}$ for some $\ell \geq 0$, $r \geq 1$, $p_r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \ldots, r-1\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\Upsilon_K(t) \leq \left(-\frac{1}{2}\left(\sum_{i=1}^{r-1}(p_i + q_i) + p_r\right) + r - 2\ell - \frac{3}{2}\right)t \quad \textit{for all } 0 \leq t \leq 1.$$

In the proof of Lemma 4.5, we will use that in B_3,

$$(ab)^{3n+1} = ab\Delta^{2n} = a^2ba^3(aba^3)^{n-1}ba^n \quad \text{for all } n \geq 1, \tag{20}$$

where $\Delta^2 = (aba)^2 = (ab)^3 = (ba)^3$; see [18, Proof of Proposition 22].

Proof of Lemma 4.5 Let $\Sigma_\gamma = \sum_{i=1}^{r-1}(p_i + q_i) + p_r$ and note that using (3),

$$g(K) = \tfrac{1}{2}(3(2\ell + 1) + \Sigma_\gamma - 2) = \tfrac{1}{2}\Sigma_\gamma + 3\ell + \tfrac{1}{2}. \tag{21}$$

If $\ell = 0$, then $\gamma = \Delta a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}$ is conjugate to

$$\gamma_1 = a^{p_1+1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r+1} b$$

and $\hat{\gamma}_1 = \hat{\gamma} = K$, so $g(\hat{\gamma}_1) = \frac{1}{2}\Sigma_\gamma + \frac{1}{2}$. By Lemma 4.4,

$$\Upsilon_K(t) \leq (-g(\hat{\gamma}_1) + r - 1)t = \left(-\tfrac{1}{2}\Sigma_\gamma + r - \tfrac{3}{2}\right)t \quad \text{for all } 0 \leq t \leq 1.$$

For $\ell \geq 1$, using $\Delta^{2\ell+1} = (ab)^{3\ell}aba = (ab)^{3\ell+1}a$,

$$\begin{aligned}
\gamma &= \Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r} \\
&= (ab)^{3\ell+1} a^{p_1+1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r} \\
&= a^2ba^3(aba^3)^{\ell-1} ba^{p_1+\ell+1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r} && \text{(by (20))} \\
&\sim a^{p_r+2}ba^3(aba^3)^{\ell-1} ba^{p_1+\ell+1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} =: \gamma_1.
\end{aligned}$$

We have $\hat{\gamma}_1 = \hat{\gamma} = K$ and $g(\hat{\gamma}_1) = \frac{1}{2}\Sigma_\gamma + 3\ell + \frac{1}{2}$ by (21). Again, Lemma 4.4 implies

$$\Upsilon_K(t) \leq (-g(\hat{\gamma}_1) + r + \ell - 1)t = \left(-\tfrac{1}{2}\Sigma_\gamma + r - 2\ell - \tfrac{3}{2}\right)t \quad \text{for all } 0 \leq t \leq 1,$$

which proves the claim of the lemma. $\square$

Lemma 4.6 *Let $\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\Upsilon_K(t) \leq \left(-\frac{1}{2} \left(\sum_{i=1}^{r} (p_i + q_i) \right) + r - 2\ell \right) t \quad \text{for all } 0 \leq t \leq 1.$$

In the proof, we will need the following statement about positive 3–braids.

Lemma 4.7 *In B_3, we have $(ab)^{3n-1} = a^{2n} b (a^2 b^2)^{n-1} a$ for all $n \geq 1$.*

Proof Starting with the left-hand side,

$$(ab)^{3n-1} = a(ba)^{3(n-1)} bab = a(ab)^{3(n-1)} aba,$$

which proves the lemma for $n = 1$. We now show by induction that

$$(ab)^{3(n-1)} a = a^{2n-1} b (a^2 b^2)^{n-2} a^2 b \quad \text{for all } n \geq 2, \tag{22}$$

which implies the lemma for all $n \geq 1$. For $n = 2$,

$$(ab)^3 a = a(ba)^3 = a(ab)^3 = a^2 babab = a^3 ba^2 b.$$

Assuming that (22) is true for some $n - 1 \geq 2$,

$$\begin{aligned} (ab)^{3(n-1)} a &= a(ba)^{3(n-1)} \\ &= a(ab)^{3(n-1)} \\ &= a^2 (ba)^{3(n-2)} babab \\ &= a^2 (ab)^{3(n-2)} aba^2 b \\ &= a^2 (a^{2n-3} b (a^2 b^2)^{n-3} a^2 b) ba^2 b \\ &= a^{2n-1} b (a^2 b^2)^{n-2} a^2 b, \end{aligned}$$

using the induction hypothesis in the second to last equality. □

Proof of Lemma 4.6 Let $\Sigma_\gamma = \sum_{i=1}^{r} (p_i + q_i)$. If $\ell = 0$, then by (3) and Lemma 4.4,

$$\Upsilon_K(t) \leq (-g(K) + r - 1)t = \left(-\tfrac{1}{2}\Sigma_\gamma + r\right)t \quad \text{for all } 0 \leq t \leq 1.$$

For $\ell \geq 1$, using $\Delta^2 = (ba)^3$ and Lemma 4.7,

$$\begin{aligned} \gamma = (ba)^{3\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r} &\sim (ab)^{3\ell-1} a^{p_1+1} b^{q_1} \cdots a^{p_r} b^{q_r+1} \\ &\sim a^{2\ell} b (a^2 b^2)^{\ell-1} a^{p_1+2} b^{q_1} \cdots a^{p_r} b^{q_r+1} =: \gamma_1. \end{aligned}$$

Note that $\hat{\gamma}_1 = \hat{\gamma} = K$ and by (3),

$$g(\hat{\gamma}_1) = g(K) = \tfrac{1}{2}(6\ell + \Sigma_\gamma - 2) = \tfrac{1}{2}\Sigma_\gamma + 3\ell - 1.$$

Again by Lemma 4.4,

$$\Upsilon_K(t) \le (-g(\hat{\gamma}_1) + r + \ell - 1)t = \left(-\tfrac{1}{2}\Sigma_\gamma + r - 2\ell\right)t \quad \text{for all } 0 \le t \le 1. \qquad \square$$

We will now focus on $\upsilon(K) = \Upsilon_K(1)$ and prove Proposition 4.2 by showing that the upper bounds on $\Upsilon_K(t)$ from Lemmas 4.5 and 4.6 for $t = 1$ are also lower bounds. We will need the following observation used in [20, Example 4.5] about 3–braids, which we prove here for completeness.

Lemma 4.8 *In B_3, $a^{2n+1}b(a^2b^2)^n = (ab)^{3n+1}$ and $b^{2n+1}a(b^2a^2)^n = (ba)^{3n+1}$ for all $n \ge 0$.*

Proof We prove the first statement by induction. For $n = 0$, the equality is clearly true. For $n = 1$, using $\Delta a = b\Delta$ and $\Delta b = a\Delta$, we have

$$a^3ba^2b^2 = a^2\Delta ab^2 = a^2ba\Delta b = a\Delta^2 b = \Delta^2 ab = (ab)^4.$$

We now assume the lemma is true for some $n - 1 \ge 0$. Using the induction hypothesis and the equality for $n = 1$,

$$\begin{aligned} a^{2n+1}b(a^2b^2)^n &= a^2(ab)^{3(n-1)+1}a^2b^2 = a^3b\Delta^{2(n-1)}a^2b^2 \\ &= \Delta^{2(n-1)}a^3ba^2b^2 = (ab)^{3(n-1)}(ab)^4 = (ab)^{3n+1}. \qquad \square \end{aligned}$$

Lemma 4.9 *Let $\gamma = \Delta^{2\ell+1}a^{p_1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r}$ for some $\ell \ge 0$, $r \ge 1$, $p_r \ge 3$ and $p_i, q_i \ge 2$ for $i \in \{1, \dots, r-1\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) = -\frac{1}{2}\left(\sum_{i=1}^{r-1}(p_i + q_i) + p_r\right) + r - 2\ell - \frac{3}{2}.$$

Proof Let $\Sigma_\gamma = \sum_{i=1}^{r-1}(p_i + q_i) + p_r$. From Lemma 4.5, it follows directly that

$$\upsilon(K) = \Upsilon_K(1) \le -\tfrac{1}{2}\Sigma_\gamma + r - 2\ell - \tfrac{3}{2},$$

so we are left to show that $\upsilon(K) \ge -\frac{1}{2}\Sigma_\gamma + r - 2\ell - \frac{3}{2}$. To that end, consider

$$\begin{aligned} \gamma &= \Delta^{2\ell+1}a^{p_1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r} \\ &\sim \Delta^{2\ell}a\Delta a^{p_1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r-1} \\ &= \Delta^{2\ell}bab^2a^{p_1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r-1} =: \gamma_1, \end{aligned}$$

where we used $a\Delta = abab = bab^2$. Note that $\hat{\gamma}_1 = \hat{\gamma} = K$. Now, define

$$\alpha := b^{2r}\gamma_1 = \Delta^{2\ell}b^{2r+1}ab^2a^{p_1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r-1}$$

and note that $\hat{\alpha}$ is a knot. By assumption, $p_r - 1 \geq 2$. There is a cobordism between $\hat{\alpha}$ and the connected sum $T_{2,2r+1} \# \hat{\gamma}_1 = T_{2,2r+1} \# K$ of genus 1 by using two saddle moves similar to the two saddle moves illustrated in Figure 2. Similarly as in (16) from Example 4.1, we have $\upsilon(K) \geq \upsilon(\hat{\alpha}) + r - 1$. In order to find a lower bound for $\upsilon(\hat{\alpha})$, note that there is a cobordism C between $\hat{\alpha}$ and the torus knot $T = T_{3,3(\ell+r)+1}$ of genus $g(C) = \frac{1}{2}\Sigma_\gamma - 2r + \frac{1}{2}$. Here we think of T as the closure of the braid word $\beta = \Delta^{2\ell} b^{2r+1} a (b^2 a^2)^r$, which is equal to $\Delta^{2\ell}(ba)^{3r+1} = (ba)^{3(\ell+r)+1}$ as 3–braids by Lemma 4.8. The cobordism C between $\hat{\alpha}$ and $T = \hat{\beta}$ can thus be realized by

$$p_1 - 2 + q_1 - 2 + \cdots + p_{r-1} - 2 + q_{r-1} - 2 + p_r - 3 = \Sigma_\gamma - 4r + 1$$

saddle moves corresponding to the deletion of the same number of generators a and b from the braid word α to obtain β. Hence the Euler characteristic of the cobordism C is $\chi(C) = -\Sigma_\gamma + 4r - 1$. Since C is connected and has two boundary components (as $\hat{\alpha}$ and $T = \hat{\beta}$ are knots), the genus of C is indeed $g(C) = \frac{1}{2}\Sigma_\gamma - 2r + \frac{1}{2}$. Now, by (15) and (12),

$$\upsilon(\hat{\alpha}) \geq \upsilon(T) - g(C) = -2(\ell + r) - \left(\tfrac{1}{2}\Sigma_\gamma - 2r + \tfrac{1}{2}\right) = -\tfrac{1}{2}\Sigma_\gamma - 2\ell - \tfrac{1}{2}.$$

It follows that

$$\upsilon(K) \geq \upsilon(\hat{\alpha}) + r - 1 \geq -\tfrac{1}{2}\Sigma_\gamma + r - 2\ell - \tfrac{3}{2}. \qquad \square$$

Lemma 4.10 *Let $\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 1$, $p_r, q_r \geq 3$ and $p_i, q_i \geq 2$ for $i \in \{1, \ldots, r-1\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) = -\frac{1}{2}\left(\sum_{i=1}^{r}(p_i + q_i)\right) + r - 2\ell.$$

Proof The proof uses similar ideas as that of Lemma 4.9. Let $\Sigma_\gamma = \sum_{i=1}^{r}(p_i + q_i)$. By Lemma 4.6, $\upsilon(K) \leq -\frac{1}{2}\Sigma_\gamma + r - 2\ell$, so it remains to show that $\upsilon(K) \geq -\frac{1}{2}\Sigma_\gamma + r - 2\ell$. To that end, we consider

$$\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r} \sim \Delta^{2\ell} b a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r - 1} =: \gamma_1.$$

Note that $\hat{\gamma}_1 = \hat{\gamma} = K$. We define

$$\alpha := a^{2r}\gamma_1 = a^{2r}\Delta^{2\ell} b a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r - 1} \sim \Delta^{2\ell} b a^{2r} b a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r - 2} =: \alpha_1.$$

Then $\hat{\alpha}_1 = \hat{\alpha}$ is a knot and by assumption we have $q_r - 2 \geq 1$. There is a cobordism between $\hat{\alpha}$ and $T_{2,2r+1} \# \hat{\gamma}_1 = T_{2,2r+1} \# K$ of genus 1 by using two saddle moves similar to the cobordism considered in Example 4.1 and in the proof of Lemma 4.9;

hence $\upsilon(K) \geq \upsilon(\hat{\alpha}_1) + r - 1$. To find a lower bound for $\upsilon(\hat{\alpha}_1)$, we observe that there is a cobordism C between the knot $\hat{\alpha}_1$ and the knot $\hat{\beta}$, where

$$\beta = \Delta^{2\ell} b a^{2r} b (a^2 b^2)^{r-1} a^3 b.$$

Using Lemma 4.8 for $n-1$, in B_3,

$$ba^{2n} b (a^2 b^2)^{n-1} a^2 = ba(ab)^{3(n-1)+1} a^2 = ba \Delta^{2(n-1)} aba^2 = \Delta^{2n} \quad \text{for all } n \geq 1.$$

We thus have $\beta = \Delta^{2\ell} \Delta^{2r} ab = (ab)^{3(\ell+r)+1}$, so the closure of β is the torus knot $T = T_{3,3(\ell+r)+1}$ with $\upsilon(T) = -2(\ell + r)$ by (12). The cobordism C between $\hat{\alpha}_1$ and $T = \hat{\beta}$ can be realized by

$$p_1 - 2 + q_1 - 2 + \cdots + p_{r-1} - 2 + q_{r-1} - 2 + p_r - 3 + q_r - 3 = \Sigma_\gamma - 4r - 2$$

saddle moves corresponding to the deletion of the same number of generators a and b from the braid word α_1 to obtain β. By a similar Euler characteristic argument as in the proofs of Lemmas 4.4 and 4.9, the genus of this cobordism is $g(C) = \frac{1}{2}\Sigma_\gamma - 2r - 1$. Note that here we used $p_r \geq 3$ and $q_r \geq 3$. Now, by (15),

$$\begin{aligned} \upsilon(\hat{\alpha}_1) &\geq \upsilon(T) - g(C) = -\tfrac{1}{2}\Sigma_\gamma - 2\ell + 1, \\ \upsilon(K) &\geq \upsilon(\hat{\alpha}_1) + r - 1 \geq -\tfrac{1}{2}\Sigma_\gamma + r - 2\ell. \end{aligned}$$

□

Lemma 4.11 *Let $\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 2$ and $p_i, q_i \geq 2$ for $i \in \{1, \ldots, r\}$. Suppose that $q_r \geq 3$ and $p_k \geq 3$ for some $1 \leq k < r$ and that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) = -\frac{1}{2}\left(\sum_{i=1}^{r}(p_i + q_i)\right) + r - 2\ell.$$

Proof We proceed as in the proof of Lemma 4.10, but here we will look at a different cobordism to obtain a lower bound for $\upsilon(\hat{\alpha}_1)$. The steps of the proof are exactly the same until then, so we consider

$$\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r} \sim \Delta^{2\ell} b a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r - 1} =: \gamma_1$$

and define

$$\alpha := a^{2r} \gamma_1 \sim \Delta^{2\ell} b a^{2r} b a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r - 2} =: \alpha_1.$$

Again, we have $\upsilon(K) \geq \upsilon(\hat{\alpha}_1) + r - 1$. Now, in order to find a lower bound for $\upsilon(\hat{\alpha}_1)$, we observe that there is a cobordism C between $\hat{\alpha}_1$ and the knot $\hat{\beta}$, where

$$\beta = \Delta^{2\ell} b a^{2r} b (a^2 b^2)^{k-1} a^3 b^2 (a^2 b^2)^{r-k-1} a^2 b.$$

We find the cobordism C by the deletion of generators from the braid word β to obtain α_1, where we use the assumptions $q_r \geq 3$ and $p_k \geq 3$. In fact, the cobordism can be realized by

$$\begin{aligned} p_1 - 2 + q_1 - 2 + \cdots + p_{k-1} - 2 + q_{k-1} - 2 + p_k - 3 + q_k - 2 \\ + p_{k+1} - 2 + q_{k+1} - 2 + \cdots + p_{r-1} - 2 + q_{r-1} - 2 \\ + p_r - 2 + q_r - 3 = \Sigma_\gamma - 4r - 2 \end{aligned}$$

saddle moves, so its genus is $g(C) = \frac{1}{2}\Sigma_\gamma - 2r - 1$. Using

$$a^{2k-1} b (a^2 b^2)^{k-1} = (ab)^{3k-2}$$

from Lemma 4.8, we have

$$\begin{aligned} \beta &= \Delta^{2\ell} b a^{2r-2k+1} (ab)^{3k-2} a^3 b^2 (a^2 b^2)^{r-k-1} a^2 b \\ &= \Delta^{2\ell} b a^{2r-2k+1} \Delta^{2(k-1)} a b a^3 b^2 (a^2 b^2)^{r-k-1} a^2 b \\ &\sim \Delta^{2(\ell+k-1)} \Delta a^2 b^2 (a^2 b^2)^{r-k-1} a^2 b^2 a^{2r-2k+1} \\ &= \Delta^{2(\ell+k-1)+1} (a^2 b^2)^{r-k+1} a^{2r-2k+1} =: \beta_1. \end{aligned}$$

Note that by our assumptions on ℓ, r and k, we have $\ell + k - 1 \geq 0$, $r - k + 1 \geq 2$ and $2r - 2k + 1 \geq 3$, so β_1 has the form of the braid words considered in Lemma 4.9. We thus have

$$\begin{aligned} \upsilon(\hat{\beta}) = \upsilon(\hat{\beta}_1) &= -\tfrac{1}{2}(4(r-k+1) + 2r - 2k + 1) + (r - k + 2) - 2(\ell + k - 1) - \tfrac{3}{2} \\ &= -2(\ell + r). \end{aligned}$$

By (15),

$$\begin{aligned} \upsilon(\hat{\alpha}_1) &\geq \upsilon(\hat{\beta}) - g(C) = -\tfrac{1}{2}\Sigma_\gamma - 2\ell + 1, \\ \upsilon(K) &\geq \upsilon(\hat{\alpha}_1) + r - 1 \geq -\tfrac{1}{2}\Sigma_\gamma + r - 2\ell. \end{aligned}$$ □

Proof of Proposition 4.2 The first case of Proposition 4.2 follows from Remark 3.6 and (12). Lemmas 4.10 and 4.11 together prove the second case, Lemma 4.9 proves the third case. Note that up to conjugation, by Remark 3.1 and the remarks in Proposition 3.2, it is no restriction to assume that $p_r \geq 3$ in Lemma 4.9 and that $q_r \geq 3$ and either $p_r \geq 3$ or $p_k \geq 3$ for some $1 \leq k < r$ in Lemmas 4.10 and 4.11, respectively. □

Before we proceed with the general case where the knot K is given as the closure of any 3–braid, let us prove the following corollaries of our results in this section.

Corollary 4.12 (Corollary 1.3) *Let K be a knot that is the closure of a positive 3–braid. Then*

$$r = g(K) + \upsilon(K) + 1$$

is minimal among all integers $r \geq 1$ such that K is the closure of a positive 3–braid $a^{p_1}b^{q_1}\cdots a^{p_r}b^{q_r}$ for integers $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$.

Proof By Lemma 4.4,

$$\upsilon(K) \leq -g(K) + r - 1 \iff g(K) + \upsilon(K) + 1 \leq r$$

whenever K is the closure of a positive 3–braid $a^{p_1}b^{q_1}\cdots a^{p_r}b^{q_r}$ for integers $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$. It remains to show that we can always find a positive braid representative for K of the form $a^{p_1}b^{q_1}\cdots a^{p_r}b^{q_r}$ with $r = g(K) + \upsilon(K) + 1$. We will use Proposition 3.2. In fact, if K is the closure of a positive braid γ of the form in (C) with $\ell \geq 0$, then $g(K) + \upsilon(K) + 1 = r + \ell$ by (3) applied to γ and Lemmas 4.10 and 4.11. Moreover,

$$\begin{aligned} \gamma &= a^{p_1}b^{q_1}\cdots a^{p_r}b^{q_r} && \text{if } \ell = 0, \\ \gamma &\sim a^{2\ell}b(a^2b^2)^{\ell-1}a^{p_1+2}b^{q_1}\cdots a^{p_r}b^{q_r+1} && \text{if } \ell \geq 1 \end{aligned}$$

by the proof of Lemma 4.6; these give the desired braid representatives for K. Furthermore, if K is represented by a positive braid γ of the form in (D) with $\ell \geq 0$, then $g(K) + \upsilon(K) + 1 = r + \ell$ by (3) and Lemma 4.9, and we have

$$\begin{aligned} \gamma &\sim a^{p_1+1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}}a^{p_r+1}b && \text{if } \ell = 0, \\ \gamma &\sim a^{p_r+2}ba^3(aba^3)^{\ell-1}ba^{p_1+\ell+1}b^{q_1}\cdots a^{p_{r-1}}b^{q_{r-1}} && \text{if } \ell \geq 1 \end{aligned}$$

by the proof of Lemma 4.5. Finally, if $K = T_{3,3\ell+k}$ for $\ell \geq 0$ and $k \in \{1, 2\}$, then by (4) and (12), we have $g(K) + \upsilon(K) + 1 = \ell + 1$ and $T_{3,3\ell+1}$ and $T_{3,3\ell+2}$ are represented by the positive 3–braids $(ab)^{3\ell+1} = a^{2\ell+1}b(a^2b^2)^{\ell}$ and $(ab)^{3\ell+2} \sim a^{2\ell+3}b(a^2b^2)^{\ell}$, respectively, by Lemmas 4.8 and 4.7. □

Corollary 4.13 (Corollary 1.4) *If K and J are concordant knots that are both closures of positive* 3*–braids, then the minimal r from Corollary 4.12 is the same for both K and J.*

Proof If K and J are concordant, then their 4–genus and their upsilon invariants are equal. So by (3) from Section 2.1 and by Corollary 4.12, positive 3–braids with closures K and J, respectively, will have the same minimal r. □

Remark 4.14 Let $\mathcal{A}_g(K)$ denote the minimal genus of a cobordism between a knot K and an alternating knot, ie the cobordism distance $d(K, \{\text{alternating knots}\})$. By [23, Theorem 8], we have $\frac{1}{2}|\tau(K)+\upsilon(K)| \leq \mathcal{A}_g(K)$ for any knot K. It thus follows from our results in this section that

$$\tfrac{1}{2}(r+\ell-1) \leq \mathcal{A}_g(K) \leq \tfrac{1}{2}(r+\ell-1+\varepsilon)$$

for any knot K that is the closure of a positive 3–braid in Garside normal form (C) or (D), where $\varepsilon \geq 0$ is an integer depending on K. The lower bound uses Proposition 4.2 and (5) from Section 2.2; see also the proof of Corollary 4.12. The upper bound follows from the proofs of Lemmas 4.5 and 4.6; see also the proof of Lemma 4.4. Note that for most positive 3–braid knots, we have $\varepsilon > 0$, so we do not get an equality.

A shorter proof of Lemma 4.4 without cobordisms follows from a result of Abe and Kishimoto on the dealternating number of positive 3–braid knots. Indeed, by (5), (24) and (27),

$$|\Upsilon_K(t)+g(K)t| = |\Upsilon_K(t)+\tau(K)t| \leq \operatorname{alt}(K)t \leq \operatorname{dalt}(K)t \leq (r-1)t \quad \text{for all } 0 \leq t \leq 1.$$

The definitions of the dealternating number $\operatorname{dalt}(K)$ and the alternation number $\operatorname{alt}(K)$ of a knot K and more details on the inequalities used here will be provided in Section 5.

4.3 Proof of Theorem 1.1

It remains to show Theorem 1.1 when K is the closure of a not necessarily positive 3–braid. We first recall a result of Murasugi, which implies that indeed all 3–braid knots except for the torus knots of braid index 3 are covered by Theorem 1.1.

Let γ be a 3–braid. Then, by [45, Proposition 2.1], γ is conjugate to one and only one of the 3–braids

(a) $\Delta^{2\ell}a^p$ or $\Delta^{2\ell+1}$ for $\ell \in \mathbb{Z}$, $p \in \mathbb{Z}$,

(b) $\Delta^{2\ell}ab$ or $\Delta^{2\ell}(ab)^2$ for $\ell \in \mathbb{Z}$,

(c) $\Delta^{2\ell}a^{-p_1}b^{q_1}\cdots a^{-p_r}b^{q_r}$ for $\ell \in \mathbb{Z}$, $r \geq 1$, $p_i, q_i \geq 1$, $i \in \{1, \dots, r\}$.

Definition 4.15 We call a braid word of the form in (a)–(c) a 3–*braid in Murasugi normal form*.

Remark 4.16 The closures of the 3–braids in Murasugi normal form (a) are links of two (if p is odd) or three components and the closures of the 3–braids in Murasugi normal form (b) are the torus knots of braid index 3 (see Remark 3.6).

If $\ell = 0$ in case (c), the braid word $\gamma = a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for integers $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$ gives rise to an alternating braid diagram. If $K = \hat{\gamma}$ is a knot, by Proposition 2.1 we thus have $\upsilon(K) = \frac{1}{2}\sigma(K)$ in that case and the statement of Theorem 1.1 follows directly from a result by Erle on the signature of 3–braid knots.

Proposition 4.17 [17, Theorem 2.6] *Let $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for integers $\ell \in \mathbb{Z}$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\sigma(K) = \sum_{i=1}^{r} (p_i - q_i) - 4\ell.$$

We still need to show Theorem 1.1 when K is the closure of a 3–braid in Murasugi normal form (c) with $\ell \neq 0$. The proof will follow from the following two lemmas.

Lemma 4.18 *Let $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for some $\ell \geq 1$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\Upsilon_K(t) \leq \left(\frac{1}{2} \left(\sum_{i=1}^{r} (p_i - q_i) \right) - 2\ell \right) t \quad \text{for all } 0 \leq t \leq 1.$$

Lemma 4.19 *Let $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \dots, r\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) \geq \frac{1}{2} \left(\sum_{i=1}^{r} (p_i - q_i) \right) - 2\ell.$$

Proof of Theorem 1.1 For $\ell \geq 1$, the statement of the theorem follows directly from Lemmas 4.18 and 4.19. If $\ell < 0$, the knot $-K$ is represented by the braid word $\Delta^{-2\ell} a^{-q_r} b^{p_r} \cdots a^{-q_1} b^{p_1}$ with $-\ell \geq 1$ and accordingly we have

$$\upsilon(-K) = \frac{1}{2} \left(\sum_{i=1}^{r} (q_i - p_i) \right) + 2\ell.$$

Using that $\upsilon(-K) = -\upsilon(K)$ by (9) from Section 2.2, this implies the claim. □

The remainder of this section is devoted to the proofs of the above lemmas.

Proof of Lemma 4.18 We first consider the case where $p_1 \geq 2$ and $\ell \geq 2$. Using $\Delta a^{-1} = ab$ and

$$(ab)^{3n+2} = b^{n+1} a (b^3 ab)^{n-1} b^3 ab^3 \quad \text{for all } n \geq 1$$

from [18, Proof of Proposition 22], we have

$$\begin{aligned}\gamma &= \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r} \\ &= \Delta^{2(\ell-1)+1} aba^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r} \\ &= (ba)^{3(\ell-1)+2} ba^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r} \\ &\sim (ab)^{3(\ell-1)+2} a^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r+1} \\ &\sim a(b^3ab)^{\ell-2} b^3ab^3 a^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r+\ell+1} =: \gamma_1.\end{aligned}$$

Now, we claim that there is a cobordism C of genus $g(C) = \frac{1}{2}(\ell + r - 1 + \varepsilon)$ between the closure K of γ_1 and the connected sum

$$J_\varepsilon = -T_{2,p_1-1-\varepsilon_1} \# -T_{2,p_2-\varepsilon_2} \# \cdots \# -T_{2,p_r-\varepsilon_r} \# T_{2,\sum_{i=1}^r q_i+5\ell-1+\varepsilon_q},$$

where we choose $\varepsilon_1, \ldots, \varepsilon_r, \varepsilon_q \in \{0,1\}$ such that J_ε is a connected sum of torus knots, ie such that $\sum_{i=1}^r q_i + 5\ell - 1 + \varepsilon_q$, $p_1 - 1 - \varepsilon_1$, $p_2 - \varepsilon_2, \ldots, p_r - \varepsilon_r$ are all odd; and $\varepsilon = \varepsilon_q + \sum_{i=1}^r \varepsilon_i$. This cobordism C can be realized using $\ell + r - 1 + \varepsilon$ saddle moves as follows. On the one hand, we add $\sum_{i=1}^r \varepsilon_i$ generators a and ε_q generators b to the braid word γ_1; on the other hand, we perform $\ell + r - 1$ saddle moves of the form as the $r-1$ saddle moves used in the proof of Lemma 4.4 to get a connected sum of torus knots. The Euler characteristic of C is $\chi(C) = -\ell - r + 1 - \varepsilon$. Since C is connected and has two boundary components (as K and J_ε are knots), the genus of C is $g(C) = -\frac{1}{2}\chi(C) = \frac{1}{2}(\ell + r - 1 + \varepsilon)$ as claimed. By (8) and (11),

$$\Upsilon_{J_\varepsilon}(t) = \left(\frac{1}{2}\left(\sum_{i=1}^r (p_i - q_i) - \varepsilon - r - 5\ell + 1\right)\right)t \quad \text{for all } 0 \le t \le 1,$$

and by (15),

$$\Upsilon_K(t) \le \Upsilon_{J_\varepsilon}(t) + g(C)t = \left(\frac{1}{2}\left(\sum_{i=1}^r (p_i - q_i)\right) - 2\ell\right)t \quad \text{for all } 0 \le t \le 1.$$

If $p_1 \ge 2$ and $\ell = 1$, then

$$\gamma \sim (ab)^2 a^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r+1} \sim ab^2 a^{-p_1+1} b^{q_1} \cdots a^{-p_r} b^{q_r+2} =: \gamma_1,$$

and similarly as above, there is a cobordism C of genus $g(C) = \frac{1}{2}(r + \varepsilon)$ between the closure K of γ_1 and the connected sum

$$J_\varepsilon = -T_{2,p_1-1-\varepsilon_1} \# -T_{2,p_2-\varepsilon_2} \# \cdots \# -T_{2,p_r-\varepsilon_r} \# T_{2,\sum_{i=1}^r q_i+4+\varepsilon_q},$$

where we choose $\varepsilon_1, \dots, \varepsilon_r, \varepsilon_q \in \{0, 1\}$ such that J_ε is a connected sum of torus knots and $\varepsilon = \varepsilon_q + \sum_{i=1}^r \varepsilon_i$. The claim follows also in this case from equations (8) and (11), and the inequality in (15).

It remains to show the claim when $p_1 = 1$. In that case, using $\Delta a^{-1} = ab$,

$$\begin{aligned}\gamma &= \Delta^{2\ell} a^{-1} b^{q_1} \cdots a^{-p_r} b^{q_r} \\ &= \Delta^{2\ell-1} a b^{q_1+1} \cdots a^{-p_r} b^{q_r} \sim \Delta^{2\ell-1} b^{q_1+1} \cdots a^{-p_r} b^{q_r+1}.\end{aligned}$$

If $\ell = 1$, then γ is conjugate to $\gamma_1 = ab^{q_1+2} a^{-p_2} b^{q_2} \cdots a^{-p_r} b^{q_r+2}$ and if $\ell \geq 2$, then using (20) from Section 4.2,

$$\begin{aligned}\gamma &\sim \Delta^{2(\ell-1)+1} b^{q_1+1} a^{-p_2} b^{q_2} \cdots a^{-p_r} b^{q_r+1} \\ &= (ba)^{3(\ell-1)+1} b^{q_1+2} a^{-p_2} b^{q_2} \cdots a^{-p_r} b^{q_r+1} \\ &\sim ab^3 (bab^3)^{\ell-2} ab^{q_1+\ell+1} a^{-p_2} b^{q_2} \cdots a^{-p_r} b^{q_r+3} =: \gamma_1.\end{aligned}$$

In both cases, there is a cobordism C of genus $g(C) = \frac{1}{2}(\ell + r - 2 + \varepsilon)$ between the closure K of γ_1 and the connected sum

$$J_\varepsilon = -T_{2,p_2-\varepsilon_2} \# \cdots \# -T_{2,p_r-\varepsilon_r} \# T_{2,\sum_{i=1}^r q_i + 5\ell - 1 + \varepsilon_q},$$

where we choose $\varepsilon_1, \dots, \varepsilon_r, \varepsilon_q \in \{0, 1\}$ such that J_ε is a connected sum of torus knots and $\varepsilon = \varepsilon_q + \sum_{i=1}^r \varepsilon_i$. Using (8), (11), and (15) again, the claim follows. □

We will need the following two technical lemmas for the proof of Lemma 4.19.

Lemma 4.20 *Let $\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 1$ and integers p_i and q_i such that $p_i < 0$ or $p_i \geq 2$, and $q_i < 0$ or $q_i \geq 2$, for any $i \in \{1, \dots, r\}$. Moreover, assume that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) \geq -\frac{1}{2}\left(\sum_{i=1}^r (p_i + q_i)\right) + r - 2\ell - \#\{i \mid p_i < 0\} - \#\{i \mid q_i < 0\},$$

where $\#A$ denotes the cardinality of the set A.

Lemma 4.21 *Let $\gamma = \Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}$ for some $\ell \geq 0$, $r \geq 1$ and integers p_i and q_i such that $p_i < 0$ or $p_i \geq 2$ for any $i \in \{1, \dots, r\}$ and $q_i < 0$ or $q_i \geq 2$ for any $i \in \{1, \dots, r-1\}$. Moreover, assume that $K = \hat{\gamma}$ is a knot. Then*

$$\upsilon(K) \geq -\frac{1}{2}\left(\sum_{i=1}^{r-1} (p_i + q_i) + p_r\right) + r - 2\ell - \frac{3}{2} - \#\{i \mid p_i < 0\} - \#\{i \mid q_i < 0\}.$$

For the proofs of Lemmas 4.20 and 4.21, we refer the reader to the very end of this section; we will first prove Lemma 4.19 using these lemmas.

Proof of Lemma 4.19 Let k be the number of exponents q_j of γ with $q_j = 1$ and let $\mathcal{J} = \{j_1, \dots, j_k\}$ for $0 \le k \le r$ be the set of indices such that $q_j = 1$ if and only if $j \in \mathcal{J}$. For all $j \in \mathcal{J}$, we rewrite the subword $a^{-p_j} b^{q_j}$ of γ using $\Delta^{-1}ab = a^{-1}$ as

$$a^{-p_j} b^{q_j} = a^{-p_j} b = a^{-p_j} a^{-1} \Delta\Delta^{-1} ab = a^{-p_j-1} \Delta a^{-1} = \Delta b^{-p_j-1} a^{-1}.$$

Note that if $j, j+1 \in \mathcal{J}$, then $a^{-p_j} b^{q_j} a^{-p_{j+1}} b^{q_{j+1}} = \Delta^2 a^{-p_j-1} b^{-p_{j+1}-2} a^{-1}$. After rewriting $a^{-p_j} b^{q_j}$ for all $j \in \mathcal{J}$, the braid γ is conjugate to $\gamma_1 = \Delta^{2\ell+k}\alpha$ for some 3–braid α which is of the form

$$\alpha = \begin{cases} a^{\tilde{p}_1} b^{\tilde{q}_1} \cdots a^{\tilde{p}_n} b^{\tilde{q}_n} \text{ for } n = r - \frac{1}{2}k & \text{if } k \text{ is even,} \\ b^{\tilde{p}_1} a^{\tilde{q}_1} \cdots b^{\tilde{p}_{n-1}} a^{\tilde{q}_{n-1}} b^{\tilde{p}_n} \text{ for } n = r - \frac{1}{2}(k-1) & \text{if } k \text{ is odd,} \end{cases}$$

where $\sum_{i=1}^{n} (\tilde{p}_i + \tilde{q}_i) = \sum_{i=1}^{r} (-p_i + q_i) - 3k$ and where the $\tilde{p}_i$ and $\tilde{q}_i$ fulfill the assumptions of Lemmas 4.20 and 4.21, respectively, ie where $\tilde{p}_i < 0$ or ≥ 2 and $\tilde{q}_i < 0$ or ≥ 2 for any i. The number of negative exponents in α equals the number of negative exponents $-p_i$ in γ, so

$$\#\{i \mid \tilde{p}_i < 0\} + \#\{i \mid \tilde{q}_i < 0\} = r.$$

If k is even, by Lemma 4.20,

$$\begin{aligned} \upsilon(\hat{\gamma}) &\ge -\frac{1}{2}\left(\sum_{i=1}^{n} (\tilde{p}_i + \tilde{q}_i)\right) + n - (2\ell + k) - \#\{i \mid \tilde{p}_i < 0\} + \#\{i \mid \tilde{q}_i < 0\} \\ &= -\frac{1}{2}\left(\sum_{i=1}^{r} (-p_i + q_i) - 3k\right) + r - \frac{k}{2} - (2\ell + k) - r \\ &= \frac{1}{2}\left(\sum_{i=1}^{r} (p_i - q_i)\right) - 2\ell. \end{aligned}$$

Similarly, if k is odd, the claim follows from Lemma 4.21. □

It remains to prove Lemmas 4.20 and 4.21.

Proof of Lemma 4.20 We will modify the braid word γ in $2r$ steps, where each step corresponds to one of the $2r$ exponents p_i or q_i, for $i \in \{1, \dots, r\}$, of γ. In every step, we will either just conjugate γ (if the corresponding exponent is positive) or perform a cobordism of genus 1 between the closure of $a^{2n}\gamma$ or $b^{2n}\gamma$ and the connected sum $T_{2,2n+1} \# \hat{\gamma}$ for some $n \ge 0$ — similar to the cobordism described in Example 4.1 and used in the proofs of Lemmas 4.9, 4.10 and 4.11. We now describe these steps in more detail. First, let $\gamma'_{0,q} = \gamma$ and define

$$\begin{aligned} a^{-p_1+2+\varepsilon_{1,p}}\gamma'_{0,q} &= \Delta^{2\ell}a^{2+\varepsilon_{1,p}}b^{q_1}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r} \\ &\sim \Delta^{2\ell}b^{q_1}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{2+\varepsilon_{1,p}} =: \gamma'_{1,p} \quad \text{if } p_1 < 0, \\ \gamma'_{0,q} &\sim \Delta^{2\ell}b^{q_1}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{p_1} =: \gamma'_{1,p} \qquad \text{if } p_1 > 0, \end{aligned}$$

so that $\gamma'_{1,p} = \Delta^{2\ell}b^{q_1}a^{p_2}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}$ for some $\tilde{p}_1 \geq 2$ (note that we assumed $p_1 < 0$ or $p_1 \geq 2$). Here, if $p_1 < 0$, we choose $\varepsilon_{1,p} \in \{0,1\}$ such that $-p_1 + 2 + \varepsilon_{1,p}$ is even and $\hat{\gamma}'_{1,p}$ is a knot. Second, let $\varepsilon_{1,q} \in \{0,1\}$ be such that $-q_1 + 2 + \varepsilon_{1,q}$ is even if $q_1 < 0$, and define

$$\begin{aligned} \gamma_{1,q} &= b^{-q_1+2+\varepsilon_{1,q}}\gamma'_{1,p} = \Delta^{2\ell}b^{2+\varepsilon_{1,q}}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1} \\ &\sim \Delta^{2\ell}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{2+\varepsilon_{1,q}} =: \gamma'_{1,q} \qquad \text{if } q_1 < 0, \\ \gamma_{1,q} &= \gamma'_{1,p} \sim \Delta^{2\ell}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{q_1} =: \gamma'_{1,q} \qquad \text{if } q_1 > 0, \end{aligned}$$

so that $\gamma'_{1,q} = \Delta^{2\ell}a^{p_2}b^{q_2}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{\tilde{q}_1}$ for some $\tilde{p}_1, \tilde{q}_1 \geq 2$. Inductively, for any $1 \leq i \leq r$, we let

$$\begin{aligned} &a^{-p_i+2+\varepsilon_{i,p}}\gamma'_{i-1,q} \\ &\qquad = \Delta^{2\ell}a^{2+\varepsilon_{i,p}}b^{q_i}a^{p_{i+1}}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{\tilde{q}_1}\cdots a^{\tilde{p}_{i-1}}b^{\tilde{q}_{i-1}} \\ &\qquad \sim \Delta^{2\ell}b^{q_i}a^{p_{i+1}}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{\tilde{q}_1}\cdots a^{\tilde{p}_{i-1}}b^{\tilde{q}_{i-1}}a^{2+\varepsilon_{i,p}} =: \gamma'_{i,p} \quad \text{if } p_i < 0, \\ &\gamma'_{i-1,q} \sim \Delta^{2\ell}b^{q_i}a^{p_{i+1}}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{\tilde{q}_1}\cdots a^{\tilde{p}_{i-1}}b^{\tilde{q}_{i-1}}a^{p_i} =: \gamma'_{i,p} \qquad \text{if } p_i > 0, \end{aligned}$$

so that

$$\gamma'_{i,p} = \Delta^{2\ell}b^{q_i}a^{p_{i+1}}\cdots a^{p_r}b^{q_r}a^{\tilde{p}_1}b^{\tilde{q}_1}\cdots a^{\tilde{p}_{i-1}}b^{\tilde{q}_{i-1}}a^{\tilde{p}_i}$$

for some integers $\tilde{p}_1, \tilde{q}_1, \ldots, \tilde{p}_{i-1}, \tilde{q}_{i-1}, \tilde{p}_i \geq 2$. Here we choose $\varepsilon_{i,p} \in \{0,1\}$ such that $-p_i + 2 + \varepsilon_{i,p}$ is even if $p_i < 0$. Moreover, for $1 \leq i \leq r$, we let $\varepsilon_{i,q} \in \{0,1\}$ be such that $-q_i + 2 + \varepsilon_{i,q}$ is even, and define

$$\gamma_{i,q} = \begin{cases} b^{-q_i+2+\varepsilon_{i,q}}\gamma'_{i,p} & \text{if } q_i < 0, \\ \gamma'_{i,p} & \text{if } q_i > 0, \end{cases}$$

and we define $\gamma'_{i,q}$ similarly as $\gamma'_{1,q}$. Inductively, after $2r$ steps, we get the positive 3–braid

$$\gamma'_{r,q} = \Delta^{2\ell}a^{\tilde{p}_1}b^{\tilde{q}_1}\cdots a^{\tilde{p}_r}b^{\tilde{q}_r}$$

with

$$\tilde{p}_i = \begin{cases} 2+\varepsilon_{i,p} & \text{if } p_i < 0, \\ p_i & \text{if } p_i > 0, \end{cases} \qquad \text{and} \qquad \tilde{q}_i = \begin{cases} 2+\varepsilon_{i,q} & \text{if } q_i < 0, \\ q_i & \text{if } q_i > 0, \end{cases}$$

for all $1 \le i \le r$; so $\tilde{p}_1, \tilde{q}_1, \dots, \tilde{p}_r, \tilde{q}_r \ge 2$. By Proposition 4.2,

$$\upsilon(\hat{\gamma}'_{r,q}) = -\frac{1}{2}\Biggl(\sum_{\substack{i=1\\p_i>0}}^{r} p_i + \sum_{\substack{i=1\\q_i>0}}^{r} q_i + \sum_{\substack{i=1\\p_i<0}}^{r} (2+\varepsilon_{i,p}) + \sum_{\substack{i=1\\q_i<0}}^{r} (2+\varepsilon_{i,q})\Biggr) + r - 2\ell.$$

Now, note that if $p_i < 0$ for some $1 \le i \le r$, then there is a cobordism of genus 1 between $\hat{\gamma}'_{i,p}$ and $T_{2,2m+1} \# \hat{\gamma}'_{i-1,q}$ by using two saddle moves, where $m = \frac{1}{2}(-p_i + 2 + \varepsilon_{i,p})$, so similarly as in (16) from Example 4.1, we have

$$\upsilon(\hat{\gamma}'_{i-1,q}) \ge \upsilon(\hat{\gamma}'_{i,p}) + m - 1 = \upsilon(\hat{\gamma}'_{i,p}) + \tfrac{1}{2}(-p_i + \varepsilon_{i,p}).$$

Similarly, if $q_i < 0$ for some $1 \le i \le r$, then $\upsilon(\hat{\gamma}'_{i,p}) \ge \upsilon(\hat{\gamma}'_{i,q}) + \frac{1}{2}(-q_i + \varepsilon_{i,q})$. In addition, if $p_i > 0$, then $\upsilon(\hat{\gamma}'_{i,p}) = \upsilon(\hat{\gamma}'_{i-1,q})$, and if $q_i > 0$, then $\upsilon(\hat{\gamma}'_{i,q}) = \upsilon(\hat{\gamma}'_{i,p})$. We conclude that

$$\begin{aligned}
\upsilon(\hat{\gamma}) = \upsilon(\hat{\gamma}'_{0,q}) &\ge \upsilon(\hat{\gamma}'_{r,q}) + \sum_{\substack{i=1\\p_i<0}}^{r} \frac{-p_i + \varepsilon_{i,p}}{2} + \sum_{\substack{i=1\\q_i<0}}^{r} \frac{-q_i + \varepsilon_{i,q}}{2} \\
&= -\frac{1}{2}\Biggl(\sum_{\substack{i=1\\p_i>0}}^{r} p_i + \sum_{\substack{i=1\\q_i>0}}^{r} q_i + \sum_{\substack{i=1\\p_i<0}}^{r} (p_i+2) + \sum_{\substack{i=1\\q_i<0}}^{r} (q_i+2)\Biggr) + r - 2\ell \\
&= -\frac{1}{2}\Biggl(\sum_{i=1}^{r} (p_i + q_i)\Biggr) + r - 2\ell - \#\{i \mid p_i < 0\} - \#\{i \mid q_i < 0\}.
\end{aligned}$$

□

Proof of Lemma 4.21 The strategy of the proof is the same as in the proof of Lemma 4.20. Here, we need $2r-1$ steps corresponding to the $2r-1$ exponents $p_1, q_1, \dots, p_{r-1}, q_{r-1}, p_r$ of γ. The steps are similar to the proof of Lemma 4.20, the only change is that we multiply $\gamma'_{i-1,q}$ by a power of b if $p_i < 0$, and $\gamma'_{i,p}$ by a power of a if $q_i < 0$ (since $a\Delta^{2\ell+1} = \Delta^{2\ell+1}b$ and $b\Delta^{2\ell+1} = \Delta^{2\ell+1}a$). Thus, starting with $\gamma'_{0,q} = \gamma$, after $2r-1$ steps we obtain the positive 3–braid

$$\gamma'_{r,p} = \Delta^{2\ell+1} a^{\tilde{p}_1} b^{\tilde{q}_1} \cdots a^{\tilde{p}_{r-1}} b^{\tilde{q}_{r-1}} a^{\tilde{p}_r}$$

with

$$\tilde{p}_i = \begin{cases} 2+\varepsilon_{i,p} & \text{if } p_i < 0, \\ p_i & \text{if } p_i > 0, \end{cases} \quad \text{and} \quad \tilde{q}_i = \begin{cases} 2+\varepsilon_{i,q} & \text{if } q_i < 0, \\ q_i & \text{if } q_i > 0. \end{cases}$$

By Lemma 4.9,

$$\upsilon(\gamma'_{r,p}) = -\frac{1}{2}\Biggl(\sum_{\substack{i=1\\p_i>0}}^{r} p_i + \sum_{\substack{i=1\\q_i>0}}^{r-1} q_i + \sum_{\substack{i=1\\p_i<0}}^{r} (2+\varepsilon_{i,p}) + \sum_{\substack{i=1\\q_i<0}}^{r-1} (2+\varepsilon_{i,q})\Biggr) + r - 2\ell - \frac{3}{2}.$$

Since the steps we performed have similar effects on $\upsilon(\hat{\gamma})$ as the ones in the proof of Lemma 4.20, we get

$$\begin{aligned}\upsilon(\hat{\gamma}) = \upsilon(\hat{\gamma}'_{0,q}) &\geq \upsilon(\hat{\gamma}'_{r,p}) + \sum_{\substack{i=1 \\ p_i<0}}^{r} \frac{-p_i+\varepsilon_{i,p}}{2} + \sum_{\substack{i=1 \\ q_i<0}}^{r-1} \frac{-q_i+\varepsilon_{i,q}}{2} \\ &= -\frac{1}{2}\Biggl(\sum_{\substack{i=1 \\ p_i>0}}^{r} p_i + \sum_{\substack{i=1 \\ q_i>0}}^{r-1} q_i + \sum_{\substack{i=1 \\ p_i<0}}^{r} (p_i+2) + \sum_{\substack{i=1 \\ q_i<0}}^{r-1} (q_i+2)\Biggr) + r - 2\ell - \frac{3}{2} \\ &= -\frac{1}{2}\Biggl(\sum_{i=1}^{r-1}(p_i+q_i) + p_r\Biggr) + r - 2\ell - \frac{3}{2} - \#\{i \mid p_i<0\} - \#\{i \mid q_i<0\}. \qquad \square\end{aligned}$$

4.4 Further discussion of Theorem 1.1

In this section, we provide some further context on our main result. In particular, in Section 4.4.2 we will discuss why it might be surprising that our proof strategy works for all 3–braid knots.

4.4.1 Comparison of upsilon and the classical signature By Theorem 1.1 and Proposition 4.17,

$$\sigma(K) = 2\upsilon(K) \tag{23}$$

for any knot K that is the closure of a 3–braid $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for integers $\ell \in \mathbb{Z}$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \ldots, r\}$. Computations of the signature for torus knots (and links) of braid index 3, first done by Hirzebruch, Murasugi and Shinora [45, Proposition 9.1, pages 34–35], together with (12) from Section 2.2 imply that the equality in (23) is in fact true for all 3–braid knots K except for the cases that $K = \pm T_{3,3\ell+1}$ for odd $\ell > 0$ or $K = \pm T_{3,3\ell+2}$ for odd $\ell > 0$. In the exceptional cases, we have $\sigma(K) = 2\upsilon(K) - 2$. As mentioned in the introduction, this improves the inequality $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right| \leq 2$ for all 3–braid knots K in [20, Proposition 4.4].

It was shown in [47, Theorem 1.2] that $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right|$ gives a lower bound on the *nonorientable smooth* 4*–genus* of a knot K, denoted by $\gamma_4(K)$, the minimal first Betti number of a nonorientable surface in B^4 that meets the boundary S^3 along K. The similarity of the invariant υ and the classical signature σ on 3–braid knots K described above clearly does not lead to a good lower bound on $\gamma_4(K)$.

However, the equality $\sigma(K) = 2\upsilon(K)$ for most 3–braid knots is actually no great surprise when noting that in fact $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right| \leq 1$ must be true for all 3–braid knots K for the following reason. It is not hard to see that for every 3–braid knot K, there is a nonorientable band move to a 2–bridge knot J, which is alternating [26]. This implies that the *nonorientable cobordism distance* $d_\gamma(K, J) = \gamma_4(K \# -J)$ between K and J is bounded from above by 1. On the other hand, using that υ and σ induce homomorphisms $\mathcal{C} \to \mathbb{Z}$ (see Section 2.2 and [44]), the inequality $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right| \le \gamma_4(K)$ implies that

$$\left|\upsilon(K) - \tfrac{1}{2}\sigma(K)\right| = \left|\upsilon(K \# -J) - \tfrac{1}{2}\sigma(K \# -J)\right| \le d_\gamma(K, J) \le 1,$$

where we used $\upsilon(J) = \frac{1}{2}\sigma(J)$ by Proposition 2.1.

Note that a similar argument shows that $\left|\upsilon(K) - \frac{1}{2}\sigma(K)\right| \le 2$ for all 4–braid knots K, using two nonorientable band moves to transform K into a 2–bridge link, which is also alternating.

4.4.2 On the proof technique As mentioned in the introduction, it came as a surprise to the author that our proof strategy works not only for positive 3–braid knots, but for all 3–braid knots. Let us make this more precise.

The proofs in Sections 4.2 and 4.3 imply, for any 3–braid knot K, the existence of cobordisms C_1 and C_2 of genus $g(C_1)$ and $g(C_2)$ between K and (connected sums of) torus knots T_1 and T_2, respectively, such that

$$g(C_1) + g(C_2) = |\upsilon(T_2) - \upsilon(T_1)|$$

and

$$\upsilon(K) = \upsilon(T_1) + g(C_1) = \upsilon(T_2) - g(C_2).$$

For example, for knots K that are closures of positive 3–braids of Garside normal form (D), the proof of Lemma 4.5 shows the existence of such a cobordism C_1 for $T_1 = J_\varepsilon$ as in the proof of Lemma 4.4; and the existence of such a cobordism C_2 between K and $T_2 = T_{3,3(\ell+r)+1} \# -T_{2,2r+1}$ follows from the proof of Lemma 4.9.

The same strategy would work to determine the concordance invariants s and τ for all positive 3–braid knots K. Indeed, every positive 3–braid knot can be realized as the slice of a cobordism C between the unknot U and a torus knot T of braid index 3 such that $g(C) = |\tau(U) - \tau(T)| = |s(U) - s(T)|$ [21, Proposition 4.1]. However, in contrast, there are 3–braid knots where this strategy provably fails to determine s and τ. A concrete example is the 3–braid knot 10_{125} — the closure of $a^{-5}ba^3b$ [36] — which is not squeezed [21, Example 3.1]. This means that every cobordism C between two connected sums of torus knots T_1 and T_2 that has 10_{125} as a slice satisfies $g(C) > |\tau(T_2) - \tau(T_1)| = |s(T_2) - s(T_1)|$.

4.4.3 Comparison of the normal forms for 3–braids An algorithm described in [11, Section 7] as Schreier's solution to the conjugacy problem [52] can be used to convert 3–braids in Garside normal form (see Definition 3.4) to 3–braids in Murasugi normal form (see Definition 4.15): if γ is a 3–braid of Garside normal form (C), then

$$\gamma \sim \Delta^{2(\ell+r)} a^{-1} b^{p_1-2} a^{-1} b^{q_1-2} \cdots a^{-1} b^{p_r-2} a^{-1} b^{q_r-2},$$

and if γ is of Garside normal form (D), then

$$\gamma \sim \Delta^{2(\ell+r)} a^{-1} b^{p_1-2} a^{-1} b^{q_1-2} \cdots a^{-1} b^{p_{r-1}-2} a^{-1} b^{q_{r-1}-2} a^{-1} b^{p_r-2}.$$

In addition, it is easy to see how 3–braids of Garside normal form (A) or (B) are conjugate to braids of Murasugi normal form (a) or (b).

5 On alternating distances of 3–braid knots

In this section, we prove Corollary 1.2 from the introduction and provide lower and upper bounds on the alternation number and dealternating number of any 3–braid knot which differ by 1.

5.1 Alternating distances of positive 3–braid knots

We will prove the following proposition.

Proposition 5.1 *Let K be a knot that is the closure of a positive 3–braid. Then*

$$\begin{aligned} \operatorname{alt}(K) = \operatorname{dalt}(K) &= \tau(K) + \upsilon(K) \\ &= \begin{cases} \ell & \text{if } K \text{ is the torus knot } T_{3,3\ell+k} \text{ for } \ell \geq 0 \text{ and } k \in \{1,2\}, \\ r+\ell-1 & \text{if } K \text{ is the closure of a braid of the form in (C) or (D)}, \end{cases} \end{aligned}$$

where (C) and (D) refer to the Garside normal forms from Proposition 3.2.

Remark 5.2 Some of the cases in Proposition 5.1 have already been proved by other authors. Indeed, Feller, Pohlmann and Zentner used the observation (25) below to show that $\operatorname{alt}(T_{3,3\ell+k}) = \ell$ for all $\ell \geq 0$ and $k \in \{1,2\}$ [22, Theorem 1.1]. The upper bound they used was provided by [30, Theorem 8]; in fact, the equality had already been shown by Kanenobu in half of the cases, namely when ℓ is even. Moreover, Abe and Kishimoto [2, Theorem 3.1] showed that $\operatorname{alt}(K) = \operatorname{dalt}(K) = r+\ell-1$ if K is a knot that is the closure of a positive 3–braid of the form in (C). However, to the best of this author's knowledge, it is new that $\operatorname{alt}(K) = g(K) + \upsilon(K)$ for all positive 3–braid knots K. Recall that $\tau(K) = g(K)$ for all positive 3–braid knots K by (5) from Section 2.1.

Before we prove Proposition 5.1, let us provide the necessary definitions and background. The *Gordian distance* $d_G(K, J)$ between two knots K and J is the minimal number of crossing changes needed to transform a diagram of K into a diagram of J, where the minimum is taken over all diagrams of K [43]. The *alternation number* $\mathrm{alt}(K)$ of a knot K is defined as the minimal Gordian distance of the knot K to the set of alternating knots [31], ie

$$\mathrm{alt}(K) = \min\{d_G(K, J) \mid J \text{ is an alternating knot}\}.$$

The *dealternating number* $\mathrm{dalt}(K)$ of a knot K is defined via a more diagrammatic approach [3]: it is the minimal number n such that K has a diagram that can be turned into an alternating diagram by n crossing changes. It follows from the definitions that

$$\mathrm{alt}(K) \leq \mathrm{dalt}(K) \tag{24}$$

for any knot K and $\mathrm{alt}(K) = \mathrm{dalt}(K) = 0$ if and only if K is alternating. Note that there are families of knots for which the difference between the alternation number and the dealternating number becomes arbitrarily large [38, Theorem 1.1].

In the proof of Proposition 5.1, we will use that

$$|\tau(K) + \upsilon(K)| \leq \mathrm{alt}(K) \tag{25}$$

for any knot K. In fact, for all alternating knots K,

$$\tau(K) = \tfrac{1}{2}s(K) = -\upsilon(K) = -\tfrac{1}{t}\Upsilon_K(t) = -\tfrac{1}{2}\sigma(K) \tag{26}$$

for any $t \in (0, 1]$ — see [46, Theorem 1.14; 48, Theorem 1.4; 50, Theorem 3] — where s denotes Rasmussen's concordance invariant from Khovanov homology [50]. It follows from [1, Theorem 2.1] — which builds on ideas of Livingston [34, Corollary 3] — that the absolute value of the difference of any two of the invariants in (26) is a lower bound on $\mathrm{alt}(K)$. It was first observed in [22] that the upsilon invariant fits very well in this context; see also [23, Lemma 8].

Another main ingredient of our proof of Proposition 5.1 is the inequality

$$\mathrm{dalt}(\hat{\gamma}) \leq r - 1 \tag{27}$$

for any positive 3–braid $\gamma = a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ with integers $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \ldots, r\}$ [2, Lemma 2.2].

Proof of Proposition 5.1 Let K be a knot that is the closure of a positive 3–braid γ of the form in (C) or (D) from Proposition 3.2 with $\ell \geq 0$. We claim that

$$r + \ell - 1 = \tau(K) + \upsilon(K) = |\tau(K) + \upsilon(K)| \leq \mathrm{alt}(K) \leq \mathrm{dalt}(K) \leq r + \ell - 1, \tag{28}$$

which implies the statement of the proposition for these knots. The two equalities in (28) directly follow from our computations of $\upsilon(K)$ in Proposition 4.2 and (5) applied to γ. The first two inequalities are direct consequences of the inequalities (25) and (24). Finally, the last inequality follows from inequality (27) applied to the particular braid representatives of K considered in the proof of Corollary 4.12.

For torus knots of braid index 3, the statement follows analogously. More precisely, if $K = T_{3,3\ell+k}$ for $\ell \geq 0$ and $k \in \{1,2\}$, then by (4) and (12), we have $|\tau(K)+\upsilon(K)| = \ell$. In addition, the inequality in (27) applied to the particular braid representatives of K considered in the proof of Corollary 4.12 implies that $\operatorname{dalt}(T_{3,3\ell+k}) \leq \ell$. □

From Proposition 5.1, it is easy to deduce that the alternating positive 3–braid knots are precisely the unknot and the connected sums $T_{2,2p+1} \# T_{2,2q+1}$ of two torus knots of braid index 2 for $p, q \geq 0$. This was already known; in fact, the stronger statement is true that the only prime alternating positive braid knots are the torus knots of braid index 2 [6, Corollary 3]. Note that by [42] — see also [11, Corollary 7.2] — the only composite 3–braid knots are the connected sums $T_{2,2p+1} \# T_{2,2q+1}$ for $p, q \in \mathbb{Z}$.

By [1, Theorem 1.1], the only torus knots with alternation number 1 are the torus knots $T_{3,4}$ and $T_{3,5}$. A knot with dealternating number 1 is called *almost alternating*.

Corollary 5.3 *A positive 3–braid knot is almost alternating if and only if it is one of the torus knots $T_{3,4}$ or $T_{3,5}$, or it is represented by a braid of the form*

$$a^{p_1} b^{q_1} a^{p_2} b^{q_2}, \quad \Delta a^{p_1} b^{q_1} a^{p_2}, \quad \Delta^2 a^{p_1} b^{q_1} \quad \text{or} \quad \Delta^3 a^{p_1}$$

for some integers $p_1, p_2, q_1, q_2 \geq 2$.

Proof This follows directly from Proposition 5.1. □

Remark 5.4 In particular, the seven positive 3–braid knots with crossing number 12 — see [36] — are all almost alternating.

Remark 5.5 Our results imply that the Turaev genus equals the alternation number for all positive 3–braid knots. Indeed, let K be a knot that is the closure of a positive braid of the form in (C) or (D) with $\ell \geq 0$. Then we have

$$g_T(K) = \operatorname{alt}(K) = \operatorname{dalt}(K) = r + \ell - 1, \tag{29}$$

where $g_T(K)$ denotes the Turaev genus of the knot K. The *Turaev genus* $g_T(K)$ of a knot K is another alternating distance [38], which was first defined in [15] as the minimal genus of a Turaev surface $F(D)$, where the minimum is taken over all diagrams D of K. The Turaev surface $F(D)$ is a closed orientable surface embedded in S^3 associated to the diagram D. It is formed by building the natural cobordism between the circles in the two extreme Kauffman states (the *all-A–state* and the *all-B–state*) of the diagram D via adding saddles for each crossing of D, and then capping off the boundary components with disks. More details on the definition can be found, for example, in a survey by Champanerkar and Kofman [13].

The equality $g_T(K) = \mathrm{dalt}(K)$ in (29) easily follows from Proposition 5.1, the inequalities $\left|\tau(K) + \frac{1}{2}\sigma(K)\right| \le g_T(K)$ [16, Theorem 1.1] and $g_T(K) \le \mathrm{dalt}(K)$ [2, Corollary 5.4], and the fact that $\sigma(K) = 2\upsilon(K)$ for all knots that are closures of positive braids of Garside normal form (C) or (D) (see Section 4.4.1).

It is not known whether the alternation number and the Turaev genus of a knot are in general comparable; namely, it is not known whether $\mathrm{alt}(K) \le g_T(K)$ for all knots K — see [38, Question 3]. However, it was shown by Abe and Kishimoto that $g_T(T_{3,3\ell+k}) = \mathrm{dalt}(T_{3,3\ell+k}) = \ell$ for all $\ell \ge 0$ and $k \in \{1,2\}$ [2, Theorem 5.9], so $g_T(K) = \mathrm{alt}(K) = \mathrm{dalt}(K)$ is true for all positive 3–braid knots.

Remark 5.6 In [23], Friedl, Livingston and Zentner introduce the invariant $\mathcal{A}_s(K)$, the minimal number of double point singularities in a generically immersed concordance from a knot K to an alternating knot. In the case that the alternating knot is the unknot, this is the well studied invariant $c_4(K)$ called the *4–dimensional clasp number* [53]. A sequence of crossing changes in a diagram of a knot K leading to a diagram of an alternating knot J realizes an immersed concordance from K to J where any crossing change gives rise to a double point singularity in the concordance. We thus have $\mathcal{A}_s(K) \le \mathrm{alt}(K)$ for any knot K, which resembles the inequality $c_4(K) \le u(K)$ between the 4–dimensional clasp number and the unknotting number $u(K)$ of K. Moreover, we have $|\upsilon(K) + \tau(K)| \le \mathcal{A}_s(K)$ for any knot K [23, Theorem 18], so Proposition 5.1 implies $\mathcal{A}_s(K) = \mathrm{alt}(K)$ for all positive 3–braid knots K.

We are now ready to prove Corollary 1.2 from the introduction.

Proof of Corollary 1.2 The corollary follows directly from Proposition 5.1 and Remarks 5.5 and 5.6. □

5.2 Bounds on the alternation number of general 3–braid knots

In the following, we turn our attention to 3–braid knots in general, which are not necessarily the closure of positive 3–braids. We will use that

$$\left|\tfrac{1}{2}s(K)+\upsilon(K)\right| \le \mathrm{alt}(K) \tag{30}$$

for any knot K, which follows from [1, Theorem 2.1]; see also (26) from Section 5.1. Rasmussen's invariant s was computed for all 3–braid knots in Murasugi normal form (see Definition 4.15) by Greene.[2]

Corollary 5.7 *Let $\gamma = \Delta^{2\ell} a^{-p_1} b^{q_1} \cdots a^{-p_r} b^{q_r}$ for some $\ell \in \mathbb{Z}$, $r \ge 1$ and $p_i, q_i \ge 1$ for $i \in \{1, \dots, r\}$ such that $K = \hat{\gamma}$ is a knot. Then*

$$|\ell| - 1 \le \mathrm{alt}(K) \le \mathrm{dalt}(K) \le |\ell| \quad \text{if } \ell \ne 0.$$

Proof The lower bound on the alternation number follows from (30), Theorem 1.1 and the values of the invariant s for $K = \hat{\gamma}$ [27, Proposition 2.4]; namely

$$s(K) = \begin{cases} -\sum_{i=1}^{r}(p_i - q_i) + 6\ell - 2 & \text{if } \ell > 0, \\ -\sum_{i=1}^{r}(p_i - q_i) + 6\ell + 2 & \text{if } \ell < 0. \end{cases}$$

Moreover, it follows from [2, Theorem 2.5] that $\mathrm{dalt}(\hat{\gamma}) \le |\ell|$. □

Remark 5.8 An alternative way to prove the upper bound on $\mathrm{dalt}(K)$ in Corollary 5.7 for $\ell \ge 1$ follows from our observations in the proof of Lemma 4.18. In fact, the braid diagrams given by the braid representatives γ_1 of $K = \hat{\gamma}$ considered in that proof can easily be transformed into alternating diagrams by ℓ crossing changes: it is enough to change the positive crossings corresponding to the single generators a in γ_1 to negative crossings; we obtain generators a^{-1} in the corresponding braid words which then correspond to alternating braid diagrams.

Remark 5.9 If K is represented by a 3–braid of Garside normal form (C) or (D) (see Definition 3.4), then using the observations in Section 4.4.3, Corollary 5.7 implies

$$\begin{aligned} |r+\ell| - 1 \le \mathrm{alt}(K) \le \mathrm{dalt}(K) \le |r+\ell| \quad & \text{if } |r+\ell| > 0, \\ \mathrm{alt}(K) = \mathrm{dalt}(K) = 0 \qquad\qquad & \text{if } r+\ell = 0. \end{aligned} \tag{31}$$

By Proposition 5.1, the lower bound in (31) is sharp whenever K is the closure of a positive 3–braid of Garside normal form (C) or (D). However, there are examples

[2] These computations were generalized to all links that are closures of 3–braids in [41].

where the upper bound in (31) is sharp. The two easiest such examples in terms of crossing number are the nonalternating knots 8_{20} and 8_{21}, which are represented by the 3–braids

$$a^3b^{-1}a^{-3}b^{-1} \sim \Delta^{-3}a^7, \quad a^3ba^{-2}b^2 \sim \Delta^{-2}a^3b^2a^2b^3,$$

respectively; see [36]. The lower bound on the alternation number from (31) is

$$|r+\ell|-1=0$$

in both cases. Indeed, by [7, Theorem 8.6] both knots are quasialternating, so all the invariants from (26) are equal [7, Proposition 1.4; 40; 46].

Remark 5.10 In a similar fashion as Corollary 5.7, the Turaev genus of all 3–braid knots was determined up to an additive error of at most 1 by Lowrance using his computation of the Khovanov width for these knots [37, Proposition 4.15]. More precisely,

$$|\ell|-1 \leq g_T(K) \leq |\ell| \quad \text{if } \ell \neq 0$$

for any knot K that is represented by $\gamma = \Delta^{2\ell}a^{-p_1}b^{q_1}\cdots a^{-p_r}b^{q_r}$ for some $\ell \in \mathbb{Z}$, $r \geq 1$ and $p_i, q_i \geq 1$ for $i \in \{1, \ldots, r\}$.

6 The fractional Dehn twist coefficient of 3–braids in Garside normal form

In this section, we compute the fractional Dehn twist coefficient of any 3–braid in Garside normal form (see Definition 3.4).

The *fractional Dehn twist coefficient* is a homogeneous quasimorphism on the braid group B_n that assigns to any n–braid γ a rational number $\omega(\gamma)$. Here, a *quasimorphism* on a group G is any map $\varphi\colon G \to \mathbb{R}$ such that

$$\sup_{(a,b)\in G\times G} |\varphi(ab)-\varphi(a)-\varphi(b)| =: D_\varphi < \infty,$$

where D_φ is called the *defect* of φ. A quasimorphism $\varphi\colon G \to \mathbb{R}$ is called *homogeneous* if $\varphi(a^k) = k\varphi(a)$ for all $k \in \mathbb{Z}$ and $a \in G$. Any homogeneous quasimorphism is invariant under conjugation, so $\omega(\gamma)$ is invariant under the conjugacy class of γ.

The fractional Dehn twist coefficient first appeared in [24] in a different language. It can be defined for mapping classes of general surfaces with boundary, where we here view braids as mapping classes of the n times punctured closed disk. Malyutin defined the fractional Dehn twist coefficient $\omega\colon B_n \to \mathbb{R}$, with $n \geq 2$, for all braid groups and showed that its defect is 1 if $n \geq 3$ and 0 if $n = 2$ [39, Theorem 6.3]. We refer the reader to [39] for a more detailed account.

Corollary 6.1 *Let γ be a 3–braid. Then its fractional Dehn twist coefficient is*

$$\omega(\gamma) = \begin{cases} \ell & \text{if } \gamma \text{ is conjugate to a braid in (A)}, \\ \frac{1}{6}(p+1)+\ell & \text{if } \gamma \text{ is conjugate to a braid in (B)}, \\ r+\ell & \text{if } \gamma \text{ is conjugate to a braid in (C) or (D)}, \end{cases}$$

where (A)–(D) refer to the Garside normal forms from Proposition 3.2.

Remark 6.2 The fractional Dehn twist coefficient was computed for 3–braids in Murasugi normal form (see Definition 4.15) in [29, Proposition 6.6].

In the proof of Corollary 6.1, we will use that the fractional Dehn twist coefficient of any 3–braid γ is completely determined by the writhe $\mathrm{wr}(\gamma)$ and the *homogenized upsilon invariant* $\tilde{\upsilon}$ of γ: we have, by [19, Theorem 1.3],

$$\omega(\gamma) = \tilde{\upsilon}(\gamma) + \tfrac{1}{2}\,\mathrm{wr}(\gamma) \tag{32}$$

for any 3–braid γ. The invariant $\tilde{\upsilon}$ is another real-valued homogeneous quasimorphism on the braid group B_3 which can be defined as

$$\tilde{\upsilon}\colon B_3 \to \mathbb{R}, \quad \gamma \mapsto \tilde{\upsilon}(\gamma) = \lim_{k\to\infty} \frac{\upsilon(\widehat{\gamma^{6k}ab})}{6k}.$$

More generally, Brandenbursky [12, Theorem 2.6] showed that a homogeneous quasimorphism $B_n \to \mathbb{R}$ can be assigned to any concordance homomorphism $\mathcal{C} \to \mathbb{R}$ that is bounded above by a constant multiple of the 4–genus. We refer the reader to [12] or [19, Appendix A] for more details on homogenized concordance invariants.

Proposition 6.3 *Let γ be a 3–braid. Then*

$$\tilde{\upsilon}(\gamma) = \begin{cases} -\frac{1}{2}p - 2\ell & \text{if } \gamma \text{ is conjugate to a braid in (A)}, \\ -\frac{1}{3}(p+1) - 2\ell & \text{if } \gamma \text{ is conjugate to a braid in (B)}, \\ -\frac{1}{2}\big(\sum_{i=1}^{r}(p_i+q_i)\big) + r - 2\ell & \text{if } \gamma \text{ is conjugate to a braid in (C)}, \\ -\frac{1}{2}\big(\sum_{i=1}^{r-1}(p_i+q_i)+p_r\big) + r - 2\ell - \frac{3}{2} & \text{if } \gamma \text{ is conjugate to a braid in (D)}. \end{cases}$$

Proof We will use that $\tilde{\upsilon}(\alpha\beta) = \tilde{\upsilon}(\alpha) + \tilde{\upsilon}(\beta)$ if α and β commute [19, Lemma A.1]. In particular, for any 3–braid γ and any $\ell \in \mathbb{Z}$,

$$\tilde{\upsilon}(\Delta^{2\ell}\gamma) = \tilde{\upsilon}(\Delta^{2\ell}) + \tilde{\upsilon}(\gamma). \tag{33}$$

Moreover, by the definition of $\tilde{\upsilon}$, equation (12) and the homogeneity of $\tilde{\upsilon}$,

$$\tilde{\upsilon}(\Delta^{2\ell}) = -2\ell \quad \text{for all } \ell \in \mathbb{Z}. \tag{34}$$

We will now compute $\tilde{\upsilon}(\gamma)$ for the positive 3–braids γ of the form (A)–(D), ie assuming $\ell \geq 0$ in (A)–(D). The statement of Proposition 6.3 will then follow from (33) and (34).

First, let $\gamma = \Delta^{2\ell} a^p$ for some $\ell \geq 0$ and $p \geq 0$. If $p = 0$, we have $\tilde{\upsilon}(\gamma) = -2\ell$ by (34). If $p \geq 1$, we have

$$\gamma^{6k} ab = \Delta^{12\ell k} a^{6pk} ab \sim \Delta^{12\ell k+1} a^{6pk-1},$$

so by Lemma 4.9, for $k \geq 1$,

$$\begin{aligned} \upsilon(\widehat{\gamma^{6k} ab}) &= -\tfrac{1}{2}(6pk-1) + 1 - 12\ell k - \tfrac{3}{2} = -3pk - 12\ell k, \\ \tilde{\upsilon}(\gamma) &= \lim_{k\to\infty} \frac{\upsilon(\widehat{\gamma^{6k} ab})}{6k} = \lim_{k\to\infty} \frac{-3pk - 12\ell k}{6k} = -\frac{p}{2} - 2\ell. \end{aligned}$$

Second, let $\gamma = \Delta^{2\ell} a^p b$ for some $\ell \geq 0$ and $p \in \{1, 2, 3\}$. We have

$$\begin{aligned} \gamma^{6k} ab &= \Delta^{12\ell k} (ab)^{6k} ab = \Delta^{12\ell k + 4k} ab && \text{if } p = 1, \\ \gamma^{6k} ab &= \Delta^{12\ell k} (a^2 b a^2 b)^{3k} ab = \Delta^{12\ell k} (ababab)^{3k} ab = \Delta^{12\ell k+6k} ab && \text{if } p = 2, \\ \gamma^{6k} ab &= \Delta^{12\ell k} (a^3 b a^3 b a^3 b)^{2k} ab = \Delta^{12\ell k} (a^2 bababa a^2 b)^{2k} ab \\ &= \Delta^{12\ell k + 8k} ab && \text{if } p = 3. \end{aligned}$$

By (12),

$$\tilde{\upsilon}(\gamma) = \lim_{k\to\infty} \frac{-12\ell k - (2p+2)k}{6k} = -2\ell - \frac{p+1}{3}.$$

Third, let $\gamma = \Delta^{2\ell} a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r}$ for some $\ell \geq 0$, $r \geq 1$ and $p_i, q_i \geq 2$ for $i \in \{1, \dots, r\}$. Then

$$\begin{aligned} \gamma^{6k} ab &= \Delta^{12\ell k} (a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r})^{6k} ab \\ &\sim \Delta^{12\ell k+1} a^{p_1 - 1} b^{q_1} \cdots a^{p_r} b^{q_r} (a^{p_1} b^{q_1} \cdots a^{p_r} b^{q_r})^{6k-1} \\ &\sim \Delta^{12\ell k+1} (b^{q_1} a^{p_2} b^{q_2} \cdots a^{p_r} b^{q_r} a^{p_1})^{6k-1} b^{q_1} a^{p_2} b^{q_2} \cdots a^{p_r} b^{p_1 + q_r - 1}, \end{aligned}$$

where $p_1 + q_r - 1 \geq 3$. By Lemma 4.9,

$$\begin{aligned} \upsilon(\widehat{\gamma^{6k} ab}) &= -3k \sum_{i=1}^{r} (p_i + q_i) + 6kr - 12\ell k - 1, \\ \tilde{\upsilon}(\gamma) &= -\frac{1}{2} \sum_{i=1}^{r} (p_i + q_i) + r - 2\ell. \end{aligned}$$

Finally, let $\gamma = \Delta^{2\ell+1} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r}$ for some $\ell \geq 0$, $r \geq 1$, $p_r \geq 2$, and $p_i, q_i \geq 2$ for $i \in \{1, \dots, r-1\}$. Then

$$\begin{aligned}
\gamma^{6k} ab &= \Delta^{12\ell k} (\Delta a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r})^{6k} ab \\
&= \Delta^{12\ell k} (\Delta^2 b^{p_1} a^{q_1} \cdots b^{p_{r-1}} a^{q_{r-1}} b^{p_r} a^{p_1} b^{q_1} \cdots a^{p_{r-1}} b^{q_{r-1}} a^{p_r})^{3k} ab \\
&= \Delta^{12\ell k+6k} (b^{p_1} \cdots b^{p_r} a^{p_1} \cdots a^{p_r})^{3k} ab \\
&\sim \Delta^{12\ell k+6k} a^{q_1} b^{p_2} \cdots b^{p_r} a^{p_1} \cdots \\
&\qquad\qquad \cdots a^{p_r} (b^{p_1} \cdots b^{p_r} a^{p_1} \cdots a^{p_r})^{3k-2} b^{p_1} \cdots b^{p_r} a^{p_1} \cdots a^{p_r+1} b^{p_1+1},
\end{aligned}$$

where $p_r + 1, p_1 + 1 \geq 3$. By Lemma 4.10,

$$\upsilon(\widehat{\gamma^{6k} ab}) = -3k \left(\sum_{i=1}^{r-1} (p_i + q_i) + p_r \right) + 6kr - 12\ell k - 9k - 1,$$

$$\tilde{\upsilon}(\gamma) = -\frac{1}{2} \left(\sum_{i=1}^{r-1} (p_i + q_i) + p_r \right) + r - 2\ell - \frac{3}{2}. \qquad \square$$

Proof of Corollary 6.1 This follows directly from Proposition 6.3, (32), and a straightforward calculation of the writhe of the braids in (A)–(D). □

Remark 6.4 If γ is a 3–braid conjugate to a braid of the form in (C) or (D) such that $\hat{\gamma}$ is a knot, then Proposition 6.3 and Theorem 1.1 imply $\tilde{\upsilon}(\gamma) = \upsilon(\hat{\gamma})$. If γ additionally is a positive 3–braid, then $\omega(\gamma) = r + \ell = g(\hat{\gamma}) + \upsilon(\hat{\gamma}) + 1$ is the minimal number from Corollary 1.3 (ie Corollary 4.12).

Remark 6.5 Our computation of $\omega(\gamma)$ in Corollary 6.1 together with [19, Theorem 1.3] completely determines $\widetilde{\Upsilon(t)}(\gamma)$ for all $0 \leq t \leq 1$ for any 3–braid γ, where $\widetilde{\Upsilon(t)}(\gamma)$ is the homogenization of the invariant $\Upsilon(t)\colon \mathcal{C} \to \mathbb{R}$, defined similarly as the homogenization $\tilde{\upsilon}$ of υ.

References

[1] **T Abe**, *An estimation of the alternation number of a torus knot*, J. Knot Theory Ramifications 18 (2009) 363–379 MR Zbl

[2] **T Abe**, **K Kishimoto**, *The dealternating number and the alternation number of a closed 3–braid*, J. Knot Theory Ramifications 19 (2010) 1157–1181 MR Zbl

[3] **C C Adams**, **J F Brock**, **J Bugbee**, **T D Comar**, **K A Faigin**, **A M Huston**, **A M Joseph**, **D Pesikoff**, *Almost alternating links*, Topology Appl. 46 (1992) 151–165 MR Zbl

[4] **J W I Alexander**, *A lemma on systems of knotted curves*, Proc. Natl. Acad. Sci. USA 9 (1923) 93–95 Zbl

[5] **E Artin**, *Theorie der Zöpfe*, Abh. Math. Sem. Univ. Hamburg 4 (1925) 47–72 MR Zbl

[6] **S Baader**, *Positive braids of maximal signature*, Enseign. Math. 59 (2013) 351–358 MR Zbl

[7] **J A Baldwin**, *Heegaard Floer homology and genus one, one-boundary component open books*, J. Topol. 1 (2008) 963–992 MR Zbl

[8] **W Ballinger**, *Concordance invariants from the $E(-1)$ spectral sequence on Khovanov homology*, preprint (2020) arXiv 2004.10807

[9] **D Bennequin**, *Entrelacements et équations de Pfaff*, from "Third Schnepfenried geometry conference, I", Astérisque 107–108, Soc. Math. France, Paris (1983) 87–161 MR Zbl

[10] **J S Birman**, **T E Brendle**, *Braids: A survey*, from "Handbook of knot theory" (W Menasco, M Thistlethwaite, editors), Elsevier, Amsterdam (2005) 19–103 MR Zbl

[11] **J S Birman**, **W W Menasco**, *Studying links via closed braids, III: Classifying links which are closed 3–braids*, Pacific J. Math. 161 (1993) 25–113 MR Zbl

[12] **M Brandenbursky**, *On quasi-morphisms from knot and braid invariants*, J. Knot Theory Ramifications 20 (2011) 1397–1417 MR Zbl

[13] **A Champanerkar**, **I Kofman**, *A survey on the Turaev genus of knots*, Acta Math. Vietnam. 39 (2014) 497–514 MR Zbl

[14] **W-L Chow**, *On the algebraical braid group*, Ann. of Math. 49 (1948) 654–658 MR Zbl

[15] **O T Dasbach**, **D Futer**, **E Kalfagianni**, **X-S Lin**, **N W Stoltzfus**, *The Jones polynomial and graphs on surfaces*, J. Combin. Theory Ser. B 98 (2008) 384–399 MR Zbl

[16] **O T Dasbach**, **A M Lowrance**, *Turaev genus, knot signature, and the knot homology concordance invariants*, Proc. Amer. Math. Soc. 139 (2011) 2631–2645 MR Zbl

[17] **D Erle**, *Calculation of the signature of a 3–braid link*, Kobe J. Math. 16 (1999) 161–175 MR Zbl

[18] **P Feller**, *Optimal cobordisms between torus knots*, Comm. Anal. Geom. 24 (2016) 993–1025 MR Zbl

[19] **P Feller**, **D Hubbard**, *Braids with as many full twists as strands realize the braid index*, J. Topol. 12 (2019) 1069–1092 MR Zbl

[20] **P Feller**, **D Krcatovich**, *On cobordisms between knots, braid index, and the upsilon-invariant*, Math. Ann. 369 (2017) 301–329 MR Zbl

[21] **P Feller**, **L Lewark**, **A Lobb**, *Squeezed knots*, preprint (2022) arXiv 2202.12289

[22] **P Feller**, **S Pohlmann**, **R Zentner**, *Alternation numbers of torus knots with small braid index*, Indiana Univ. Math. J. 67 (2018) 645–655 MR Zbl

[23] **S Friedl**, **C Livingston**, **R Zentner**, *Knot concordances and alternating knots*, Michigan Math. J. 66 (2017) 421–432 MR Zbl

[24] **D Gabai**, **U Oertel**, *Essential laminations in 3–manifolds*, Ann. of Math. 130 (1989) 41–73 MR Zbl

[25] **F A Garside**, *The braid group and other groups*, Quart. J. Math. Oxford Ser. 20 (1969) 235–254 MR Zbl

[26] **R E Goodrick**, *Two bridge knots are alternating knots*, Pacific J. Math. 40 (1972) 561–564 MR Zbl

[27] **J E Greene**, *Donaldson's theorem, Heegaard Floer homology, and knots with unknotting number one*, Adv. Math. 255 (2014) 672–705 MR Zbl

[28] **J Hom**, *A survey on Heegaard Floer homology and concordance*, J. Knot Theory Ramifications 26 (2017) art. id. 1740015 MR Zbl

[29] **D Hubbard**, **K Kawamuro**, **F C Kose**, **G Martin**, **O Plamenevskaya**, **K Raoux**, **L Truong**, **H Turner**, *Braids, fibered knots, and concordance questions*, from "Research directions in symplectic and contact geometry and topology" (B Acu, C Cannizzo, D McDuff, Z Myer, Y Pan, L Traynor, editors), Assoc. Women Math. Ser. 27, Springer (2021) 293–324 MR Zbl

[30] **T Kanenobu**, *Upper bound for the alternation number of a torus knot*, Topology Appl. 157 (2010) 302–318 MR Zbl

[31] **A Kawauchi**, *On alternation numbers of links*, Topology Appl. 157 (2010) 274–279 MR Zbl

[32] **P B Kronheimer**, **T S Mrowka**, *Gauge theory for embedded surfaces, I*, Topology 32 (1993) 773–826 MR Zbl

[33] **L Lewark**, *Rasmussen's spectral sequences and the $\mathfrak{sl}_N$–concordance invariants*, Adv. Math. 260 (2014) 59–83 MR Zbl

[34] **C Livingston**, *Computations of the Ozsváth–Szabó knot concordance invariant*, Geom. Topol. 8 (2004) 735–742 MR Zbl

[35] **C Livingston**, *Notes on the knot concordance invariant upsilon*, Algebr. Geom. Topol. 17 (2017) 111–130 MR Zbl

[36] **C Livingston**, **A H Moore**, *KnotInfo: table of knot invariants*, online resource (2021) Available at `http://knotinfo.math.indiana.edu`

[37] **A Lowrance**, *The Khovanov width of twisted links and closed 3–braids*, Comment. Math. Helv. 86 (2011) 675–706 MR Zbl

[38] **A M Lowrance**, *Alternating distances of knots and links*, Topology Appl. 182 (2015) 53–70 MR Zbl

[39] **A V Malyutin**, *Writhe of (closed) braids*, Algebra i Analiz 16 (2004) 59–91 MR Zbl In Russian; translated in St. Petersburg Math. J. 16 (2005) 791–813

[40] **C Manolescu**, **P Ozsváth**, *On the Khovanov and knot Floer homologies of quasi-alternating links*, from "Proceedings of the 14th Gökova geometry-topology conference" (S Akbulut, T Önder, R J Stern, editors), GGT, Gökova (2008) 60–81 MR Zbl

[41] **G Martin**, *Annular Rasmussen invariants: Properties and 3–braid classification*, preprint (2019) arXiv 1909.09245

[42] **H R Morton**, *Closed braids which are not prime knots*, Math. Proc. Cambridge Philos. Soc. 86 (1979) 421–426 MR Zbl

[43] **H Murakami**, *Some metrics on classical knots*, Math. Ann. 270 (1985) 35–45 MR Zbl

[44] **K Murasugi**, *On a certain numerical invariant of link types*, Trans. Amer. Math. Soc. 117 (1965) 387–422 MR Zbl

[45] **K Murasugi**, *On closed 3–braids*, Mem. Amer. Math. Soc. 151, Amer. Math. Soc., Providence, RI (1974) MR Zbl

[46] **P S Ozsváth**, **A I Stipsicz**, **Z Szabó**, *Concordance homomorphisms from knot Floer homology*, Adv. Math. 315 (2017) 366–426 MR Zbl

[47] **P S Ozsváth**, **A I Stipsicz**, **Z Szabó**, *Unoriented knot Floer homology and the unoriented four-ball genus*, Int. Math. Res. Not. 2017 (2017) 5137–5181 MR Zbl

[48] **P Ozsváth**, **Z Szabó**, *Knot Floer homology and the four-ball genus*, Geom. Topol. 7 (2003) 615–639 MR Zbl

[49] **J A Rasmussen**, *Floer homology and knot complements*, PhD thesis, Harvard University (2003) MR arXiv math/0306378

[50] **J Rasmussen**, *Khovanov homology and the slice genus*, Invent. Math. 182 (2010) 419–447 MR Zbl

[51] **L Rudolph**, *Quasipositivity as an obstruction to sliceness*, Bull. Amer. Math. Soc. 29 (1993) 51–59 MR Zbl

[52] **O Schreier**, *Über die gruppen $A^a B^b = 1$*, Abh. Math. Sem. Univ. Hamburg 3 (1924) 167–169 MR Zbl

[53] **T Shibuya**, *Some relations among various numerical invariants for links*, Osaka Math. J. 11 (1974) 313–322 MR Zbl

[54] **H F Trotter**, *Homology of group systems with applications to knot theory*, Ann. of Math. 76 (1962) 464–498 MR Zbl

Department of Mathematics, ETH Zurich
Zurich, Switzerland

paulagtruoel@gmail.com

Received: 20 September 2021 Revised: 22 March 2022

Algebraic & Geometric Topology **23**:8 (2023) 3805–3834
DOI: 10.2140/agt.2023.23.3805
Published: 5 November 2023

Cusps and commensurability classes of hyperbolic 4–manifolds

CONNOR SELL

There are six orientable compact flat 3–manifolds that can occur as cusp cross-sections of hyperbolic 4–manifolds. We provide criteria for exactly when a given commensurability class of arithmetic hyperbolic 4–manifolds contains a representative with a given cusp type. In particular, for three of the six cusp types, we provide infinitely many examples of commensurability classes that contain no manifolds with cusps of the given type; no such examples were previously known for any cusp type.

57M50; 11E20, 11F06, 16H05, 57K50

1 Introduction

Let $M = \mathbb{H}^n/\,\Gamma$ be a finite-volume noncompact hyperbolic n–manifold. A cusp of M is homeomorphic to $B \times \mathbb{R}^+$, where B is a compact flat $(n-1)$–manifold. If M is orientable, then B must be orientable. In [13], Long and Reid proved that every compact flat $(n-1)$–manifold, up to homeomorphism, must occur as a cusp cross-section of a hyperbolic n–orbifold; this result was upgraded from n–orbifolds to n–manifolds by McReynolds in [15]. Long and Reid [13] give a constructive algorithm which, given a compact flat $(n-1)$–manifold, outputs an arithmetic hyperbolic n–orbifold with a cusp with the specified cross-section. We discuss this algorithm in more detail in Section 5.

For ease of notation, we may refer to a cusp with cross-section B as a *cusp of type* B, as the cross-section of a cusp determines its homeomorphism class. We may also refer to a homeomorphism class of cusps, or "cusp type", by its cross-section. See Section 4 for a description of the six possible cusp types for hyperbolic 4–manifolds, and the names used below.

The above results tell us that each compact flat $(n-1)$–manifold occurs as a cusp of some hyperbolic n–manifold, but little is known about which conditions give rise to each cusp type. To investigate the occurrence of cusp types further, it makes sense to look at compact flat 3–manifolds in finite-volume hyperbolic 4–manifolds, as this is the lowest dimension in which multiple orientable cusp types can occur. It is well known that the 3–torus occurs as a cusp in every commensurability class of cusped hyperbolic 4–manifolds. Indeed, in every commensurability class of cusped hyperbolic 4–manifolds, manifolds where all cusp types are the 3–torus occur; see McReynolds, Reid, and Stover [16]. A striking result by Kolpakov and Martelli [10] showed that there exist one-cusped hyperbolic 4–manifolds having cusp type the 3–torus. Furthermore, Kolpakov and Slavich [11] showed that the $\frac{1}{2}$–twist also occurs as the cusp type of a one-cusped hyperbolic 4–manifold. On the other hand, the $\frac{1}{3}$–twist and $\frac{1}{6}$–twist have been obstructed from occurring as cusps of one-cusped manifolds; see Long and Reid [12]. Although it is as yet unknown whether the Hantzsche–Wendt manifold occurs as a cusp type of a one-cusped hyperbolic 4–manifold, it was shown by Ferrari, Kolpakov, and Slavich [9] that there exists a finite-volume hyperbolic 4–manifold where all cusp types are the Hantzsche–Wendt manifold. We also note that the isometry classes within each homeomorphism class that occur geometrically as cusps of hyperbolic 4–manifolds are dense in the moduli space of any compact flat 3–manifold; see Nimershiem [20].

We provide the first known examples of commensurability classes that avoid three cusp types. In fact, we provide infinitely many such examples, obtaining the result below. Furthermore, given any commensurability class C of cusped arithmetic hyperbolic 4–manifolds and any cusp type B, we give conditions on when C contains a manifold with a cusp of type B in Theorem 5.1. Notably, three cusp types occur in every such class. We refer to Section 2 for terminology used in Theorems 1.1 and 1.2.

Theorem 1.1 *Every commensurability class of arithmetic hyperbolic 4–manifolds contains manifolds with the 3–torus, the $\frac{1}{2}$–twist, and the Hantzsche–Wendt manifold as cusp types. There exist infinitely many commensurability classes C of hyperbolic 4–manifolds such that no manifold in C has a cusp of type $\frac{1}{3}$–twist. The same holds for cusps of type $\frac{1}{4}$–twist and $\frac{1}{6}$–twist.*

Additionally, we can use "inbreeding" of arithmetic hyperbolic 4–manifolds (see Agol [1]) to construct some nonarithmetic manifolds that avoid some cusp types, up to commensurability.

Theorem 1.2 *There exist infinitely many commensurability classes of finite-volume cusped nonarithmetic hyperbolic 4–manifolds that avoid each of the following cusp types: the $\frac{1}{3}$–twist, the $\frac{1}{4}$–twist, and the $\frac{1}{6}$–twist.*

We briefly review the organization of the paper. In Sections 2, 3, and 4, we provide preliminary information about quadratic forms, quaternion algebras, arithmetic hyperbolic manifolds, and the six orientable compact flat 3–manifolds that are the cusp types of orientable hyperbolic 4–manifolds. In Sections 5 and 6, we prove Theorem 1.1 and generalize it to give complete conditions on when a given commensurability class contains a manifold with a cusp of given type. In Section 7, we use this result to show that there are some commensurability classes of hyperbolic 5–manifolds that avoid some compact flat 4–manifold cusp types, and explain why we can't make the same argument in higher dimensions. In Section 8, we show that there are commensurability classes of nonarithmetic hyperbolic manifolds in both 4 and 5 dimensions that avoid certain cusp types as well, proving Theorem 1.2.

Acknowledgements The author wishes to thank his PhD advisor Alan Reid for his guidance and useful discussions. This paper is partially supported by NSF grant DMS-1745670.

2 Quadratic forms and quaternion algebras

2.1 Quadratic forms

Definition 2.1 (quadratic form) A *quadratic form* over a field K is a homogeneous polynomial of degree 2 with coefficients in K.

A quadratic form $q(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j$ in n variables is said to have *rank* n, and can be written as an $n \times n$ symmetric matrix Q such that $q(x) = x^t Q x$. This can be accomplished by defining the entries by $Q_{ii} = a_{ii}$ and $Q_{ij} = \frac{1}{2} a_{ij}$ when $i \neq j$.

For any quadratic form q of rank n and ring R, we can define the orthogonal group $O(q, R)$ to be the group of all invertible $n \times n$ matrices A with entries in R such that $q(x) = q(Ax)$ for any $x \in R^n$. We can similarly define the special orthogonal group $\mathrm{SO}(q, R)$ to be the subgroup of $O(q, R)$ of matrices with determinant 1. Note that $\mathrm{SO}(q, \mathbb{R})$ is a Lie group, and thus has an identity component $\mathrm{SO}_0(q, \mathbb{R})$. Then, for any subring $R \subset \mathbb{R}$, we define $\mathrm{SO}_0(q, R) = \mathrm{SO}_0(q, \mathbb{R}) \cap \mathrm{SO}(q, R)$. Our focus is quadratic forms over $\mathbb{Q}$ and the corresponding groups $\mathrm{SO}_0(q, \mathbb{Z})$.

Definition 2.2 (rational equivalence) Quadratic forms given by symmetric matrices $Q_1, Q_2 \in \mathrm{GL}(n, \mathbb{Q})$ are *rationally equivalent* (or *equivalent over* $\mathbb{Q}$) if there exists $T \in \mathrm{GL}(n, \mathbb{Q})$ such that $T^t Q_1 T = Q_2$.

All quadratic forms over $\mathbb{Q}$ are rationally equivalent to a *diagonal* quadratic form, by which we mean a quadratic form whose corresponding matrix is diagonal. Thus, when working with a rational equivalence class of quadratic forms, we will always choose a diagonal representative. For ease of notation, we will denote diagonal quadratic forms $q(x) = \sum_{i=1}^{n} a_i x_i^2$ by writing their coefficients $\langle a_1, \ldots, a_n \rangle$. Here all quadratic forms will be nondegenerate; that is, $a_i \neq 0$ for all i.

There is another relevant notion of equivalence, which is closely related to rational equivalence [18]:

Definition 2.3 (projective equivalence) Quadratic forms q_1 and q_2 are *projectively equivalent over* $\mathbb{Q}$, or just "projectively equivalent", if there are nonzero integers a and b such that aq_1 and bq_2 are rationally equivalent.

Let q_1 and q_2 be quadratic forms of odd rank with the same signature and discriminant. We can check for projective equivalence by scaling q_1 and q_2 so they have the same discriminant, and then checking for rational equivalence.

A complete set of invariants for diagonal quadratic forms up to rational equivalence is given by the signature, discriminant, and the Hasse–Witt invariants over all primes p. A quadratic form $q = \langle a_1, \ldots, a_n \rangle$ of *signature* (a, b) has a positive coefficients and b negative coefficients. The *discriminant* $d \in \mathbb{Q}/(\mathbb{Q}^\times)^2$ is given by $d = \prod_{i=1}^{n} a_i$; note that it is defined only up to multiplication by squares. The Hasse–Witt invariants are a little harder to define, and contain the bulk of the number-theoretic information. For integers a and b and prime p, we first define the Hilbert symbol

$$(a,b)_p = \begin{cases} 1 & \text{if } z^2 = ax^2 + by^2 \text{ has a solution in } \mathbb{Q}_p, \\ -1 & \text{otherwise.} \end{cases}$$

Here $\mathbb{Q}_p$ denotes the p–adic field at p, or $\mathbb{R}$ if $p = \infty$.

Definition 2.4 (Hasse–Witt invariant) For a diagonal quadratic form $q = \langle a_1, \ldots, a_n \rangle$ over $\mathbb{Q}$ and a prime p, possibly ∞, the *Hasse–Witt invariant of q at p* is given by

$$\epsilon_p(q) = \prod_{1 \le i < j \le n} (a_i, a_j)_p.$$

Every Hasse–Witt invariant must have value 1 or -1. There is a closed-form equation that allows us to easily compute a Hilbert symbol; thus, a Hasse–Witt invariant is easy to compute as well. Let $a = p^\alpha u$ and $b = p^\beta v$ with u and v both relatively prime to p in $\mathbb{Z}$. Then for $p > 2$

$$(a,b)_p = (-1)^{\alpha\beta\tau(p)}\left(\frac{u}{p}\right)^\beta\left(\frac{v}{p}\right)^\alpha,$$

and, for $p = 2$,

$$(a,b)_p = (-1)^{\tau(u)\tau(v)+\alpha\omega(v)+\beta\omega(u)}.$$

Here we use the Legendre symbol and the functions $\tau(x) = \frac{1}{2}(x-1)$ and $\omega(x) = \frac{1}{8}(x^2-1)$, both of which only need to be defined modulo 2 [25, Chapter III, Theorem 1].

We can see from these equations that $(a,b)_p$ can only be -1 if either a or b is divisible by p an odd number of times. This means that $\epsilon_p(q) = 1$ for all primes p that don't occur as a factor of a coefficient of q. In particular, for any given quadratic form q, $\epsilon_p(q) = 1$ for all but finitely many values of p.

Additionally, Hilbert's reciprocity law states that the Hilbert symbols satisfy the identity $\prod_p (a,b)_p = 1$, where the product is taken over all places p of $\mathbb{Q}$, including $p = \infty$ [25, Chapter III, Theorem 3]. From this, we deduce the identity $\prod_p \epsilon_p(q) = 1$ for any quadratic form q. Since $(a,b)_\infty$ depends on the existence of a nonzero solution to $z^2 = ax^2 + by^2$ over the field $\mathbb{Q}_\infty = \mathbb{R}$, we know $(a,b)_\infty = -1$ if and only if both a and b are negative. We'll be working mostly with quadratic forms of signature $(4,1)$, so in this case $\prod_{1\le i<j\le n}(a_i,a_j)_\infty = 1$, as no pair (a_i,a_j) of distinct coefficients are both negative. As a result, the identity $\prod_p \epsilon_p(q) = 1$ holds when we consider only finite places p for quadratic forms of signature $(4,1)$.

2.2 Quaternion algebras

Definition 2.5 (quaternion algebra) A *quaternion algebra* over a field F with $\mathrm{char}(F) \neq 2$ is an algebra consisting of elements $w+xi+yj+zij$, with $w,x,y,z \in F$, equipped with relations $i^2 = a$, $j^2 = b$, and $ij = -ji$ for some fixed $a,b \in F$. We write this as $((a,b)/F)$.

Alternatively, a quaternion algebra Q over F is any central simple algebra of dimension 4 over F. Every such $Q = ((a,b)/F)$ has a *norm form*, given by $N(w+xi+yj+zij) = w^2 - ax^2 - by^2 + abz^2$, which is compatible with multiplication in Q.

The *pure quaternions* Q_0 of Q are the elements $w + xi + yj + zij$ with $w = 0$. Restricted to the pure quaternions, the norm form of Q (or, for short, the norm form of Q_0) becomes $N(xi + yj + zij) = -ax^2 - by^2 + abz^2$. Note that any quadratic form of rank 3 and discriminant 1 is rationally equivalent to such a form. To see this, observe that if $\langle a, b, c\rangle$ has discriminant 1 then $c = ab$ up to multiplication by a square. In particular, the quadratic form $\langle a, b, ab\rangle$ coincides with the norm form of $((-a, -b)/\mathbb{Q})_0$. We will make use of quadratic forms of signature $(4, 1)$ that are the direct sum of a positive definite norm form of some Q_0 and $\langle 1, -1\rangle$.

Definition 2.6 (quaternion type) A *quadratic form of quaternion type* is a quadratic form $q = \langle a, b, ab, 1, -1\rangle$ for some positive $a, b \in \mathbb{Z}$.

Lemma 2.7 *Every quadratic form q over $\mathbb{Q}$ of signature $(4, 1)$ is projectively equivalent to a quadratic form q' of quaternion type.*

In order to prove this lemma we'll need to use Conway's p–excesses, as described in [7, Chapter 15]. These will not appear in the rest of the paper, so readers not interested in the proof of this lemma may ignore these definitions.

Definition 2.8 (p–excess of rank-1 quadratic form) Let $p \neq 2$ be a prime, possibly ∞, and let $q = \langle a\rangle$ be a rank-1 quadratic form such that $a = p^k u$ with u relatively prime to p. If $p = \infty$, then let p^k be the sign of a and u its magnitude. Then we define the p–excess of q to be

$$e_p(q) \equiv \begin{cases} p^k + 3 \pmod 8 & \text{if } k \text{ is odd and } u \text{ is a quadratic nonresidue modulo p,} \\ p^k - 1 \pmod 8 & \text{otherwise.} \end{cases}$$

If $p = 2$, then

$$e_p(q) \equiv \begin{cases} -u - 3 \pmod 8 & \text{if k is odd and } u \equiv 3, 5 \pmod 8, \\ -u + 1 \pmod 8 & \text{otherwise.} \end{cases}$$

Definition 2.9 (p–excess of arbitrary quadratic form) Let p be a prime, possibly ∞, and let $q = \langle a_1, \dots, a_n\rangle$ be a diagonal quadratic form. Then we define the p–excess of q to be

$$e_p(q) \equiv \sum_{i=1}^{n} e_p(\langle a_i\rangle) \pmod 8.$$

The most notable properties of the p–excesses are that they are additive under direct sum of quadratic forms, and that they are invariant under rational equivalence. In fact, p–excesses are part of a complete invariant of quadratic forms up to rational equivalence, together with the signature and, in the case of forms of even rank, the discriminant

[7, Section 15.5.1, Theorem 3]. We can also extract the Hasse–Witt invariants of a quadratic form q from the discriminant d and p–excesses $e_p(q)$ [7, Section 15.5.3]:

$$\epsilon_p(q) = \begin{cases} 1 & \text{if } e_p(q) = e_p(\langle d, 1, \dots, 1\rangle), \\ -1 & \text{otherwise.} \end{cases}$$

To prove Lemma 2.7, we'll use the additivity of e_p to construct a rank-3 form q_3 of discriminant 1 such that $q_3 \oplus \langle 1, -1\rangle$ has certain desired Hasse–Witt invariants. We'll also use the following lemma, which can be found in greater generality in [25, Chapter IV, Proposition 7].

Lemma 2.10 *Let d, r, s, and n be integers, and ϵ_p be 1 or -1 for each prime p, including ∞. Then there exists a rank-n quadratic form q of discriminant d, signature (r, s), and Hasse–Witt invariants ϵ_p if and only if the following conditions are satisfied:*

(1) *$\epsilon_p = 1$ for almost all p and $\prod \epsilon_p = 1$ over all primes p.*

(2) *$\epsilon_p = 1$ if $n = 1$, or if $n = 2$ and the image of d in $\mathbb{Q}_p^*/(\mathbb{Q}_p^*)^2$ is -1.*

(3) *$r, s \geq 0$ and $n = r + s$.*

(4) *The sign of d is equal to $(-1)^s$.*

(5) *$\epsilon_\infty = (-1)^{s(s-1)/2}$.*

Proof of Lemma 2.7 We can scale q to ensure it has discriminant -1 by multiplying the entire form by $-d$, where d is its discriminant. This will multiply the product of the terms by $-d^5$, and thus we'll obtain the new discriminant $-d^6 \equiv -1$. Note that scaling a form does not change its projective equivalence class.

Now, compute the p–excesses $e_p(q)$ and set $e'_p = e_p(q) - e_p(\langle 1, -1\rangle)$. By definition $e_p(\langle a_1, a_2, \dots, a_n\rangle) = \sum_{i=1}^n e_p(a_i)$, so if we can find a quadratic form q_3 of signature $(3, 0)$, discriminant 1, and p–excesses equal to e'_p, then $q' = q_3 \oplus \langle 1, -1\rangle$ will have p–excesses equal to those of q and discriminant -1. Since it will also have signature $(4, 1)$, q' will be rationally equivalent to q.

It suffices, then, to show that q_3 exists. Lemma 2.10 gives five conditions on the Hasse–Witt invariants, signature, and discriminant under which a quadratic form must exist. Conditions (2)–(4) hold trivially for q_3, either because they only apply to forms of rank 2 or less, or because they merely require that the signature is valid and agrees with the sign of the discriminant. This is true because the desired signature of q_3 is $(3, 0)$ and the discriminant is 1.

Conditions (1) and (5) concern the desired Hasse–Witt invariants ϵ'_p of q_3, which can be determined from the desired discriminant 1 and desired p–excesses e'_p. We will show that $\epsilon'_p = \epsilon_p(q)$ for all p, and thus that conditions (1) and (5) are satisfied.

Recall that $\epsilon'_p = 1$ if $e'_p = e_p(\langle d(q_3), 1, 1\rangle) = e_p(\langle 1, 1, 1\rangle)$, and -1 otherwise. We can similarly compute the Hasse–Witt invariants of q to be $\epsilon_p(q) = 1$ if and only if $e_p(q) = e_p(\langle -1, 1, 1, 1, 1\rangle)$. By construction, $e'_p = e_p(q) - e_p(\langle 1, -1\rangle)$. Then note that $e_p(\langle 1, 1, 1\rangle) = e_p(\langle -1, 1, 1, 1, 1\rangle) - e_p(\langle 1, -1\rangle)$ by additivity of p–excesses. Thus, $\epsilon'_p = \epsilon_p(q)$ for all p.

In particular, for any quadratic form q, $\epsilon_p(q) = 1$ for all but finitely many p, and $\prod \epsilon_p(q) = 1$ over all primes p. These same properties must hold for ϵ'_p, so condition (1) holds. Similarly, $\epsilon_\infty(q) = 1$ since q has signature $(4, 1)$, so $\epsilon'_\infty = 1$ as well, satisfying condition (5). Now we can apply Lemma 2.10 to deduce that a valid quadratic form q_3 exists with signature $(3, 0)$, discriminant 1, and Hasse–Witt invariants $\epsilon_p(q_3) = \epsilon'_p$.

As stated above, we can take the form $q' = q_3 \oplus \langle 1, -1\rangle$, which is rationally equivalent to q, has discriminant -1, and is of the form $\langle a, b, c, 1, -1\rangle$, where $q_3 = \langle a, b, c\rangle$. □

On the pure quaternions of any quaternion algebra, we can define the orthogonal group

$$O(N, Q_0) = \{f : Q_0 \to Q_0 \mid f \text{ is linear and } N(f(x)) = N(x) \text{ for all } x \in Q_0\}$$

as the set of linear transformations on Q_0 that preserve the norm form. These transformations can be described as conjugation by the units Q^* of Q. This is the intuition behind the following theorem from [14, Section 2.4]:

Theorem 2.11 *Let $Q = ((-a, -b)/\mathbb{Q})$ and $q = \langle a, b, ab\rangle$. Then $\mathrm{SO}(q, \mathbb{Q})$ is isomorphic to $Q^*/Z(Q^*)$, where $Z(G)$ denotes the center of G.*

There are three more theorems from [14] that are used in our argument. We state them here, along with a relevant definition, and remark that it will be important to obstruct certain torsion from occurring in $Q^*/Z(Q^*)$.

Definition 2.12 (ramification) A prime p ramifies a quaternion algebra Q over $\mathbb{Q}$ if $Q \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is isomorphic to the unique division algebra of dimension 4 over $\mathbb{Q}_p$. Otherwise, $Q \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is isomorphic to the algebra of 2×2 matrices $M_2(\mathbb{Q}_p)$, and we say Q splits over p.

Theorem 2.13 [14, Lemma 12.5.6] *Let ξ_n for $n > 2$ be a primitive n^{th} root of unity, and Q be a quaternion algebra over $\mathbb{Q}$. Then $Q^*/Z(Q^*)$ contains an element of order n if and only if $\xi_n + \xi_n^{-1} \in \mathbb{Q}$ and $\mathbb{Q}(\xi_n)$ embeds in Q.*

Theorem 2.14 [14, Theorem 7.3.3] *Given a quaternion algebra Q over $\mathbb{Q}$ and a quadratic extension L of $\mathbb{Q}$, then L embeds in Q if and only if, for each prime p that ramifies Q, p does not split in L.*

Theorem 2.15 [14, Theorem 2.6.6] *Let $p \neq 2, \infty$ be a prime in $\mathbb{Q}$. Consider the quaternion algebra $Q = ((a,b)/\mathbb{Q})$, with both a and b squarefree.*

(1) *If p does not divide a or b, then p does not ramify Q.*

(2) *If p divides a but not b, then p ramifies Q if and only if b is a quadratic nonresidue modulo p.*

(3) *If p divides both a and b, then p ramifies Q if and only if $-a^{-1}b$ is a quadratic nonresidue modulo p.*

3 Arithmetic hyperbolic manifolds

3.1 Hyperbolic manifolds

Let $q = x_1^2 + \cdots + x_n^2 - x_{n+1}^2$ be a quadratic form of signature $(n,1)$. We define hyperbolic space using the hyperboloid model $\mathbb{H}^n = \{x \in \mathbb{R}^{n+1} \mid q(x) = -1, x_{n+1} > 0\}$, equipped with the metric derived from the inner product

$$x \circ y = \sqrt{x_1 y_1 + \cdots + x_n y_n - x_{n+1} y_{n+1}},$$

so that $(x \circ x)^2 = q(x)$. A hyperplane in $\mathbb{H}^n$ is an intersection of a subspace $V \subset \mathbb{R}^{n+1}$ with $\mathbb{H}^n$, and $\mathbb{H}^n$ has a boundary $\partial\mathbb{H}^n$ consisting of 1–dimensional subspaces of *light-like vectors* $y \in \mathbb{R}^{n+1}$ such that $q(y) = 0$. The isometries of $\mathbb{H}^n$ must preserve q, and in fact $\mathrm{Isom}^+(\mathbb{H}^n) = \mathrm{SO}_0(q, \mathbb{R})$.

Observe that we can perform this construction with any form q' of signature $(n,1)$ in place of q. The resulting space $\mathbb{H}^n_{q'}$ is isometric to $\mathbb{H}^n$, although both are different subsets of $\mathbb{R}^{n+1}$ and points in $\mathbb{Q}^{n+1}$ in one model may not correspond to points in $\mathbb{Q}^{n+1}$ in the other. Thus, $\mathrm{Isom}^+(\mathbb{H}^n)$ is isomorphic to $\mathrm{Isom}^+(\mathbb{H}^n_{q'})$. In particular, there is a linear transformation A that maps any $\mathbb{H}^n_{q'}$ to $\mathbb{H}^n$ isometrically, so any isometry $\gamma \in \mathrm{Isom}^+(\mathbb{H}^n_{q'})$ can be said to sit in $\mathrm{Isom}^+(\mathbb{H}^n)$ as $A\gamma A^{-1}$. We will sometimes abuse notation and refer to any $\mathbb{H}^n_{q'}$ as $\mathbb{H}^n$ when it is clear which quadratic form is being used.

We will use the notion of hyperplanes P sitting *rationally* inside $\mathbb{H}^n_q$. By this, we mean P is the intersection of $\mathbb{H}^n_q$ with a subspace $V \subset \mathbb{R}^{n+1}$ determined by a system of equations with rational coefficients. This notion depends on our choice of q, which in our case will always have coefficients in $\mathbb{Z}$.

A hyperbolic n–manifold is a quotient $\mathbb{H}^n/\Gamma$ of hyperbolic n–space by a discrete, torsion-free group Γ acting on $\mathbb{H}^n$ via isometries. If Γ is not torsion-free, a hyperbolic orbifold results instead. A cusp of a finite-volume hyperbolic n–manifold or orbifold is a subset of the manifold homeomorphic to $B \times \mathbb{R}^+$ for some cross-section B. Cusps result from the parabolic elements of Γ that fix a single point y of $\partial\mathbb{H}^n$. Specifically, since $\mathrm{Stab}_\Gamma(y)$ acts on a horosphere centered at y, which has a flat geometry, the cross-section of the corresponding cusp is given by $B = \mathbb{E}^{n-1}/\mathrm{Stab}_\Gamma(y)$. We consider only finite-volume hyperbolic manifolds, so B is compact. Furthermore, if $\mathbb{H}^n/\Gamma$ is orientable then so is B. For more information on cusps of hyperbolic manifolds and the *thick–thin decomposition* we refer the reader to [23, Chapter 12].

Definition 3.1 (commensurability) Two subgroups Γ_1 and Γ_2 of a group Γ are *commensurable* if $\Gamma_1 \cap \Gamma_2$ has finite index in both Γ_1 and Γ_2. Two hyperbolic orbifolds $\mathbb{H}^n/\Gamma_1$ and $\mathbb{H}^n/\Gamma_2$ are *commensurable* if $\gamma\Gamma_1\gamma^{-1}$ and Γ_2 are commensurable in $\mathrm{Isom}(\mathbb{H}^n)$ for some $\gamma \in \mathrm{Isom}(\mathbb{H}^n)$.

Note that two orbifolds are commensurable if and only if they share a finite cover.

3.2 Arithmetic manifolds

Since we are working solely with cusped hyperbolic manifolds, all arithmetic hyperbolic manifolds in this paper are of *simplest type*. This allows us to use a simpler definition of arithmetic hyperbolic manifolds than the more involved general definition. This is stated, for example, in [19, Proposition 6.4.2] with the condition $n \neq 3, 7$, although this condition is unnecessary.

Definition 3.2 (arithmetic hyperbolic orbifold/arithmetic group) Let M be a finite-volume cusped hyperbolic n–orbifold with $\pi_1(M) = \Gamma < \mathrm{Isom}(\mathbb{H}^n)$. Then M is arithmetic if there exists a quadratic form q of signature $(n,1)$ such that $A^{-1}\Gamma A < \mathrm{Isom}(\mathbb{H}^n_q)$ is commensurable to $\mathrm{SO}_0(q,\mathbb{Z})$, where A is the linear transformation that maps $\mathbb{H}^n_q$ to $\mathbb{H}^n$ isometrically. We say Γ is arithmetic under the same condition, that is, when Γ is conjugate to a subgroup of $\mathrm{Isom}^+(\mathbb{H}^n_q)$ commensurable to $\mathrm{SO}_0(q,\mathbb{Z})$.

A hyperbolic arithmetic n–manifold is a hyperbolic arithmetic n–orbifold that is also a hyperbolic manifold. Henceforth we may refer to the arithmetic orbifold $\mathbb{H}^n_q/\mathrm{SO}_0(q,\mathbb{Z})$ as $\mathbb{H}^n/\mathrm{SO}_0(q,\mathbb{Z})$ using this particular embedding, without ambiguity.

To any cusped arithmetic hyperbolic n–orbifold M we can associate the (nonunique) quadratic form q from the definition. There are easily checkable conditions on quadratic forms q_1 and q_2 that determine whether $\Gamma_1 = \mathrm{SO}_0(q_1, \mathbb{Z})$ and $\Gamma_2 = \mathrm{SO}_0(q_2, \mathbb{Z})$ are commensurable as subgroups of $\mathrm{Isom}(\mathbb{H}^n)$, identifying both $\mathrm{Isom}(\mathbb{H}^n_{q_1})$ and $\mathrm{Isom}(\mathbb{H}^n_{q_2})$ with $\mathrm{Isom}(\mathbb{H}^n)$, and are thus associated to the same orbifolds.

Proposition 3.3 [18, Theorem 1] *Let M_1 and M_2 be arithmetic hyperbolic orbifolds with associated quadratic forms q_1 and q_2, respectively. Then M_1 and M_2 are commensurable if and only if q_1 and q_2 are projectively equivalent.*

One way to determine whether two quadratic forms q_1 and q_2 of signature $(4, 1)$ are projectively equivalent is to scale both so they have the same discriminant, and then compare Hasse–Witt invariants. In particular, since such forms have odd rank, if q_i has discriminant $-d_i$ then the form $d_i q_i$ must have discriminant -1. Thus, we can deal with rational equivalence rather than projective equivalence by associating to a commensurability class of arithmetic hyperbolic 4–manifolds a (nonunique) quadratic form q of discriminant -1. Furthermore, by Lemma 2.7 we can take q to be of quaternion type. We summarize this discussion:

Corollary 3.4 *Every commensurability class C of cusped arithmetic hyperbolic 4–orbifolds has an associated quadratic form q of quaternion type such that $\mathbb{H}^4/\mathrm{SO}_0(q, \mathbb{Z})$ lies in C.*

3.3 Systoles

Definition 3.5 (systole length) The systole length of a manifold M is the minimal length of a closed geodesic in M.

The arithmetic n–manifolds we deal with have a minimum bound on the systole length. The following proposition is an application of Corollary 1.3 or 1.8 from [8], depending on whether n is even or odd:

Proposition 3.6 *There is a lower bound on the systole length of a cusped arithmetic hyperbolic 4–manifold.*

We will use this fact to show that certain finite-volume hyperbolic n–manifolds are nonarithmetic.

4 Compact flat 3–manifolds

Recall from Section 3.1 that finite-volume cusped hyperbolic n–manifolds $M = \mathbb{H}^n/\Gamma$ have compact flat $(n-1)$–manifolds B for the cross-sections of their cusps, and if M

M	$\pi_1(M)$	$\mathrm{Hol}(\pi_1(M))$
3–torus	$\mathbb{Z}^3 = \langle t_1, t_2, t_3 \mid t_i t_j = t_j t_i \rangle$	$\mathbf{1}$
$\frac{1}{2}$–twist	$\langle \alpha, t_1, t_2, t_3 \mid t_i t_j = t_j t_i, \alpha^2 = t_1, \alpha t_2 \alpha^{-1} = t_2^{-1}, \alpha t_3 \alpha^{-1} = t_3^{-1} \rangle$	$\mathbb{Z}/2\mathbb{Z}$
H–W	$\langle x, y, z \mid xy^2x^{-1}y^2 = 1, yx^2y^{-1}x^2 = 1, xyz = 1 \rangle$ [4]	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$
$\frac{1}{3}$–twist	$\langle \alpha, t_1, t_2, t_3 \mid t_i t_j = t_j t_i, \alpha^3 = t_1, \alpha t_2 \alpha^{-1} = t_3, \alpha t_3 \alpha^{-1} = t_2^{-1} t_3^{-1} \rangle$	$\mathbb{Z}/3\mathbb{Z}$
$\frac{1}{4}$–twist	$\langle \alpha, t_1, t_2, t_3 \mid t_i t_j = t_j t_i, \alpha^4 = t_1, \alpha t_2 \alpha^{-1} = t_3, \alpha t_3 \alpha^{-1} = t_2^{-1} \rangle$	$\mathbb{Z}/4\mathbb{Z}$
$\frac{1}{6}$–twist	$\langle \alpha, t_1, t_2, t_3 \mid t_i t_j = t_j t_i, \alpha^6 = t_1, \alpha t_2 \alpha^{-1} = t_3, \alpha t_3 \alpha^{-1} = t_2^{-1} t_3 \rangle$	$\mathbb{Z}/6\mathbb{Z}$

Table 1: The six orientable compact flat 3–manifolds [21].

is orientable then so is B. Considering only orientable manifolds, this means that hyperbolic 2– and 3–manifolds only have one type of cusp cross-section each: S^1 and T^2, respectively. However, there are six orientable compact flat 3–manifolds up to homeomorphism, which means there are six possible cusp cross-sections for an orientable finite-volume hyperbolic 4–manifold. We give a brief description of each in Table 1 and Figure 1.

In the images depicting the fundamental domains, a face without a label is paired with its opposite face via translation, and labeled faces are paired in such a way that the labels align. Note that all but the Hantzsche–Wendt manifold differ from the 3–torus by at most a twist on one of the face pairings. All six flat manifolds are commensurable, and are in fact finitely covered by the 3–torus.

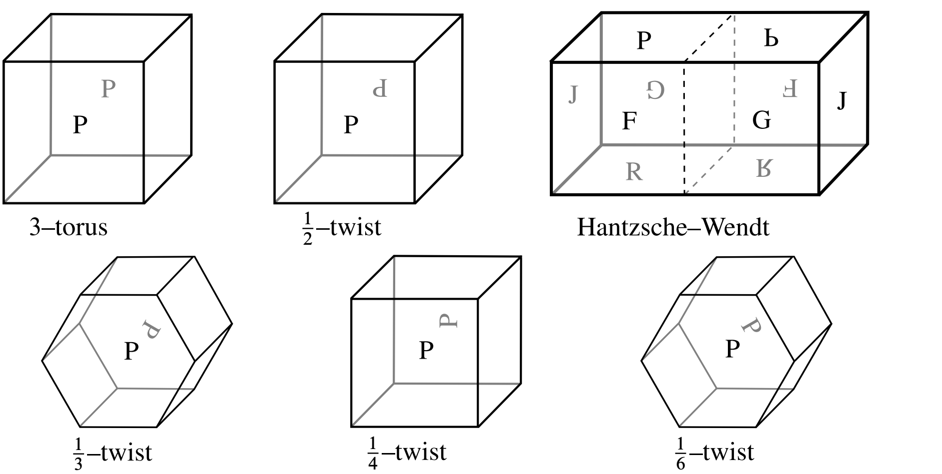


Figure 1: The fundamental domains for the manifolds in Table 1.

Every isometry of Euclidean 3–space $\mathbb{E}^3$ is an affine transformation $v \mapsto Av + w$ for some $A \in \mathrm{SO}(3)$. For a group $G < \mathrm{Isom}(\mathbb{E}^3)$, the holonomy of G is given by

$$\mathrm{Hol}(G) = \{A \in \mathrm{SO}(3) \mid (v \mapsto Av + w) \in G \text{ for some } w \in \mathbb{R}^3\}.$$

$\mathrm{Hol}(G)$ is independent of the faithful representation of G into $\mathrm{Isom}(\mathbb{E}^3)$.

5 Classes with a given cusp

One goal of the next two sections is to prove Theorem 1.1. In fact, we generalize Theorem 1.1 to a full description of exactly when a commensurability class of cusped arithmetic hyperbolic 4–manifolds contains a manifold with a given cusp type.

Theorem 5.1 *Let C be a commensurability class of cusped arithmetic hyperbolic 4–manifolds, with associated quadratic form q, scaled so that the discriminant of q is -1. Then:*

- *C must contain a manifold with a 3–torus cusp, a manifold with a $\frac{1}{2}$–twist cusp, and a manifold with a Hantzsche–Wendt cusp.*
- *C contains a manifold with a $\frac{1}{4}$–twist cusp if and only if $\epsilon_p(q) = 1$ for all $p \equiv 1 \pmod 4$.*
- *C contains a manifold with a $\frac{1}{3}$–twist cusp if and only if $\epsilon_p(q) = 1$ for all $p \equiv 1 \pmod 3$. C contains a manifold with a $\frac{1}{6}$–twist cusp under the same condition.*

In this section, we prove the positive portion of the theorem, namely that C does indeed contain certain cusp types.

Proposition 5.2 *Let C be a commensurability class of arithmetic hyperbolic 4–manifolds, with associated quadratic form q of discriminant -1. Then:*

- *C must contain a manifold with a 3–torus cusp, a manifold with a $\frac{1}{2}$–twist cusp, and a manifold with a Hantzsche–Wendt cusp.*
- *If $\epsilon_p(q) = 1$ for all $p \equiv 1 \pmod 4$, then C contains a manifold with a $\frac{1}{4}$–twist cusp.*
- *If $\epsilon_p(q) = 1$ for all $p \equiv 1 \pmod 3$, then C contains a manifold with a $\frac{1}{3}$–twist cusp and a manifold with a $\frac{1}{6}$–twist cusp.*

Our primary tool for showing that a commensurability class must contain a given cusp type is the algorithm given by Long and Reid [13]. Given a compact flat n–manifold B, this algorithm yields an arithmetic hyperbolic $(n+1)$–orbifold with a cusp of type B. We can always find an $(n+1)$–manifold with a cusp of type B covering this orbifold by [15].

Given a cusp type B of dimension n, the algorithm works as follows. Consider the holonomy group of $\pi_1(B)$. We can find a faithful representation of $\mathrm{Hol}(\pi_1(B))$ into $\mathrm{GL}(n,\mathbb{Z})$, which yields an embedding $\mathrm{Hol}(\pi_1(B)) \subset \mathrm{GL}(n,\mathbb{Z})$. Further, we can choose a signature-$(n,0)$ quadratic form q_n that is invariant under $\mathrm{Hol}(\pi_1(B))$ by considering an arbitrary signature-$(n,0)$ quadratic form r and taking the average of all the quadratic forms $r \circ A$ over $A \in \mathrm{Hol}(\pi_1(B))$, since $\mathrm{Hol}(\pi_1(B))$ is finite. Then, using linear algebra, the algorithm extends the representation into $\mathrm{GL}(n+2,\mathbb{Z})$ in such a way that $\mathrm{Hol}(\pi_1(B))$ leaves a quadratic form q' rationally equivalent to $q_n \oplus \langle 1,-1\rangle$ invariant. As a result, we see that some cover of $\mathbb{H}^{n+1}/\mathrm{SO}_0(q',\mathbb{Z})$ must contain a cusp of type B, and is commensurable to $\mathbb{H}^{n+1}/\mathrm{SO}_0(q_n \oplus \langle 1,-1\rangle,\mathbb{Z})$.

By investigating properties of quadratic forms q_n invariant under $\mathrm{Hol}(\pi_1(B))$, we characterize the commensurability classes of arithmetic hyperbolic manifolds that can be output by this algorithm. Since we're working with flat 3–manifolds and hyperbolic 4–manifolds, we apply the algorithm with $n = 3$.

Proof of Proposition 5.2 Given the commensurability class C, we can choose a quadratic form $q = \langle x, y, xy, 1, -1\rangle$ of quaternion type such that $\mathbb{H}^4/\mathrm{SO}_0(q,\mathbb{Z}) \in C$ by Lemma 2.7. Note that q has discriminant -1. We can compute the Hasse–Witt invariants $\epsilon_p(q)$.

First let B be the 3–torus, $\frac{1}{2}$–twist, or Hantzsche–Wendt manifold. These have holonomy groups of 1, $\mathbb{Z}/2\mathbb{Z}$, and $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, respectively. Each holonomy group has a representation into $\mathrm{GL}(3,\mathbb{R})$ consisting solely of diagonal matrices with ± 1 along the diagonal. In particular, these representations fix any quadratic form $\langle a, b, ab\rangle$ of rank 3. Thus, we can apply the Long–Reid algorithm to find a representation of the corresponding Bieberbach group into $\mathrm{SO}_0(\langle a, b, ab, k, -k\rangle,\mathbb{Z})$. Set $a = x$ and $b = y$. Then $\langle a, b, ab, k, -k\rangle$ is rationally equivalent to $\langle a, b, ab, 1, -1\rangle = \langle x, y, xy, 1, -1\rangle$. This yields an orbifold commensurable to $\mathbb{H}^4/\mathrm{SO}_0(q,\mathbb{Z})$ that has the desired cusp type. By [15], there is also a manifold with the desired cusp type.

Next, consider the $\frac{1}{4}$–twist cusp. This flat manifold has holonomy group $\mathbb{Z}/4\mathbb{Z}$, and has a representation ρ into $\mathrm{SL}(3,\mathbb{Z})$ mapping its generator g_4 to

$$\rho(g_4) = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This holonomy preserves any quadratic form $q_3 = \langle a, a, b\rangle$, so the Long–Reid algorithm finds a representation of B into $\mathrm{SO}_0(\langle a, a, b, k, -k\rangle, \mathbb{Z})$, which is commensurable to $\mathrm{SO}_0(\langle ab, ab, 1, 1, -1\rangle, \mathbb{Z})$. The Hasse–Witt invariant at p of the form $q' = \langle ab, ab, 1, 1, -1\rangle$ is equal to the Hilbert symbol $(ab, ab)_p$. Let $ab = up^\alpha$, where u is an integer not divisible by p. By definition, for $p > 2$ and $\tau(p) = \frac{1}{2}(p-1)$,

$$(ab, ab)_p = (-1)^{\tau(p)\alpha\alpha}\left(\frac{u}{p}\right)^\alpha\left(\frac{u}{p}\right)^\alpha = (-1)^{\tau(p)\alpha}.$$

Note that $\tau(p)$ is even if $p \equiv 1 \pmod 4$ and odd if $p \equiv 3 \pmod 4$.

So if $p \equiv 1 \pmod 4$, we always have $\epsilon_p(q') = (ab, ab)_p = 1$. But if $p \equiv 3 \pmod 4$, then $\epsilon_p(q') = -1$ if and only if p divides ab an odd number of times. Given the finite set of primes $p_i > 2$ such that $\epsilon_p(q) = -1$, as long as there is no p_i such that $p_i \equiv 1 \pmod 4$, we can now ensure that there is a quadratic form $q'' = \left\langle \prod p_i, \prod p_i, 1, 1, -1\right\rangle$ such that $\epsilon_p(q'') = \epsilon_p(q)$. Note that the identity $\prod \epsilon_p(q) = 1$ ensures that $\epsilon_2(q'') = \epsilon_2(q)$ as well. Thus q'' and q both have the same Hasse–Witt invariants, as well as discriminant -1 and signature $(4, 1)$. Hence q'' is rationally equivalent to q and, taking $ab = \prod p_i$, we see that $\mathbb{H}^4/\mathrm{SO}_0(q'', \mathbb{Z})$ must have a finite cover with a $\frac{1}{4}$–twist cusp. Thus we can construct a manifold in C with a $\frac{1}{4}$–twist cusp.

The arguments for the $\frac{1}{3}$–twist and the $\frac{1}{6}$–twist cusps are similar. The holonomy groups $\mathbb{Z}/3\mathbb{Z}$ and $\mathbb{Z}/6\mathbb{Z}$ have representations ρ_3 and ρ_6 into $\mathrm{SL}(3, \mathbb{Z})$ mapping the respective generators g_3 and g_6 as

$$\rho_3(g_3) = \begin{bmatrix} -1 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \rho_6(g_6) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Under this representation, both holonomy groups preserve quadratic forms of the form $q'(x) = 4ax_1^2 + 4ax_2^2 - 4ax_1x_2 + 3bx_3^2$. With some effort, we can show that this form is projectively equivalent to $q'' = \langle ab, 3ab, 3, 1, -1\rangle$. We then compute that $\epsilon_p(q'') = (ab, 3ab)_p(3, 3)_p$. The second Hilbert symbol $(3, 3)_p$ is equal to -1 at $p = 2, 3$ and equal to 1 everywhere else. To compute the first Hilbert symbol $(ab, 3ab)_p$, we consider the case $p = 3$ separately from $p \neq 2, 3$. (We'll ignore $p = 2$ for now since the identity $\prod \epsilon_p(q) = \prod \epsilon_p(q'') = 1$ will ensure that $\epsilon_2(q'') = \epsilon_2(q)$ if all other Hasse–Witt invariants are equal.)

For $p = 3$, suppose $ab = 3^\alpha u$ where u is not divisible by 3. Then $3ab = 3^{\alpha+1}u$, so

$$(ab, 3ab)_3 = (-1)^{\alpha(\alpha+1)\tau(3)}\left(\frac{u}{3}\right)^\alpha\left(\frac{u}{3}\right)^{\alpha+1} = \left(\frac{u}{3}\right).$$

Thus $(ab, 3ab)_3 = 1$ if $u \equiv 1 \pmod 3$ and -1 if $u \equiv 2 \pmod 3$.

For $p \neq 3$, let $ab = p^\alpha u$ where u is not divisible by p, so that $3ab = p^\alpha(3u)$. Then

$$(ab, 3ab)_p = (-1)^{\alpha\alpha\tau(p)} \left(\frac{u}{p}\right)^\alpha \left(\frac{3u}{p}\right)^\alpha = \left((-1)^{\tau(p)}\left(\frac{3}{p}\right)\right)^\alpha.$$

Combining $(-1)^{\tau(p)}$ with quadratic reciprocity, we can see that, for $p > 2$, $(ab, 3ab)_p = 1^\alpha$ if $p \equiv 1 \pmod 3$ and $(-1)^\alpha$ if $p \equiv 2 \pmod 3$. Consider the finite set of primes $p_i > 2$ such that $\epsilon_p(q) = -1$. As long as there is no p_i such that $p_i \equiv 1 \pmod 3$, we can take $ab = \prod p_i$ over all $p_i \equiv 2 \pmod 3$. Additionally, we can multiply ab by 2 if necessary to set $ab \equiv 1$ or $2 \pmod 3$ to obtain the desired value of $\epsilon_3(q'')$. Now $\epsilon_p(q) = \epsilon_p(q'')$ for all $p > 2$ and, as before, $\epsilon_2(q) = \epsilon_2(q'')$ due to the identity $\prod \epsilon_p(q) = \prod \epsilon_p(q'') = 1$. Now q'' is rationally equivalent to q, and from $\mathrm{SO}_0(q'', \mathbb{Z})$ we can construct a manifold in C with a $\frac{1}{3}$–twist or $\frac{1}{6}$–twist cusp, as desired. □

Remark 5.3 In addition to the six orientable compact flat 3–manifolds, there are four nonorientable ones: two double-covered by the 3–torus and two double-covered by the $\frac{1}{2}$–twist. A thorough description of these manifolds can be found in [6]. Notably, all of them have holonomies generated by orthogonal reflections. In particular, this means each fundamental group has a holonomy representation into $\mathrm{GL}(3, \mathbb{R})$ with image consisting of diagonal matrices with ± 1 along the diagonal. Thus, for the same reasons as the 3–torus, $\frac{1}{2}$–twist, and Hantzsche–Wendt manifold, all four nonorientable compact flat 3–manifolds occur as a cusp cross-section in every commensurability class of arithmetic hyperbolic 4–manifolds.

6 Classes without a given cusp

The goal of this section is to prove the negative part of Theorem 5.1, that is, to obstruct some cusp types from occurring in some commensurability classes of hyperbolic 4–manifolds. This obstruction will yield infinitely many commensurability classes that avoid each of the $\frac{1}{3}$–twist, $\frac{1}{6}$–twist, and $\frac{1}{4}$–twist.

Proposition 6.1 *Let C be a commensurability class of arithmetic hyperbolic 4–manifolds, with associated quadratic form q with discriminant -1. Then:*

- *If $\epsilon_p(q) \neq 1$ for some $p \equiv 1 \pmod 4$, then C does not contain a manifold with a $\frac{1}{4}$–twist cusp.*
- *If $\epsilon_p(q) \neq 1$ for some $p \equiv 1 \pmod 3$, then C contains neither a manifold with a $\frac{1}{3}$–twist cusp, nor a manifold with a $\frac{1}{6}$–twist cusp.*

Proof By Lemma 2.7, we can take q to be of quaternion form. Thus, without loss of generality, we can set $q = \langle a, b, ab, 1, -1\rangle$ for some positive integers a and b.

Let B be the cusp type that we want to obstruct, and let $\Delta = \pi_1(B)$. We will show that it suffices to obstruct the existence of an injective homomorphism $\Delta \to \mathrm{SO}_0(q, \mathbb{Q})$.

For the sake of contradiction, suppose C contains a manifold M with the cusp type in question. This yields an embedding $\Delta \to \pi_1(M) = \Gamma$. Because Γ is an arithmetic lattice in $\mathrm{SO}(4,1)$, we know that Γ lies in the $\mathbb{Q}$–points of some quadratic form q' [3]. Because $M \in C$, q and q' are projectively equivalent. Thus by Proposition 3.3, there exists a matrix $F \in \mathrm{GL}(5, \mathbb{Q})$ such that $F\Delta F^{-1}$ is commensurable with $\mathrm{SO}_0(q, \mathbb{Z})$ and embeds into $\mathrm{SO}_0(q, \mathbb{Q})$. Note that Δ acts on a horosphere centered at some point y in $\partial\mathbb{H}^4$.

Since y is fixed by isometries that lie in $\mathrm{SO}_0(q, \mathbb{Z})$, we can take y itself to lie in $\mathbb{Q}^5$. Additionally, since $\mathrm{SO}_0(q, \mathbb{Q})$ acts transitively on the rational points of $\partial\mathbb{H}^4_q$, we can choose y to be $(0,0,0,1,1)$ without loss of generality. Specifically, we can conjugate the image of Δ by some matrix $A' \in \mathrm{SO}_0(q, \mathbb{Q})$ such that $A'y = y_0 = (0,0,0,1,1)$ to get a new rational representation of Δ acting on a horosphere H centered at y_0.

Let $q_3 = \langle a, b, ab\rangle$ be the quadratic form such that $q_3 \oplus \langle 1, -1\rangle = q$. Given any affine transformation $\varphi \in \mathrm{Isom}(\mathbb{E}^3)$, we can write the isometry as $\varphi(v) = Av + w$, with $A \in \mathrm{SO}_0(q_3, \mathbb{R})$. Then we can map φ to an action $\rho(\varphi)$ on H by taking ρ to be induced by an isometry from $\mathbb{E}^3$ to H. Imitating [22], we can write ρ as

$$\rho(\varphi)\colon v \mapsto \begin{bmatrix} A & w & -w \\ f(w)^t A & 1 + \frac{1}{2}q_3(w) & -\frac{1}{2}q_3(w) \\ f(w)^t A & \frac{1}{2}q_3(w) & 1 - \frac{1}{2}q_3(w) \end{bmatrix} v.$$

Here f is the linear function $f(x) = (ax_1, bx_2, abx_3)^t$ such that $f(x)^t x = q_3(x)$ for any $x \in \mathbb{R}^3$. Since one can recover A from the top left and $-w$ from the top right of $\rho(\varphi)$, we see that ρ must be injective. One can check through manual calculation that ρ is a homomorphism, and that all elements in $\rho(\mathrm{Isom}(\mathbb{E}^3))$ preserve both q and y_0, and thus act on H. All isometries of H must be of the form $\rho(\varphi)$ above for some $\varphi \in \mathrm{Isom}(\mathbb{E}^3)$, so in particular, every element of $\rho(\Delta)$ has this form.

If Δ is the fundamental group of the $\frac{1}{3}$–twist, $\frac{1}{4}$–twist, or $\frac{1}{6}$–twist cusp, it has holonomy group $\mathbb{Z}/3\mathbb{Z}$, $\mathbb{Z}/4\mathbb{Z}$, or $\mathbb{Z}/6\mathbb{Z}$, respectively. The holonomy is represented by the matrix A above, so in order to embed Δ into $\mathrm{SO}_0(q, \mathbb{Q})$ there must exist an isometry φ with A that is 3–torsion or 4–torsion. Since A is a submatrix of $\rho(\varphi)$, which has rational entries, it must have rational entries. Thus, if we can obstruct 3–torsion or 4–torsion from $\mathrm{SO}_0(q_3, \mathbb{Q})$, then we can obstruct the existence of an embedding $\Delta \to \mathrm{SO}_0(q, \mathbb{Q})$.

Now consider the quaternion algebra $Q = ((-a,-b)/\mathbb{Q})$. The norm form of Q_0 is given by $ax_1^2+bx_2^2+abx_3^2 = q_3(x)$, so by Theorem 2.11, $\mathrm{SO}(q_3,\mathbb{Q})$ is isomorphic to $Q^*/Z(Q^*)$. Thus, if we obstruct torsion of some degree from appearing in $Q^*/Z(Q^*)$, then we obstruct it from $\mathrm{SO}_0(q_3,\mathbb{Q}) < \mathrm{SO}(q_3,\mathbb{Q})$ as well.

Now we apply Theorem 2.13. For $n=3$ and $n=4$, clearly $\xi_n+\xi_n^{-1}\in\mathbb{Q}$. So there are no order-$n$ elements of $Q^*/Z(Q^*)$ if and only if $\mathbb{Q}(\xi_n)$ does not embed in Q. Furthermore, by Theorem 2.14, the field $\mathbb{Q}(\xi_n)$ embeds in Q if and only if $\mathbb{Q}(\xi_n)\otimes_{\mathbb{Q}}\mathbb{Q}_p$ is a field for each $p\in\mathrm{Ram}(Q)$. The latter occurs exactly when p does not split in $\mathbb{Q}(\xi_n)$. Thus, in order to obstruct n–torsion, we wish to show there is some $p\in\mathrm{Ram}(Q)$ such that p splits in $\mathbb{Q}(\xi_n)$.

To check this condition, we must first determine when $p\in\mathrm{Ram}(Q)$. If neither $-a$ nor $-b$ is divisible by p an odd number of times, then p does not ramify by Theorem 2.15(1). Note that if both $-a$ and $-b$ are divisible by p an odd number of times, then ab is not. Since a, b, and ab are interchangeable when constructing Q, in this case we can pass to $Q' = ((-a,-ab)/\mathbb{Q})$ to ensure that p divides only one of $-a$ and $-b$ an odd number of times. Without loss of generality, say p divides $-a$ but not $-b$. Then, by Theorem 2.15(2), p ramifies if and only if b is a nonsquare modulo p.

We claim that p ramifies over Q exactly when the Hasse–Witt invariant $\epsilon_p(q)$ equals -1. Using the definitions of the Hasse–Witt invariant and the Hilbert symbol, we can expand $\epsilon_p(q)$. Let $a = p^\alpha j$ and $b = p^\beta k$ with j and k relatively prime to p. Then $ab = p^{\alpha+\beta}jk$, so

$$\begin{aligned}\epsilon_p(q) &= \epsilon_p(\langle a,b,ab,1,-1\rangle) = (a,b)_p(ab,ab)_p\\ &= \left[(-1)^{\alpha\beta\tau(p)}\left(\frac{j}{p}\right)^{\beta}\left(\frac{k}{p}\right)^{\alpha}\right]\left[(-1)^{(\alpha+\beta)(\alpha+\beta)\tau(p)}\left(\frac{jk}{p}\right)^{\alpha+\beta}\left(\frac{jk}{p}\right)^{\alpha+\beta}\right]\\ &= (-1)^{\tau(p)(\alpha\beta+\alpha+\beta)}\left(\frac{j}{p}\right)^{\beta}\left(\frac{k}{p}\right)^{\alpha}.\end{aligned}$$

If both α and β are even, then $\epsilon_p(q) = (-1)^0(j/p)^0(k/p)^0 = 1$. As shown above, p does not ramify over Q in this case.

If both α and β are odd, then we can choose to use $Q' = ((-a,-ab)/\mathbb{Q})$ as before. So, unless both α and β are even, without loss of generality we can assume α is odd and β is even. Then $\epsilon_p(q) = (-1)^{\tau(p)}(k/p)$. Note that $(-1/p)$ is 1 when $p\equiv 1 \pmod 4$ and -1 when $p\equiv 3 \pmod 4$, so $(-1)^{\tau(p)} = (-1/p)$. Thus, since $b = p^\beta k$ with β even, we have $\epsilon_p(q) = (-1/p)(k/p) = (-k/p) = (-b/p)$. We already showed that

p ramifies over Q exactly when $-b$ is a nonsquare modulo p in this case, which is equivalent to the condition $\epsilon_p(q) = (-b/p) = -1$. Now, in all cases, p ramifies over Q exactly when $\epsilon_p(q) = -1$.

Next, we investigate when p splits in $\mathbb{Q}(\xi_n)$. When $n = 3$ or 6 we have $\mathbb{Q}(\xi_n) = \mathbb{Q}(\sqrt{-3})$, and if $n = 4$ we have $\mathbb{Q}(\xi_n) = \mathbb{Q}(\sqrt{-1})$. It is well known that p splits in $\mathbb{Q}(\sqrt{a})$ if and only if a is a quadratic residue modulo p, so p splits in $\mathbb{Q}(\sqrt{-3})$ exactly when $p \equiv 1 \pmod 3$ and in $\mathbb{Q}(\sqrt{-1})$ exactly when $p \equiv 1 \pmod 4$.

Now, suppose there is some prime p such that $\epsilon_p(q) = -1$ and $p \equiv 1 \pmod 4$. Then, p ramifies over Q and p splits in $\mathbb{Q}(\xi_n)$. Thus, as stated above, $\mathrm{SO}_0(q_3, \mathbb{Q}) \cong Q^*/Z(Q^*)$ has no 4–torsion. As a result, the $\frac{1}{4}$–twist group B cannot possibly embed into $\mathrm{SO}_0(q, \mathbb{Q})$, so there is no $\frac{1}{4}$–twist cusp in the associated commensurability class of hyperbolic 4–manifolds. In fact, this same argument suffices to show there are no $\frac{1}{4}$–twist cusps in the class of orbifolds, either.

By similar logic, we can also see that if there is a prime p such that $\epsilon_p(q) = -1$ and $p \equiv 1 \pmod 3$, then there is no 3–torsion in $\mathrm{SO}_0(q_3, \mathbb{Q}) \cong Q^*/Z(Q^*)$. Thus the commensurability class of hyperbolic 4–manifolds (or orbifolds) associated to q must avoid $\frac{1}{3}$–twist cusps and $\frac{1}{6}$–twist cusps. □

Between Propositions 5.2 and 6.1, we've exhausted all possible commensurability classes for each cusp type. This suffices to prove Theorem 5.1. Theorem 1.1 follows.

Example 6.2 For $q_6 = \langle 1, 1, 7, 7, -1 \rangle$ the commensurability class of $\mathbb{H}^4/\mathrm{SO}_0(q_6, \mathbb{Z})$ avoids the $\frac{1}{3}$–twist and $\frac{1}{6}$–twist, since $\epsilon_7(q_6) = -1$ and $7 \equiv 1 \pmod 3$.

Example 6.3 For $q_4 = \langle 1, 2, 5, 10, -1 \rangle$ the commensurability class of $\mathbb{H}^4/\mathrm{SO}_0(q_4, \mathbb{Z})$ avoids the $\frac{1}{4}$–twist, since $\epsilon_5(q_4) = -1$ and $5 \equiv 1 \pmod 4$.

7 Obstructions in higher dimensions

Using Theorem 5.1, we can prove a version of Theorem 1.1 one dimension higher. Namely, some commensurability classes of hyperbolic 5–manifolds avoid some cusp types associated to flat 4–manifolds. Our strategy will be to show that an arithmetic hyperbolic 5–manifold with cusp $B \times S^1$ must contain a 4–dimensional totally geodesic submanifold with cusp B, and then manipulate Hasse–Witt invariants to show that, sometimes, no such submanifold can contain B as a cusp.

Proposition 7.1 *Let B be either the $\frac{1}{3}$–twist, $\frac{1}{4}$–twist, or $\frac{1}{6}$–twist. Then any arithmetic hyperbolic 5–manifold M with $B \times S^1$ as a cusp cross-section contains an immersed finite-volume totally geodesic submanifold W of codimension 1 with B as a cusp cross-section.*

Proof Let Γ be the fundamental group of M. As M is arithmetic, it is commensurable to some orbifold $\mathbb{H}^5/\mathrm{SO}_0(q,\mathbb{Z})$. Let y be a light-like vector in $\mathbb{H}^5_q$ that lies above the $B \times S^1$ cusp under the universal covering map of M.

The parabolic elements of Γ that fix y act on a horosphere E centered at y which is isomorphic to $\mathbb{E}^4$. Without loss of generality, we can take E to be the horosphere passing through $(0,0,0,0,0,1)$ by conjugating by an element of $\mathrm{SO}_0(q,\mathbb{Q})$. Note that $\mathrm{Stab}_\Gamma(y)$ is isomorphic to $\pi_1(B \times S^1) = \pi_1(B) \times \mathbb{Z}$, which acts on $\mathbb{E}^3 \times \mathbb{E}^1$. We can choose a flat subspace $P' \subset E$ of dimension 3 such that $H = \mathrm{Stab}_\Gamma(y) \cap \mathrm{Stab}_\Gamma(P')$ is isomorphic to $\pi_1(B)$. Let γ_1, γ_2, and γ_3 be three translations that generate the translation subgroup of H.

Unlike in $\mathrm{Isom}^+(\mathbb{H}^4)$, we can't assume that each γ_i lies in $\mathrm{SO}(q,\mathbb{Q})$ for $i = 1,2,3$. However, we can argue as follows. The γ_i act by translation on E, and so are parabolic translations. One can check, by applying ρ from Proposition 6.1 to any translation $v \mapsto Iv + w$, that this means each γ_i must be unipotent as an element of $\mathrm{SO}_0(q,\mathbb{R})$. For each γ_i, there is some positive integer k such that γ_i^k lies in $\mathrm{SO}_0(q,\mathbb{Z})$, since Γ is commensurable to $\mathrm{SO}_0(q,\mathbb{Z})$. Hence, the field of coefficients of γ_i^k, denoted by $F(\gamma_i^k)$, is $\mathbb{Q}$. This allows us to argue that $F(\gamma_i) = \mathbb{Q}$, and so $\gamma_i \in \mathrm{SO}_0(q,\mathbb{Q})$. The justification of the previous sentence is somewhat technical, so we defer it to Lemma 7.6.

The three translations γ_i act on the three-dimensional subspace $P' \subset E$. Since each $\gamma_i \in \mathrm{SO}_0(q,\mathbb{Q})$, P' must sit rationally in $E \subset \mathbb{H}^5_q$. To see this, pick any rational point in E, say $O = (0,0,0,0,1)$, and notice that $\gamma_i(O) \in \mathbb{Q}^5$ for all i.

The four points O, $\gamma_1(O)$, $\gamma_2(O)$, and $\gamma_3(O)$, together with y a rational line in $\partial\mathbb{H}^5_q \subset \mathbb{R}^6$, determine a four-dimensional hyperplane P which must also sit rationally in $\mathbb{H}^5_q$. Hence, after an appropriate change of basis over $\mathbb{Q}$, the quadratic form q restricts to a rank-5 form f on the 5–dimensional subspace $V \subset \mathbb{R}^6$ containing P. Then since P consists of exactly the points in V satisfying $f(x) = q(x) = -1$ and $x_6 > 0$, P sits in V as $\mathbb{H}^4_f$. In particular, this means $\mathrm{Isom}^+(P) = \mathrm{SO}_0(f,\mathbb{R})$, so $\mathrm{Isom}^+(P) \cap \mathrm{SO}_0(q,\mathbb{Z}) = \mathrm{SO}_0(f,\mathbb{Z})$. Note that this group is commensurable to $\mathrm{Isom}^+(P) \cap \Gamma$, as $\mathrm{SO}_0(q,\mathbb{Z})$ is commensurable to Γ. Thus $\mathrm{Isom}^+(P) \cap \Gamma$ is arithmetic and its action on P has finite covolume. Furthermore, $\mathrm{Isom}^+(P) \cap \mathrm{Stab}_\Gamma(y) = \mathrm{Isom}^+(P') \cap \mathrm{Stab}_\Gamma(y) = H$, so $W = P/(\mathrm{Isom}^+(P) \cap \Gamma)$ has a cusp at y with cross-section B. Now, W is an immersed finite-volume totally geodesic submanifold of M with cusp B. □

This completes the first half of the proof. For the second, using Hasse–Witt invariants we prove that we should not find any totally geodesic 4–manifolds in our 5–manifold class with a cusp of type B, yielding a contradiction. The next step, then, is to find the Hasse–Witt invariants associated to such submanifolds.

Proposition 7.2 *Let q be a quadratic form of signature $(5,1)$, discriminant -1, and Hasse–Witt invariants $\epsilon_p(q)$, and let M be a hyperbolic 5–manifold commensurable to $\mathbb{H}^5/\mathrm{SO}_0(q,\mathbb{Z})$. Then any immersed finite-volume totally geodesic 4–dimensional submanifold $W \subset M$ must be commensurable to $\mathbb{H}^4/\mathrm{SO}_0(f,\mathbb{Z})$, where f is a quadratic form of signature $(4,1)$, discriminant -1, and Hasse–Witt invariants $\epsilon_p(f) = \epsilon_p(q)$.*

Proof Since M is arithmetic, W is also arithmetic [17, Theorem 3.2]. Thus, we know W is commensurable to $\mathbb{H}^4/\mathrm{SO}_0(f,\mathbb{Z})$ for some quadratic form f of signature $(4,1)$, which we can scale to ensure discriminant -1. All that remains to be shown is that $\epsilon_p(f) = \epsilon_p(q)$ at all primes p.

Let $f = \langle a,b,c,d,-abcd\rangle$ over a quadratic space with basis $\{v_1,\dots,v_5\}$. Since W is an arithmetic manifold commensurable to $\mathbb{H}^4/\mathrm{SO}_0(f,\mathbb{Z})$, we know $\pi_1(W) < \mathrm{SO}_0(f,\mathbb{Q})$ [3]. In particular, $\pi_1(W)$ acts on $\mathbb{H}^5$ in such a way that it preserves f and a 4–dimensional hyperplane P. Taking a vector w transverse to P and adding it to the basis above, we have a basis $\{v_1,\dots,v_5,w\}$ upon which we can define our quadratic form q. Though q may not be diagonal, we can use the Gram–Schmidt process to find a basis which makes q diagonal. And, since q restricted to $\mathrm{span}(\{v_1,\dots,v_5\})$ is already diagonal, the only basis element that is affected is w. Thus, since q has signature $(5,1)$, it can be written as a diagonal form $\langle a,b,c,d,-abcd,e\rangle$ for some positive $e \in \mathbb{Z}$. Since we started with the assumption that the discriminant of q is -1, we can conclude $e = 1$.

It is now easy to show that the Hasse–Witt invariants of $f = \langle a,b,c,d,-abcd\rangle$ are equal to the Hasse–Witt invariants of $q = \langle a,b,c,d,-abcd,1\rangle$. Since any Hilbert symbol $(1,x)_p$ equals 1,

$$\begin{aligned}
\epsilon_p(q) &= (a,b)_p(a,c)_p(a,d)_p(a,-abcd)_p(b,c)_p(b,d)_p(b,-abcd)_p(c,d)_p \\
&\qquad \cdot (c,-abcd)_p(d,-abcd)_p(1,a)_p(1,b)_p(1,c)_p(1,d)_p(1,-abcd)_p \\
&= (a,b)_p(a,c)_p(a,d)_p(a,-abcd)_p(b,c)_p(b,d)_p(b,-abcd)_p(c,d)_p \\
&\qquad \cdot (c,-abcd)_p(d,-abcd)_p \\
&= \epsilon_p(f).
\end{aligned}$$

□

Theorem 7.3 *Let B be either the $\frac{1}{3}$–twist, $\frac{1}{4}$–twist, or $\frac{1}{6}$–twist. Then there exist infinitely many commensurability classes of arithmetic hyperbolic 5–manifolds that contain no manifolds with cusp cross-section given by $B \times S^1$.*

Proof Consider any quadratic form q of signature $(5,1)$ and discriminant -1. We claim that if $\epsilon_p(q) = -1$ for any $p \equiv 1 \pmod 3$ then the commensurability class C of $\mathbb{H}^5/\mathrm{SO}_0(q,\mathbb{Z})$ cannot contain $B \times S^1$ for B the $\frac{1}{3}$–twist or the $\frac{1}{6}$–twist, and if $\epsilon_p(q) = -1$ for any $p \equiv 1 \pmod 4$ then this commensurability class cannot contain $B \times S^1$ for B the $\frac{1}{4}$–twist.

By Proposition 7.1, any manifold M in C with a $B \times S^1$ cusp must contain an immersed totally geodesic submanifold W with a B cusp. By Proposition 7.2, W must be commensurable to some $\mathbb{H}^4/\mathrm{SO}_0(q',\mathbb{Z})$ with $\epsilon_p(q') = \epsilon_p(q)$ for all primes p. But by Theorem 5.1, a manifold with these Hasse–Witt invariants cannot have a cusp with cross-section B. Thus we've reached a contradiction, and such an M cannot exist in C. □

It is tempting to apply this argument repeatedly to find commensurability classes in higher-dimensional hyperbolic manifolds that avoid certain cusp types. Unfortunately, this argument fails to work even in dimension 6, because Proposition 7.2 fails to generalize. Proposition 7.2 relies on the fact that we can rescale a quadratic form of rank 5 to control the discriminant. In rank 6, rescaling a quadratic form by k multiplies the discriminant by k^6, so the discriminant does not change in $\mathbb{Q}^*/(\mathbb{Q}^*)^2$.

In fact, we can prove that repeatedly taking products of a compact flat manifold B with S^1 will eventually yield a manifold that occurs as a cusp cross-section in all arithmetic hyperbolic manifolds of the appropriate dimension. Thus, if we want to find cusp types with obstructions in higher dimensions, we'll have to use nontrivial high-dimensional flat manifolds.

Theorem 7.4 *Let B be a compact flat n–manifold. Then $B \times (S^1)^k$ occurs as a cusp cross-section in every commensurability class C of cusped arithmetic hyperbolic $(n+k+1)$–manifolds of simplest type for sufficiently high k.*

Proof First, we prove the result for $n+k+1$ even. When $n+k+1$ is even, any commensurability class C is associated with a quadratic form q of discriminant -1, since q has odd rank and we can scale q to control the discriminant.

Note that $B \times (S^1)^k$ has the same associated holonomy group as B. Since B is a flat manifold, the holonomy of its fundamental group $\mathrm{Hol}(\pi_1(B))$ must be finite. As such, $\mathrm{Hol}(\pi_1(B))$ must be a subgroup of a symmetric group S_m. Let q_m denote the quadratic form $\langle 1,\dots,1\rangle$ of rank m. The natural representation σ of S_m into permutation matrices in $\mathrm{GL}(m,\mathbb{Z})$ clearly preserves q_m. Restricting σ to $\mathrm{Hol}(\pi_1(B))$, we have a representation of $\mathrm{Hol}(\pi_1(B))$ that preserves q_m and must have entries in $\mathbb{Z}$. Let $q_m' = q_m \oplus \langle 1,-1\rangle$. We can use the Long–Reid algorithm [13] as in Proposition 5.2 to construct an orbifold with cusp cross-section $B \times (S^1)^k$ in the commensurability class of $\mathbb{H}^{n+k+1}/\mathrm{SO}_0(q' \oplus q_m', \mathbb{Z})$ for any positive definite quadratic form q' of rank $n+k-m \geq 0$.

Now, if m is even, let $k = m-n+3$ so that $n+k+1 = m+4$, and if m is odd, let $k = m-n+4$ so that $n+k+1 = m+5$. This ensures $n+k+1$ is even. Consider the class C of $(n+k+1)$–manifolds with quadratic form q of discriminant -1. We can show that q must be rationally equivalent to a quadratic form $f = \langle a,b,c\rangle \oplus q_m'$ (or $f = \langle a,b,c,1\rangle \oplus q_m'$ if m is odd) by the same argument used to prove Lemma 2.7, with q_m' in the place of $\langle 1,-1\rangle$. Then $\mathbb{H}^{n+k+1}/\mathrm{SO}_0(f,\mathbb{Z})$ lies in C and is commensurable to a manifold with a cusp of type $B \times (S^1)^k$.

When $n+k+1$ is odd, we cannot control the discriminant of the quadratic form q associated to C. However, we can take a rank-$(n+k)$ subform q' of q such that $q = q' \oplus \langle x\rangle$ for some positive integer x. Then we can scale q' by y so that it has discriminant -1, and, as in the paragraph above, q' is rationally equivalent to $f = \langle a,b,c\rangle \oplus q_m'$ or $f = \langle a,b,c,1\rangle \oplus q_m'$. But now $yq = yq' \oplus \langle xy\rangle$ is rationally equivalent to $f \oplus \langle xy\rangle$, and we can conclude that $\mathbb{H}^{n+k+1}/\mathrm{SO}_0(f \oplus \langle xy\rangle)$ lies in C and is commensurable to a manifold with a cusp of type $B \times (S^1)^k$, as before. □

Corollary 7.5 *Every commensurability class C of cusped arithmetic hyperbolic 8–manifolds contains a manifold with a cusp of type $B \times (S^1)^3$, where B is any compact flat 3–manifold.*

Proof According to Theorem 5.1, every B occurs in the commensurability class of $\mathbb{H}^4/\mathrm{SO}(\langle 1,1,1,1,-1\rangle,\mathbb{Z})$. The result follows from the third paragraph of the proof of Theorem 7.4, using $m = 3$. □

7.1 Fields of coefficients of unipotent matrices

In proving Proposition 7.1, we used the fact that, for a unipotent matrix M, the field of coefficients $F(M)$, defined to be the number field obtained by adjoining the entries of M to $\mathbb{Q}$, is unchanged under powers of M. We prove this result here:

Lemma 7.6 *For any unipotent matrix M and any positive integer k, $F(M) = F(M^k)$.*

Proof Because the entries of M^k are polynomial in the entries of M, $F(M^k) \subset F(M)$. This holds for any M, so in particular, $F(M^{ak}) \subset F(M^k)$ for any nonnegative integer a. We will show that M can be written as a linear combination over $\mathbb{Q}$ of matrices M^{ak}, and thus that each entry in M is polynomial in entries of M^k. This will suffice to show $F(M) \subset F(M^k)$.

By definition, a unipotent matrix M can be written as $M = I + T$, where T is a nilpotent matrix. There is a positive integer l such that $T^l = 0$. Now we can expand $M^k = (I+T)^k$ using binomial coefficients:

$$M^k = \sum_{i=0}^{k} \binom{k}{i} T^i = \sum_{i=0}^{l-1} \binom{k}{i} T^i.$$

Consider the vector space V over $\mathbb{Q}$ consisting of the matrices spanned by all T^i for nonnegative integers i. V must have dimension at most l, since only l of the T^i are nonzero. We will show that if $T^l = 0$ then the $l+1$ matrices M^{ak} for $a \in \{0, 1, \dots, l\}$ span V. Since $M \in V$, this will show that M is a linear combination of these M^{ak}.

Choose some $n \in \mathbb{Z}^+$, and consider the linear combination of matrices M^{ak}

$$\begin{aligned}
\sum_{a=0}^{n} (-1)^{n+a} \binom{n}{a} M^{ak} &= \sum_{a=0}^{n} (-1)^{n+a} \binom{n}{a} \left[\sum_{b=0}^{ak} \binom{ak}{b} T^b \right] \\
&= \sum_{a=0}^{n} \sum_{b=0}^{ak} (-1)^{n+a} \binom{n}{a} \binom{ak}{b} T^b \\
&= \sum_{b=0}^{nk} \sum_{a=\lceil b/k \rceil}^{n} (-1)^{n+a} \binom{n}{a} \binom{ak}{b} T^b \\
&= \sum_{b=0}^{nk} (-1)^n \left[\sum_{a=0}^{n} (-1)^a \binom{n}{a} \binom{ak}{b} \right] T^b.
\end{aligned}$$

Note that when we interchange the summations in line three, we see that a is indexed from $\lceil b/k \rceil$ to n. However, when $ak < b$, $\binom{ak}{b} = 0$ anyway, so we can start a at 0 in line four to get the same value.

The coefficient of T^b in this sum is given by $\sum_{a=0}^{n} (-1)^{n+a} \binom{n}{a} \binom{ak}{b}$. Note that for fixed b, $\binom{t}{b}$ is a degree-b polynomial in t, defined over all nonnegative integers t. When

$b < n$, the coefficient of T^b is 0; we apply Lemma 7.7, proven below, with $f(t) = \binom{t}{b}$ and $y = k$. Since the function $g_f^{n,k}(x)$ is uniformly 0, it is 0 at $x = 0$ in particular. Furthermore, when $b = n$, $\binom{t}{b}$ is a degree-n polynomial, $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$. The coefficient of T^b then must be $a_n n!(-k)^n \neq 0$, by Lemma 7.7.

Now we can use induction on i to construct each T^i as a linear combination of M^{ak}. For the base case, consider $i = l-1$. Choose $n = l-1$, and in the above summation, T^b has coefficient 0 when $b < n = l-1$, and $T^b = 0$ when $b > n$ because $b \geq l$. Thus we've obtained a rational multiple of T^{l-1}, which we can rescale to write T^{l-1} as a linear combination of M^{ak}.

For the induction step, assume T^j can be written as such a linear combination for all $i < j \leq l-1$. Consider the linear combination above with $n = i$. Then, by Lemma 7.7, the coefficients of T^b are 0 for $b < i$, nonzero for $b = i$, and $T^b = 0$ for $b \geq l$. Since T^b can already be written as a linear combination for $i < b \leq l-1$ by the induction hypothesis we can subtract out the appropriate linear combinations to leave only a multiple of T^i.

This suffices to show that every T^i is a linear combination of M^{ak}, and thus $M = T^0 + T^1$ is some linear combination of matrices M^{ak}. Since $F(M^{ak}) \subset F(M^k)$ for all a and we have already proven $F(M^k) \subset F(M)$, we conclude $F(M) = F(M^k)$. □

Finally, we prove here the technical result that allowed us to conclude certain coefficients were zero or nonzero:

Lemma 7.7 *Let $f:\mathbb{R} \to \mathbb{R}$ be a function, and fix $y \in \mathbb{R}$ and $n \in \mathbb{Z}^+$. Let*

$$g_f^{n,y}(x) = \sum_{a=0}^{n} (-1)^a f(x+ay)\binom{n}{a}.$$

If f is a polynomial of degree less than n, then $g_f^{n,y} = 0$ uniformly. Furthermore, if $f(x) = x^n$, then $g_f^{n,y}$ is the constant function $n!(-y)^n$.

Proof First, we prove that $g_f^{n,y} = 0$ when f is a polynomial of degree less than n by induction on n. For the base case, consider $n = 1$. In order for f to be a polynomial of degree less than 1 it must be a constant function $f(x) = c$. Then

$$g_f^{1,y}(x) = \sum_{a=0}^{1} (-1)^a f(x+ay)\binom{1}{a} = f(x) - f(x+y) = c - c = 0.$$

Now assume the statement holds for $n-1$. We can compute

$$
\begin{aligned}
g_f^{n,y}(x) &= \sum_{a=0}^{n}(-1)^a f(x+ay)\binom{n}{a}\\
&= \sum_{a=1}^{n}(-1)^a f(x+ay)\binom{n-1}{a-1}+\sum_{a=0}^{n-1}(-1)^a f(x+ay)\binom{n-1}{a}\\
&= -\sum_{a=0}^{n-1}(-1)^a f(x+y+ay)\binom{n-1}{a}+\sum_{a=0}^{n-1}(-1)^a f(x+ay)\binom{n-1}{a}\\
&= -g_f^{n-1,y}(x+y)+g_f^{n-1,y}(x)=\int_{x+y}^{x}\frac{\partial}{\partial t}[g_f^{n-1,y}(t)]\,dt=\int_{x+y}^{x} g_{f'}^{n-1,y}(t)\,dt
\end{aligned}
$$

The second line above follows from the identity $\binom{n}{a}=\binom{n-1}{a-1}+\binom{n-1}{a}$. The final equality follows from the fact that $g_f^{n,y}$ is a particular linear combination of $f(x+ay)$, with fixed coefficients depending on n; concisely, g is linear in f. Since f is a polynomial of degree less than n, its derivative f' is a polynomial of degree less than $n-1$. Thus, $g_{f'}^{n-1,y}(t)=0$ everywhere by induction, and therefore $g_f^{n,y}=0$.

Next, we prove that $g_f^{n,y}=n!(-y)^n$ for $f=x^n$ by induction on n. For the base case, consider $n=1$. Then

$$g_f^{1,y}(x)=\sum_{a=0}^{1}(-1)^a(x+ay)\binom{1}{a}=(x)-(x+y)=-y.$$

Now assume the statement holds for $n-1$. Let $f(x)=x^n$ and $h(x)=x^{n-1}$, so that $f'=nh$. Then

$$
\begin{aligned}
g_f^{n,y}(x) &= \int_{x+y}^{x} g_{f'}^{n-1,y}(t)\,dt = n\int_{x+y}^{x} g_h^{n-1,y}(t)\,dt\\
&= n\int_{x+y}^{x}(n-1)!(-y)^{n-1}\,dt = n!(-y)^n.
\end{aligned}
$$

We proved the first equality in the first part of this proof. The rest follows from the fact that $g_f^{n,y}$ is linear in f, and the induction hypothesis. □

8 Commensurability classes of nonarithmetic manifolds

We can turn arithmetic commensurability classes that avoid certain cusp types into nonarithmetic ones by "inbreeding" the arithmetic manifolds with themselves, in a manner introduced by Agol [1]. We mimic the argument in [1] to construct a manifold with arbitrarily short geodesic, which must be nonarithmetic by Proposition 3.6. Further, this nonarithmetic group is constructed in such a way that it still lies in the $\mathbb{Q}$–points of the original quadratic form, so we can conclude by the same argument as our proof of Proposition 6.1 that it avoids the same cusps. Since this construction can be performed on any of the infinitely many classes that avoid the $\frac{1}{3}$–twist, $\frac{1}{4}$–twist, and $\frac{1}{6}$–twist cusps, there are infinitely many nonarithmetic commensurability classes that avoid such cusps.

Proof of Theorem 1.2 Let q be a quadratic form such that the commensurability class of $\mathbb{H}^4/\mathrm{SO}_0(q,\mathbb{Z})$ does not contain any manifolds or orbifolds with a certain cusp B. Let M be any manifold in this commensurability class, and Γ its fundamental group. By [5, Theorem 4.2], there exist infinitely many closed totally geodesic hyperbolic 3–manifolds immersed in M. These 3–manifolds lift to copies of $\mathbb{H}^3$ in $\mathbb{H}^4$; pick one such copy and call it P. Since the immersed 3–manifold is compact, $H = \mathrm{Isom}(P) \cap \Gamma$ acts cocompactly on P.

By Margulis' commensurability criterion for arithmeticity [19, Theorem 16.3.3], since Γ is arithmetic, its commensurator $\mathrm{Comm}(\Gamma)$ contains $\mathrm{PO}(q,\mathbb{Q})$. Thus for any $\epsilon > 0$, we can choose $\gamma \in \mathrm{Comm}(\Gamma)$ such that $\gamma(P)$ is disjoint from P and the distance $d(P,\gamma(P))$ is less than $\frac{1}{2}\epsilon$. Since $\gamma \in \mathrm{Comm}(\Gamma)$, the stabilizer of $\gamma(P)$, namely $(\gamma H\gamma^{-1}) \cap \Gamma$, acts cocompactly on $\gamma(P)$. Then $H_\gamma = \mathrm{Isom}(\gamma(P)) \cap \Gamma$ must act cocompactly on $\gamma(P)$, since $(\gamma H\gamma^{-1}) \cap \Gamma < H_\gamma$.

Let g be the geodesic segment orthogonal to both P and $\gamma(P)$ intersecting P at p_1 and $\gamma(P)$ at p_2. Because H is discrete and residually finite, as a finitely generated linear group we can choose a finite-index subgroup $H_1 < H$ such that $d(p_1, h(p_1)) > 2\,\mathrm{arctanh}\big(\mathrm{sech}\big(\frac{1}{4}\epsilon\big)\big)$ for all nonidentity $h \in H_1$. Similarly, choose $H_2 < H_\gamma$ such that $d(p_2, h(p_2)) > 2\,\mathrm{arctanh}\big(\mathrm{sech}\big(\frac{1}{4}\epsilon\big)\big)$ for all nonidentity $h \in H_2$. Let $\Sigma_1 = P/H_1$ and $\Sigma_2 = \gamma(P)/H_2$. Let $E_i \subset \mathbb{H}^4$ be the Dirichlet domain of H_i centered at p_i.

Now, $U = \Sigma_1 \cup_{p_1} g \cup_{p_2} \Sigma_2$ is an embedded compact spine for $E_1 \cap E_2$, with one component of $\mathbb{H}^4 - P$ retracting to Σ_1, the opposite component of $\mathbb{H}^4 - \gamma(P)$ retracting to Σ_2, and the space in between P and $\gamma(P)$ retracting to g.

We claim $G := \langle H_1, H_2\rangle = H_1 * H_2$ and G is geometrically finite, and defer the proof to Lemma 8.1.

Then G is separable in Γ [2]. By Scott's separability criterion [24], for some finite index subgroup $\Gamma_1 < \Gamma$, U embeds in $\mathbb{H}^4/\Gamma_1$. Thus, Σ_1 and Σ_2 embed in $\mathbb{H}^4/\Gamma_1$. Now let $N = (\mathbb{H}^4/\Gamma_1) - (\Sigma_1 \cup \Sigma_2)$, and D be the double of N along its boundary.

D is a hyperbolic manifold, since N is a hyperbolic manifold with totally geodesic boundary. Note that the double of g is a closed geodesic of length bounded by ϵ, since g is perpendicular to Σ_1 and Σ_2. Through choice of ϵ, we can construct D so that it has a geodesic of arbitrarily small length. Thus, by Proposition 3.6, we can construct D to be nonarithmetic.

Next, we claim that $\pi_1(D) < \mathrm{SO}_0(q, \mathbb{Q})$. First, note that the universal cover of N is $\mathbb{H}^4$ with some half-spaces removed, with its group action given by Γ_1. By construction, $\Gamma_1 < \Gamma < \mathrm{SO}_0(q, \mathbb{Q})$. Thus, we can find a fundamental domain S for N such that all the face pairings of S lie in $\mathrm{SO}_0(q, \mathbb{Q})$. We can construct a fundamental domain for D by taking two copies of S glued together at one of the boundary faces F that lifts to Σ_1, and pairing the remaining boundary faces by mapping each to its counterpart in the other copy of S. We will show that $\pi_1(D) < \mathrm{SO}_0(q, \mathbb{Q})$ by showing that these face pairings, which generate $\pi_1(D)$, each lie in $\mathrm{SO}_0(q, \mathbb{Q})$.

By construction, the face pairings ϕ_i on the original copy of S must lie in $\mathrm{SO}_0(q, \mathbb{Q})$. The corresponding face pairings in the other copy of S are given by $r_P \phi_i r_P$, where r_P is reflection across P. Recall that P was constructed as a hyperplane perpendicular to some $v \in \mathbb{Q}^5$, so the reflection r_P across P lies in $\mathrm{SO}_0(q, \mathbb{Q})$. Thus each $r_P \phi_i r_P$ must also lie in $\mathrm{SO}_0(q, \mathbb{Q})$.

The remaining face pairings are the new ones formed from identifying boundary components of N. To pair a boundary component C with its corresponding mirror component, we can use the isometry $r_P r_F$, where r_F is the reflection across the hyperplane F containing C. Note that F must be the image of $\gamma(P)$ under some isometry $\alpha \in \pi_1(N)$, so $r_F = \alpha^{-1} r_{\gamma(P)} \alpha = \alpha^{-1} \gamma^{-1} r_P \gamma \alpha$. Since we chose γ to lie in $\mathrm{SO}_0(q, \mathbb{Q})$ and α must be an element of $\pi_1(N)$, r_F lies in $\mathrm{SO}_0(q, \mathbb{Q})$ as well. Now $r_F r_P$ lies in $\mathrm{SO}_0(q, \mathbb{Q})$, and thus every face pairing does as well. Therefore, $\pi_1(D)$ is generated by elements of $\mathrm{SO}_0(q, \mathbb{Q})$, and so $\pi_1(D) < \mathrm{SO}_0(q, \mathbb{Q})$.

Now, if we choose the quadratic form q in such a way that the commensurability class of $\mathbb{H}^4/\mathrm{SO}_0(q, \mathbb{Q})$ avoids cusps with cross-section B, then D cannot have cusps with cross-section B, using the same argument as in the proof of Proposition 6.1. In this way, we use Theorem 1.1 to construct infinitely many commensurability classes of nonarithmetic manifolds that avoid the $\frac{1}{3}$–twist, $\frac{1}{4}$–twist, and $\frac{1}{6}$–twist. $\square$

The same proof can be applied to provide examples of commensurability classes of nonarithmetic hyperbolic 5–manifolds that avoid certain cusp types, with Theorem 7.3. We finish the proof by proving the claim we deferred:

Lemma 8.1 *Let H_1 and H_2 be as above. Then $G = \langle H_1, H_2\rangle$ is isomorphic to $H_1 * H_2$, and is geometrically finite.*

Proof As in the proof of Theorem 1.2, we let g be the geodesic segment connecting $p_1 \in P$ with $p_2 \in \gamma(P)$, meeting both planes perpendicularly. Let L be the 3–plane that perpendicularly bisects g, and consider the projections $\mathrm{pr}_1 : \mathbb{H}^4 \to P$ and $\mathrm{pr}_2 : \mathbb{H}^4 \to \gamma(P)$ that map each point in $\mathbb{H}^4$ to the closest point on the target 3–plane. Using hyperbolic geometry (see Theorem 3.5.10 in [23]), we can see that $\mathrm{pr}_i(L)$ is a disk with radius bounded by $\operatorname{arcsinh}\bigl(\operatorname{csch}\bigl(\frac{1}{4}\epsilon\bigr)\bigr) = \operatorname{arctanh}\bigl(\operatorname{sech}\bigl(\frac{1}{4}\epsilon\bigr)\bigr)$ centered at p_i. We defined H_1 so that $d(p_1, h(p_1)) > 2\operatorname{arctanh}\bigl(\operatorname{sech}\bigl(\frac{1}{4}\epsilon\bigr)\bigr)$ for all nonidentity $h \in H_1$, so $\mathrm{pr}_1(L)$ must lie inside E_1, the Dirichlet domain of H_1 centered at p_1. Thus, since $H_1 < \mathrm{Isom}(P)$, L must lie inside of E_1. Similarly, L lies in E_2 as well. Now L splits $\mathbb{H}^4$ into two parts, with ∂E_1 lying in the part with P, and ∂E_2 lying in the part with $\gamma(P)$. Thus $\partial E_1 \cap \partial E_2 = \varnothing$. Since E_1 and E_2 are each geometrically finite, $E_1 \cap E_2$, the fundamental domain of G, is geometrically finite too. Also, note that $E_1 \cap E_2 = E_1 \# E_2$, with the two sets glued along L, so it's a fundamental domain of $H_1 * H_2$. We can conclude that G is geometrically finite and $G = H_1 * H_2$. □

References

[1] **I Agol**, *Systoles of hyperbolic 4–manifolds*, preprint (2006) arXiv

[2] **N Bergeron**, **F Haglund**, **D T Wise**, *Hyperplane sections in arithmetic hyperbolic manifolds*, J. Lond. Math. Soc. 83 (2011) 431–448 MR Zbl

[3] **A Borel**, *Density and maximality of arithmetic subgroups*, J. Reine Angew. Math. 224 (1966) 78–89 MR Zbl

[4] **G Chelnokov**, **A Mednykh**, *On the coverings of Hantzsche–Wendt manifold*, Tohoku Math. J. 74 (2022) 313–327 MR Zbl

[5] **M Chu**, **A W Reid**, *Embedding closed hyperbolic 3–manifolds in small volume hyperbolic 4–manifolds*, Algebr. Geom. Topol. 21 (2021) 2627–2647 MR Zbl

[6] **J H Conway**, **J P Rossetti**, *Describing the platycosms*, preprint (2003) arXiv

[7] **J H Conway**, **N J A Sloane**, *Sphere packings, lattices and groups*, Grundl. Math. Wissen. 290, Springer (1988) MR Zbl

[8] **V Emery**, **J G Ratcliffe**, **S T Tschantz**, *Salem numbers and arithmetic hyperbolic groups*, Trans. Amer. Math. Soc. 372 (2019) 329–355 MR Zbl

[9] **L Ferrari**, **A Kolpakov**, **L Slavich**, *Cusps of hyperbolic 4–manifolds and rational homology spheres*, Proc. Lond. Math. Soc. 123 (2021) 636–648 MR Zbl

[10] **A Kolpakov**, **B Martelli**, *Hyperbolic four-manifolds with one cusp*, Geom. Funct. Anal. 23 (2013) 1903–1933 MR Zbl

[11] **A Kolpakov**, **L Slavich**, *Hyperbolic 4–manifolds, colourings and mutations*, Proc. Lond. Math. Soc. 113 (2016) 163–184 MR Zbl

[12] **D D Long**, **A W Reid**, *On the geometric boundaries of hyperbolic 4–manifolds*, Geom. Topol. 4 (2000) 171–178 MR Zbl

[13] **D D Long**, **A W Reid**, *All flat manifolds are cusps of hyperbolic orbifolds*, Algebr. Geom. Topol. 2 (2002) 285–296 MR Zbl

[14] **C Maclachlan**, **A W Reid**, *The arithmetic of hyperbolic 3–manifolds*, Graduate Texts in Math. 219, Springer (2003) MR Zbl

[15] **D B McReynolds**, *Peripheral separability and cusps of arithmetic hyperbolic orbifolds*, Algebr. Geom. Topol. 4 (2004) 721–755 MR Zbl

[16] **D B McReynolds**, **A W Reid**, **M Stover**, *Collisions at infinity in hyperbolic manifolds*, Math. Proc. Cambridge Philos. Soc. 155 (2013) 459–463 MR Zbl

[17] **J S Meyer**, *Totally geodesic spectra of arithmetic hyperbolic spaces*, Trans. Amer. Math. Soc. 369 (2017) 7549–7588 MR Zbl

[18] **J M Montesinos-Amilibia**, *On odd rank integral quadratic forms: canonical representatives of projective classes and explicit construction of integral classes with square-free determinant*, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. 109 (2015) 199–245 MR Zbl

[19] **D W Morris**, *Introduction to arithmetic groups*, Deductive (2015) MR Zbl

[20] **B E Nimershiem**, *All flat three-manifolds appear as cusps of hyperbolic four-manifolds*, Topology Appl. 90 (1998) 109–133 MR Zbl

[21] **F Pfäffle**, *The Dirac spectrum of Bieberbach manifolds*, J. Geom. Phys. 35 (2000) 367–385 MR Zbl

[22] **PyRulez**, *Mapping an element of the euclidean group to an isometry of a horosphere in the hyperboloid model*, Mathematics Stack Exchange thread (2020) Available at `https://math.stackexchange.com/q/3812423`

[23] **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, Graduate Texts in Math. 149, Springer (1994) MR Zbl

[24] **P Scott**, *Subgroups of surface groups are almost geometric*, J. London Math. Soc. 17 (1978) 555–565 MR Zbl

[25] **J-P Serre**, *A course in arithmetic*, Graduate Texts in Math. 7, Springer (1973) MR Zbl

Department of Mathematics, Rice University
Houston, TX, United States

csell@rice.edu

Received: 11 October 2021 Revised: 17 May 2022

Algebraic & Geometric Topology **23**:8 (2023) 3835–3847
DOI: 10.2140/agt.2023.23.3835
Published: 5 November 2023

The group of quasi-isometries of the real line cannot act effectively on the line

Shengkui Ye
Yanxin Zhao

We prove that the group $\mathrm{QI}^+(\mathbb{R})$ of orientation-preserving quasi-isometries of the real line is a left-orderable, nonsimple group, which cannot act effectively on the real line $\mathbb{R}$.

20F65

1 Introduction

A function $f : X \to Y$ between metric spaces X and Y is a quasi-isometry if there exist real numbers $K \geq 1$ and $C \geq 0$ such that

$$\frac{1}{K} d(x_1, x_2) - C \leq d(f(x_1), f(x_2)) \leq K d(x_1, x_2) + C$$

for any $x_1, x_2 \in X$, and $d(\mathrm{Im}\, f, y) \leq C$ for any $y \in Y$. Two quasi-isometries f and g are called equivalent if they are of bounded distance; ie $\sup_{x \in X} d(f(x), g(x)) < \infty$. The quasi-isometry group $\mathrm{QI}(X)$ is the group of all equivalence classes $[f]$ of quasi-isometries $f : X \to X$ under composition. The notion of quasi-isometries is one of the fundamental concepts in geometric group theory. In this note, we consider the quasi-isometry group $\mathrm{QI}(\mathbb{R})$ of the real line. Gromov and Pansu [3, Section 3.3B] noted that the group of bi-Lipschitz homeomorphisms has a full image in $\mathrm{QI}(\mathbb{R})$. Sankaran [9] proved that the orientation-preserving subgroup $\mathrm{QI}^+(\mathbb{R})$ is torsion-free and many large groups, like Thompson groups and free groups of infinite rank, can be embedded into $\mathrm{QI}^+(\mathbb{R})$.

Recall that a group G is left-orderable if there is a total order $\leq$ on G such that $g \leq h$ implies $fg \leq fh$ for any $f \in G$. We will prove the following.

Theorem 1.1 *The quasi-isometry group* $\mathrm{QI}^+(\mathbb{R})$ — *or* $\mathrm{QI}([0, +\infty))$ — *is not simple.*

Theorem 1.2 *The quasi-isometry group* $\mathrm{QI}^+(\mathbb{R})$ *— or* $\mathrm{QI}([0,+\infty))$ *— is left-orderable.*

Theorem 1.3 *The quasi-isometry group* $\mathrm{QI}^+(\mathbb{R})$ *cannot act effectively on the real line* $\mathbb{R}$.

Other (uncountable) left-orderable groups that cannot act on the line are been known. For example, the germ group $\mathcal{G}_\infty(\mathbb{R})$, due to Mann [4] and Rivas; and the compact supported diffeomorphism group $\mathrm{Diff}_c(\mathbb{R}^n)$ for $n>1$, due to Chen and Mann [1].

2 The group structure of QI(ℝ)

Let $\mathrm{QI}(\mathbb{R}_+)$ (resp. $\mathrm{QI}(\mathbb{R}_-)$) be the quasi-isometry group of the ray $[0,+\infty)$ (resp. $(-\infty,0]$), viewed as subgroup of $\mathrm{QI}(\mathbb{R})$ fixing the negative (resp. positive) part.

Lemma 2.1 $\mathrm{QI}(\mathbb{R})=(\mathrm{QI}(\mathbb{R}_+)\times\mathrm{QI}(\mathbb{R}_-))\rtimes\langle t\rangle$, *where* $t\in\mathrm{QI}(\mathbb{R})$ *is the reflection* $t(x)=-x$ *for any* $x\in\mathbb{R}$.

Proof Sankaran [9] proves that the group $\mathrm{PL}_\delta(\mathbb{R})$ consisting of piecewise linear homeomorphisms with bounded slopes has a full image in $\mathrm{QI}(\mathbb{R})$. Since every homeomorphism $f\in\mathrm{PL}_\delta(\mathbb{R})$ is of bounded distance to the map $f-f(0)\in\mathrm{PL}_\delta(\mathbb{R})$, we see that the subgroup

$$\mathrm{PL}_{\delta,0}(\mathbb{R})=\{f\in\mathrm{PL}_\delta(\mathbb{R})\mid f(0)=0\}$$

also has full image in $\mathrm{QI}(\mathbb{R})$. Let

$$\mathrm{PL}_{\delta,+}(\mathbb{R})=\{f\in\mathrm{PL}_\delta(\mathbb{R})\mid f(x)=x,\ x\le 0\},$$
$$\mathrm{PL}_{\delta,-}(\mathbb{R})=\{f\in\mathrm{PL}_\delta(\mathbb{R})\mid f(x)=x,\ x\ge 0\}.$$

Since $\mathrm{PL}_{\delta,+}(\mathbb{R})\cap\mathrm{PL}_{\delta,-}(\mathbb{R})=\{\mathrm{id}_\mathbb{R}\}$, we see that $\mathrm{PL}_{\delta,+}(\mathbb{R})\times\mathrm{PL}_{\delta,-}(\mathbb{R})$ has a full image in $\mathrm{QI}^+(\mathbb{R})$, the orientation-preserving subgroup of $\mathrm{QI}(\mathbb{R})$. It's obvious that $\mathrm{PL}_{\delta,+}(\mathbb{R})$ (resp. $\mathrm{PL}_{\delta,-}(\mathbb{R})$) has a full image in $\mathrm{QI}(\mathbb{R}_+)$ (resp. $\mathrm{QI}(\mathbb{R}_-)$). Therefore, $\mathrm{QI}(\mathbb{R})=(\mathrm{QI}(\mathbb{R}_+)\times\mathrm{QI}(\mathbb{R}_-))\rtimes\langle t\rangle$. □

Let $\mathrm{Homeo}_+(\mathbb{R})$ be the group of orientation-preserving homeomorphisms of the real line. Two functions $f,g\in\mathrm{Homeo}_+(\mathbb{R})$ are of bounded distance if

$$\sup_{|x|\ge M}|f(x)-g(x)|<\infty$$

for a sufficiently large real number M. This means when we study elements $[f]$ in $\mathrm{QI}(\mathbb{R})$, we don't need to care too much about the function values $f(x)$ for x with small absolute values. We will implicitly use this fact in the following context. As $\mathrm{PL}_\delta(\mathbb{R})$ has a full image in $\mathrm{QI}(\mathbb{R})$ (by Sankaran [9]), we take representatives of quasi-isometries which are homeomorphisms in the rest of the article.

2.1 $\mathrm{QI}(\mathbb{R}_+)$ is not simple

Let $\mathrm{QI}(\mathbb{R}_+)$ be the quasi-isometry group of the half-line $[0,+\infty)$. Note that the quasi-isometry group $\mathrm{QI}^+(\mathbb{R}) = \mathrm{QI}(\mathbb{R}_+)\times\mathrm{QI}(\mathbb{R}_-)$ and $\mathrm{QI}(\mathbb{R}_+)\cong\mathrm{QI}(\mathbb{R}_-)$, by Lemma 2.1. Let $H=\{[f]\in\mathrm{QI}(\mathbb{R}_+)\mid \lim_{x\to\infty}(f(x)-x)/x=0\}$. Theorem 1.1 follows from the following theorem.

Theorem 2.2 *H is a proper normal subgroup of* $\mathrm{QI}(\mathbb{R}_+)$. *In particular,* $\mathrm{QI}(\mathbb{R}_+)$ *is not simple.*

Proof For any $[f],[g]\in H$,

$$\frac{f(g(x))-x}{x}=\frac{f(g(x))-g(x)}{g(x)}\frac{g(x)}{x}+\frac{g(x)-x}{x}.$$

Since g is a quasi-isometry, we know that $(1/K)x-C\le g(x)-g(0)\le Kx+C$. Therefore, $1/K-1\le g(x)/x\le K+1$ for sufficiently large x. When $x\to\infty$, we have $g(x)\to\infty$. This means $(f(g(x))-g(x))/g(x)\to 0$. Therefore, $(f(g(x))-x)/x\to 0$ as $x\to\infty$. This proves that $[fg]\in H$.

Note that

$$\frac{|f^{-1}(x)-x|}{x}=\frac{|f^{-1}(x)-f^{-1}(f(x))|}{x}\le\frac{K|x-f(x)|+C}{x}.$$

Therefore,

$$\lim_{x\to\infty}\frac{|f^{-1}(x)-x|}{x}=0.$$

This means $[f^{-1}]\in H$ and that H is a subgroup.

For any quasi-isometric homeomorphism $g\in\mathrm{Homeo}(\mathbb{R}_+)$ and any $[f]\in H$,

$$\begin{aligned}\frac{g^{-1}(f(g(x)))-x}{x}&=\frac{g^{-1}(f(g(x)))-g^{-1}(g(x))}{x}\\&=\frac{g^{-1}(f(g(x)))-g^{-1}(g(x))}{g(x)}\frac{g(x)}{x}.\end{aligned}$$

Note that when $x\to\infty$, the function $g(x)/x$ is bounded. Let $y=g(x)$. We have

$$\frac{|g^{-1}(f(y))-g^{-1}(y)|}{y}\le\frac{K|f(y)-y|+C}{y}\to 0,\quad x\to\infty.$$

Therefore, $[g^{-1}fg]\in H$.

It's obvious that the function f defined by $f(x) = 2x$ is not an element in H. The function defined by $g(x) = x + \ln(x+1)$ gives a nontrivial element in H. Thus H is a proper normal subgroup of $\mathrm{QI}(\mathbb{R}_+)$. □

Lemma 2.3 *Let*

$$W(\mathbb{R}) = \Bigl\{ f \in \mathrm{Diff}(\mathbb{R}) \mid \sup_{x\in\mathbb{R}} |f(x) - x| < \infty,\ \sup_{x\in\mathbb{R}} |f'(x)| < \infty \Bigr\}$$

be the group consisting of diffeomorphisms with bounded derivatives and of bounded distance from the identity. Define a homeomorphism $h \colon \mathbb{R} \to \mathbb{R}$ by $h(x) = e^x$ when $x \geq 1$, $h(x) = -h(-x)$ when $x \leq -1$, and $h(x) = ex$ when $-1 \leq x \leq 1$. Then hfh^{-1} is a quasi-isometry for any $f \in W(\mathbb{R})$.

Proof For any $f \in W(\mathbb{R})$ and sufficiently large $x > 0$, its derivative satisfies that

$$\begin{aligned} |hfh^{-1}(x)'| &= |(e^{f(\ln x)})'| \\ &= |(xe^{f(\ln x) - \ln x})'| \\ &= |e^{f(\ln x) - \ln x}(1 + f'(\ln x) - 1)| \\ &= |e^{f(\ln x) - \ln x} f'(\ln x)| \\ &\leq e^{\sup_{x\in\mathbb{R}} |f(x) - x|} \cdot \sup_{x\in\mathbb{R}} |f'(x)|. \end{aligned}$$

The case for negative $x < 0$ can be calculated similarly. This proves that hfh^{-1} is a quasi-isometry. □

The following result was proved by Sankaran [9].

Corollary 2.4 *The quasi-isometry group* $\mathrm{QI}(\mathbb{R})$ *contains* $\mathrm{Diff}_{\mathbb{Z}}(\mathbb{R})$ (*the lift of* $\mathrm{Diff}(S^1)$ *to* $\mathrm{Homeo}(\mathbb{R})$).

Proof For any $f \in \mathrm{Diff}_{\mathbb{Z}}(\mathbb{R})$, we have $f(x+1) = f(x) + 1$ for any $x \in \mathbb{R}$. This means $\sup_{x\in\mathbb{R}} |f(x) - x| < +\infty$. Since $f(x) - x$ is periodic, we know that $f'(x)$ is bounded. Suppose that $f(x) > x$ for some $x \in [0, 1]$. Take $y_n = e^{x+n}$ for $n > 0$. Let h be the function defined in Lemma 2.3. We have

$$|hfh^{-1}(y_n) - y_n| = |e^{f(x+n)} - e^{x+n}| = |e^{f(x)} - e^x| e^n \to \infty,$$

which means $[hfh^{-1}] \neq [\mathrm{id}] \in \mathrm{QI}(\mathbb{R})$. □

Lemma 2.5 $\mathrm{QI}(\mathbb{R})$ *contains the semidirect product* $\mathrm{Diff}_{\mathbb{Z}}(\mathbb{R}) \ltimes H$.

Proof Since H is normal, it's enough to prove that $\mathrm{Diff}_{\mathbb{Z}}(\mathbb{R}) \cap H = \{e\}$, the trivial subgroup. Actually, for any $f \in \mathrm{Diff}_{\mathbb{Z}}(\mathbb{R})$, the conjugate hfh^{-1} is a quasi-isometry as in the proof of Corollary 2.4. If $hfh^{-1} \in H$, then

$$\lim_{x\to\infty} \frac{hfh^{-1}(x)}{x} = \lim_{x\to\infty} \frac{xe^{f(\ln x)-\ln x}}{x} = \lim_{x\to\infty} e^{f(\ln x)-\ln x} = 1.$$

Since $f(x)-x$ is periodic, we know that $f(\ln x) = \ln x$ for any sufficiently large x. But this means that $f(y) = y$ for any y, so f is the identity. □

2.2 Affine subgroups of QI($\mathbb{R}$)

Lemma 2.6 *The quasi-isometry group* $\mathrm{QI}(\mathbb{R}_+)$ (*actually, the semidirect product* $\mathrm{Diff}_{\mathbb{Z}}(\mathbb{R}) \ltimes H$) *contains the semidirect product* $\mathbb{R}_{>0} \ltimes \big(\bigoplus_{i\in\mathbb{R}_{\ge 1}} \mathbb{R}\big)$, *generated by* A_t *and* $B_{i,s}$ *for* $t \in \mathbb{R}_{>0}$, $i \in \mathbb{R}_{\ge 1} = [1,\infty)$ *and* $s \in \mathbb{R}$ *satisfying*

$$A_t B_{i,s} A_t^{-1} = B_{i,st^{\frac{i}{i+1}}}, \quad B_{i,s_1} B_{i,s_2} = B_{i,s_1+s_2},$$
$$A_{t_1} A_{t_2} = A_{t_1 t_2}, \qquad B_{i,s_1} B_{j,s_2} = B_{j,s_2} B_{i,s_1},$$

for any $t_1, t_2 \in \mathbb{R}_{>0}$, $i, j \in \mathbb{R}_{\ge 1}$ *and* $s_1, s_2 \in \mathbb{R}$.

Proof Let

$$A_t(x) = tx, \qquad t \in \mathbb{R}_{>0},$$
$$B_{i,s}(x) = x + sx^{\frac{1}{i+1}}, \quad s \in \mathbb{R},$$

for $x \ge 0$. We define $A_t(x) = B_{i,s}(x) = x$ for $x \le 0$. Since the derivatives

$$A_t'(x) = t, \quad B_{i,s}'(x) = 1 + \frac{s}{i+1} x^{\frac{-i}{i+1}}$$

are bounded for sufficiently large x, we know that A_t and $B_{i,s}$ are quasi-isometries. For any $x \ge 1$,

$$A_t B_{i,s} A_t^{-1}(x) = A_t B_{i,s}\Big(\frac{x}{t}\Big) = A_t\Big(\frac{x}{t} + s\Big(\frac{x}{t}\Big)^{\frac{1}{i+1}}\Big) = x + st^{\frac{i}{i+1}} x^{\frac{1}{i+1}} = B_{i,st^{\frac{i}{i+1}}}(x).$$

For any $x \ge 1$,

$$B_{i,s_1} B_{i,s_2}(x) = B_{i,s_1}(x + s_2 x^{\frac{1}{i+1}}) = x + s_2 x^{\frac{1}{i+1}} + s_1 (x + s_2 x^{\frac{1}{i+1}})^{\frac{1}{i+1}}$$

and

$$|B_{i,s_1} B_{s_2}(x) - B_{i,s_1+s_2}(x)| = |s_1((x + s_2 x^{\frac{1}{i+1}})^{\frac{1}{i+1}} - x^{\frac{1}{i+1}})| \le \left| s_1 \frac{s_2 x^{\frac{1}{i+1}}}{x^{\frac{i}{i+1}}} \right| \le |s_1 s_2|$$

by Newton's binomial theorem. This means that $B_{i,s_1} B_{i,s_2}$ and B_{i,s_1+s_2} are of bounded distance. It is obvious that $A_{t_1} A_{t_2} = A_{t_1 t_2}$.

When $i < j$ are distinct natural numbers,

$$\begin{aligned}
&|B_{i,s_1}B_{j,s_2}(x) - B_{j,s_2}B_{i,s_1}(x)| \\
&\qquad = |x + s_2 x^{\frac{1}{j+1}} + s_1(x + s_2 x^{\frac{1}{j+1}})^{\frac{1}{i+1}} - (x + s_1 x^{\frac{1}{i+1}} + s_2(x + s_1 x^{\frac{1}{i+1}})^{\frac{1}{j+1}})| \\
&\qquad = |s_1((x + s_2 x^{\frac{1}{j+1}})^{\frac{1}{i+1}} - x^{\frac{1}{i+1}}) + s_2(x^{\frac{1}{j+1}} - (x + s_1 x^{\frac{1}{i+1}})^{\frac{1}{j+1}})| \\
&\qquad \leq \left| s_1 \frac{s_2 x^{\frac{1}{j+1}}}{x^{\frac{i}{i+1}}} \right| + \left| s_2 \frac{s_1 x^{\frac{1}{i+1}}}{x^{\frac{j}{j+1}}} \right| \\
&\qquad \leq 2|s_1 s_2|
\end{aligned}$$

for any $x \geq 1$. This proves that images $[A_t], [B_{i,s}] \in \mathrm{QI}(\mathbb{R}_{\geq 0})$ satisfy the relations. By abuse of notation, we still denote the classes by the same letters.

We prove that the subgroup generated by $\{B_{i,s} \mid i \in \mathbb{R}_{\geq 1}, s \in \mathbb{R}\}$ is the infinite direct sum $\bigoplus_{i \in \mathbb{R}_{\geq 1}} \mathbb{R}$. It's enough to prove that $B_{i_1,s_1}, B_{i_2,s_2}, \ldots, B_{i_k,s_k}$ are $\mathbb{Z}$–linearly independent for distinct $i_1, i_2, \ldots, i_k$ and nonzero $s_1, s_2, \ldots, s_k \in \mathbb{R}$. This can directly checked. For integers $n_1, n_2, \ldots, n_k$, suppose that $B^{n_1}_{i_1,s_1} \circ B^{n_2}_{i_2,s_2} \circ \cdots \circ B^{n_k}_{i_k,s_k} = \mathrm{id} \in \mathrm{QI}(\mathbb{R}_{\geq 0})$. We have

$$\begin{aligned}
&\sup_{x \in \mathbb{R}_{>0}} |B^{n_1}_{i_1,s_1} \circ B^{n_2}_{i_2,s_2} \circ \cdots \circ B^{n_k}_{i_k,s_k}(x) - x| \\
&\quad = \sup_{x \in \mathbb{R}_{>0}} |n_k s_k x^{\frac{1}{i_k+1}} + n_{k-1}s_{k-1}(x + n_k s_k x^{\frac{1}{i_k+1}})^{\frac{1}{i_{k-1}+1}} + \cdots + n_1 s_1 (x + \cdots)^{\frac{1}{i_1+1}}| \\
&\quad < +\infty,
\end{aligned}$$

which implies $n_1 = n_2 = \cdots = n_k = 0$, considering the exponents.

The subgroup $\mathbb{R}_{>0} \ltimes \left(\bigoplus_{i \in \mathbb{R}_{\geq 1}} \mathbb{R}\right)$ lies in $\mathrm{Diff}_{\mathbb{Z}}(\mathbb{R}) \ltimes H$ by the following construction. Let $a_t, b_{i,s} \colon \mathbb{R} \to \mathbb{R}$ be defined by $a_t(x) = x + \ln t$ and $b_{i,s}(x) = \ln(e^x + s e^{\frac{x}{i+1}})$ for $t \in \mathbb{R}_{>0}$, $i \in \mathbb{R}_{\geq 1}$ and $s \in \mathbb{R}$. It can be directly checked that $a_t \in \mathrm{Diff}_{\mathbb{Z}}(\mathbb{R})$ and $b_{i,s} \in W(\mathbb{R})$ (defined in Lemma 2.3). Let $h(x) = e^x$. A direct calculation shows that $h a_t h^{-1} = A_t$ and $h b_{i,s} h^{-1} = B_{i,s}$, as elements in $\mathrm{QI}(\mathbb{R}_+)$. □

3 Left-orderability

The following is well known; for a proof, see [7, Proposition 1.4]:

Lemma 3.1 *A group G is left-orderable if and only if, for every finite collection of nontrivial elements $g_1, \ldots, g_k$, there exist choices $\varepsilon_i \in \{1, -1\}$ such that the identity is not an element of the semigroup generated by $\{g_i^{\varepsilon_i} \mid i = 1, 2, \ldots, k\}$.*

The proof of Theorem 1.2 follows a similar strategy used by Navas to prove the left-orderability of the group $\mathcal{G}_\infty$ of germs at ∞ of homeomorphisms of $\mathbb{R}$; cf [2, Remark 1.1.13] or [4, Proposition 2.2].

Proof of Theorem 1.2 It's enough to prove that $\mathrm{QI}(\mathbb{R}_+)$ is left-orderable. Let $f_1, f_2, \dots, f_n \in \mathrm{QI}(\mathbb{R}_+)$ be any finitely many nontrivial elements. Note that any $1 \neq [f] \in \mathrm{QI}(\mathbb{R}_+)$ has $\sup_{x>0} |f(x) - x| = \infty$. This property doesn't depend on the choice of $f \in [f]$. Without confusion, we still denote $[f]$ by f. Choose a sequence $\{x_{1,k}\} \subset \mathbb{R}_+$ such that $\sup_{k\in\mathbb{N}} |f_1(x_{1,k}) - x_{1,k}| = \infty$. For each $i > 1$, we have either $\sup_{k\in\mathbb{N}} |f_i(x_{1,k}) - x_{1,k}| = \infty$ or $\sup_{k\in\mathbb{N}} |f_i(x_{1,k}) - x_{1,k}| \le M$ for a real number M. After passing to subsequences, we assume for each $i = 1, 2, \dots, n$ that either $f_i(x_{1,k}) - x_{1,k} \to +\infty$, $f_i(x_{1,k}) - x_{1,k} \to -\infty$ or $\sup_{k\in\mathbb{N}} |f_i(x_{1,k}) - x_{1,k}| \le M$. We assign $\varepsilon_i = 1$ for the first case and $\varepsilon_i = -1$ for the second case. For the third case, let

$$S_1 = \{f_i \mid \sup_{k\in\mathbb{N}} |f_i(x_{1,k}) - x_{1,k}| \le M\}.$$

Note that $f_1 \notin S_1$. Choose $f_{i_0} \in S_1$ if S_1 is not empty. We choose another sequence $\{x_{2,k}\}$ such that $\sup_{k\in\mathbb{N}} |f_{i_0}(x_{2,k}) - x_{2,k}| = \infty$. Similarly, after passing to a subsequence, we have for each $f \in S_1$ that either $f(x_{2,k}) - x_{2,k} \to +\infty$, $f(x_{2,k}) - x_{2,k} \to -\infty$ or $\sup_{k\in\mathbb{N}} |f(x_{2,k}) - x_{2,k}| \le M'$ for another real number M'. Assign $\varepsilon_i = 1$ for the first case and $\varepsilon_i = -1$ for the second case. Continue this process to define $S_2, S_3, \dots$ and choose sequences $\{x_{i,k}\}, i = 3, 4, \dots$ to assign ε_i for each f_i. Note that the process will stop at n times, as the number of elements without assignment is strictly decreasing.

For an element $f \in \mathrm{QI}(\mathbb{R}_+)$ satisfying $f(x_i) - x_i \to \infty$ as $i \to \infty$ for some sequence $\{x_i\}$, we assume that $f(x_i) - x_i > 0$ for each i. Since f and f^{-1} are orientation-preserving,

$$\begin{aligned} f^{-1}(x_i) - x_i &= -(x_i - f^{-1}(x_i)) \\ &= -(f^{-1}(f(x_i)) - f^{-1}(x_i)) \le -\Big(\frac{1}{K}(f(x_i) - x_i) - C\Big) \to -\infty. \end{aligned}$$

Let $w = f_{i_1}^{\varepsilon_{i_1}} \cdots f_{i_m}^{\varepsilon_{i_m}} \in \langle f_1, f_2, \dots, f_n \rangle$ be a nontrivial word. If $\{i_1, \dots, i_m\} \not\subseteq S_1$, we have $w(x_{1,k}) - x_{1,k} \to \infty$. Otherwise, $\sup_{k\in\mathbb{N}} |w(x_{1,k}) - x_{1,k}| < \infty$. Suppose that $\{i_1, \dots, i_m\} \subset S_t$, but $\{i_1, \dots, i_m\} \not\subseteq S_{t+1}$ with the assumption that $S_0 = \{f_1, f_2, \dots, f_n\}$. We have $w(x_{t+1,k}) - x_{t+1,k} \to \infty$ as $k \to \infty$. This proves that $w \neq 1 \in \mathrm{QI}(\mathbb{R}_+)$. Therefore, $\mathrm{QI}(\mathbb{R}_+)$ is left-orderable by Lemma 3.1. □

Lemma 3.2 *The group* $\mathrm{QI}(\mathbb{R}_+)$ *is not locally indicable.*

Proof Note that $\mathrm{QI}(\mathbb{R}_+)$ contains the lift $\widetilde{\Gamma}$ of $\mathrm{PSL}(2,\mathbb{R}) < \mathrm{Diff}(S^1)$ to $\mathrm{Homeo}(\mathbb{R})$ (Corollary 2.4). But this lift $\widetilde{\Gamma}$ contains a subgroup $\Gamma = \langle f, g, h : f^2 = g^3 = h^7 = fgh\rangle$, the lift of the $(2,3,7)$–triangle group. There are no nontrivial maps from Γ to $(\mathbb{R},+)$; for more details see [2, page 94]. □

4 The quasi-isometric group cannot act effectively on the line

The following was proved by Mann [4, Proposition 6].

Lemma 4.1 *Consider the affine group $\mathbb{R}_{>0} \ltimes \mathbb{R}$, generated by A_t and B_s for $t \in \mathbb{R}_{>0}$ and $s \in \mathbb{R}$ satisfying*

$$A_t B_s A_t^{-1} = B_{ts}, \quad B_{s_1} B_{s_2} = B_{s_1+s_2}, \quad A_{t_1} A_{t_2} = A_{t_1 t_2}.$$

The affine group $\mathbb{R}_{>0} \ltimes \mathbb{R}$ cannot act effectively on the real line $\mathbb{R}$ by homeomorphisms with A_t a translation for each t.

Proof Suppose that $\mathbb{R}_{>0} \ltimes \mathbb{R}$ acts effectively on the real line $\mathbb{R}$ with each A_t a translation. After passing to an index-2 subgroup, we assume that the group is orientation-preserving. If B_1 acts freely on $\mathbb{R}$, then it is conjugate to the translation $T\colon \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto x+1$. In such a case, we have $A_2 T A_2^{-1} = T^2$. Therefore, $A_2^{-1}(x+2) = A_2^{-1}(x) + 1$ for any x. Since A_2^{-1} maps intervals of length 2 to an interval of length 1, it is a contracting map, and thus has a fixed point.

If B_1 has a nonempty fixed point set $\mathrm{Fix}(B_1)$, choose I to be a connected component of $\mathbb{R} \setminus \mathrm{Fix}(B_1)$. Suppose that $A_2(x) = x + a$, a translation by some real number $a > 0$. Since $A_2 = A_{2^{1/n}}^n$, we have $A_{2^{1/n}}(x) = x + a/n$ for each positive integer n. For each n, let $F_n = A_{2^{1/n}} B_1 A_{2^{1/n}}^{-1}$. Since $A_{2^{1/n}} B_1 A_{2^{1/n}}^{-1}$ commutes with B_1, we see that $F_n \mathrm{Fix}(B_1) = \mathrm{Fix}(B_1)$. This means that either $F_n(I) = I$ or $F_n(I) \cap I = \varnothing$. Since $F_n(x) = B_1(x - a/n) + a/n$ for any $x \in \mathbb{R}$, we know that $F_n(I) = I$ for sufficiently large n. Without loss of generality, we assume that I is of the form (x,y) or $(-\infty, y)$. Choose a sufficiently large n such that $y - a/n \in I$. We have

$$A_{2^{1/n}} B_1 A_{2^{1/n}}^{-1}(y) = B_1\Big(y - \frac{a}{n}\Big) + \frac{a}{n} \neq y,$$

which is a contradiction to the fact that $F_n(I) = I$. □

Definition 4.2 A topologically diagonal embedding of a group $G < \mathrm{Homeo}(\mathbb{R})$ is a homomorphism $\phi\colon G \to \mathrm{Homeo}_+(\mathbb{R})$ defined as follows. Choose a collection of disjoint open intervals $I_n \subset \mathbb{R}$ and homeomorphisms $f_n\colon \mathbb{R} \to I_n$. Define ϕ by $\phi(g)(x) = f_n g f_n^{-1}(x)$ when $x \in I_n$ and $\phi(g)(x) = x$ when $x \notin I_n$.

The following is similar to a result proved by Militon [6].

Lemma 4.3 (Militon [6]) *Let $\Gamma = \mathrm{PSL}_2(\mathbb{R})$ and $\widetilde{\Gamma} < \mathrm{Homeo}_+(\mathbb{R})$ be the lift of Γ to the real line. Any effective action $\phi\colon \widetilde{\Gamma} \hookrightarrow \mathrm{Homeo}_+(\mathbb{R})$ of $\widetilde{\Gamma}$ on the real line $\mathbb{R}$ is a topological diagonal embedding.*

Proof After passing to an index-2 subgroup, we assume the action is orientation-preserving. Let $\tau\colon \mathbb{R} \to \mathbb{R}$ be the translation $x \mapsto x+1$. Suppose that $\mathrm{Fix}(\phi(\tau)) \neq \varnothing$. Note that τ lies in the center of $\widetilde{\Gamma}$. The quotient group $\Gamma = \widetilde{\Gamma}/\langle\tau\rangle$ acts on the fixed point set $\mathrm{Fix}(\phi(\tau))$. For any $f \in \Gamma$ and $x \in \mathrm{Fix}(\phi(\tau))$, we denote the action by $f(x)$ without confusion. Choose any torsion-element $f \in \Gamma$ and any $x \in \mathrm{Fix}(\phi(\tau))$. We must have $x = f(x)$, for otherwise $x < f(x) < f^2(x) < \cdots < f^k(x)$ for any k. Since Γ is simple, we know that the action of $\widetilde{\Gamma}$ on $\mathrm{Fix}(\tau)$ is trivial. For each connected component $I_i \subset \mathbb{R} \setminus \mathrm{Fix}(\phi(\tau))$, we know that $\tau|_{I_i}$ is conjugate to a translation. The group $\Gamma = \widetilde{\Gamma}/\langle\tau\rangle$ acts on $I_i/\langle\phi(\tau)\rangle = S^1$. A result of Matsumoto [5, Theorem 5.2] says that the group Γ is conjugate to the natural inclusion $\mathrm{PSL}_2(\mathbb{R}) \hookrightarrow \mathrm{Homeo}_+(S^1)$ by a homeomorphism $g \in \mathrm{Homeo}_+(S^1)$. Therefore, the group $\phi(\widetilde{\Gamma})|_{I_i}$ is conjugate to the image of the natural inclusion $\widetilde{\Gamma} \hookrightarrow \mathrm{Homeo}_+(\mathbb{R})$. □

For a real number $a \in \mathbb{R}$, let

$$t_a\colon \mathbb{R} \to \mathbb{R}, \quad x \mapsto x + a$$

be the translation. Denote by $A = \langle t_a : a \in \mathbb{R}\rangle$, the subgroup of translations in the lift $\widetilde{\Gamma}$ of $\mathrm{PSL}_2(\mathbb{R})$.

Corollary 4.4 *For any injective group homomorphism $\phi\colon \widetilde{\Gamma} \to \mathrm{Homeo}(\mathbb{R})$, the image $\phi(A)$ is a continuous one-parameter subgroup; ie $\lim_{a\to a_0} \phi(t_a) = \phi(t_{a_0})$ for any $a_0 \in \mathbb{R}$.*

Proof If ϕ is injective, the previous lemma says that ϕ is a topological diagonal embedding. Therefore, $\phi(A)$ is continuous. □

We will need the following elementary fact.

Lemma 4.5 *Let $\phi\colon (\mathbb{R},+) \to (\mathbb{R},+)$ be a group homomorphism. If ϕ is continuous at any $x \neq 0$, then ϕ is $\mathbb{R}$–linear.*

Proof For any nonzero integer n, we have $\phi(n) = n\phi(1)$ and $\phi(1) = \phi\big(\frac{1}{n}n\big) = n\phi\big(\frac{1}{n}\big)$. Since ϕ is additive, we have $\phi\big(\frac{m}{n}\big) = m\phi\big(\frac{1}{n}\big) = \frac{m}{n}\phi(1)$ for any integers $m, n \neq 0$.

For any nonzero real number $a \in \mathbb{R}$, choose a rational sequence $r_i \to a$. When ϕ is continuous, we have that $\phi(r_i) \to \phi(a)$ and $\phi(r_i) = r_i\phi(1) \to a\phi(1) = \phi(a)$. □

The following is the classical theorem of Hölder: a group acting freely on $\mathbb{R}$ is semiconjugate to a group of translations; see Navas [8, Section 2.2.4].

Lemma 4.6 *Let Γ be a group acting freely on the real line $\mathbb{R}$. There is an injective group homomorphism $\phi\colon \Gamma \to (\mathbb{R},+)$ and a continuous nondecreasing map $\varphi\colon \mathbb{R} \to \mathbb{R}$ such that*

$$\varphi(h(x)) = \varphi(x) + \phi(h)$$

for any $x \in \mathbb{R}$ and $h \in \Gamma$.

Corollary 4.7 *Suppose that the affine group $\mathbb{R}_{>0} \ltimes \mathbb{R} = \langle a_t : t \in \mathbb{R}_{>0}\rangle \ltimes \langle b_s : s \in \mathbb{R}\rangle$ acts on the real line $\mathbb{R}$ by homeomorphisms satisfying*

(1) *the action of the subgroup $\mathbb{R} = \langle b_s : s \in \mathbb{R}\rangle$ is free;*
(2) *for any fixed $x \in \mathbb{R}$, $a_t(x)$ is continuous with respect to $t \in \mathbb{R}_{>0}$.*

Let $\phi\colon \langle b_s : s \in \mathbb{R}\rangle \to (\mathbb{R},+)$ be the additive map in Lemma 4.6 for $\Gamma = \langle b_s : s \in \mathbb{R}\rangle$. Then ϕ is an $\mathbb{R}$–linear map.

Proof Note that $a_t b_s a_t^{-1} = b_{ts}$. We have

$$\varphi(b_{ts}(x)) = \varphi(x) + \phi(b_{ts}).$$

Since $b_{ts}(x) = a_t b_s a_t^{-1}(x) \to b_s(x)$ when $t \to 1$, we have that

$$\varphi(x) + \phi(b_{ts}) \to \varphi(b_s(x)) = \varphi(x) + \phi(b_s).$$

This implies that $\phi(b_{ts}) \to \phi(b_s)$ as $t \to 1$. For any nonzero $x \in \mathbb{R}$ and sequence $x_n \to x$,

$$\phi(b_{x_n}) = \phi(b_{\frac{x_n}{x}x}) \to \phi(b_x).$$

The map ϕ is $\mathbb{R}$–linear by Lemma 4.5. □

Theorem 4.8 *Consider $G = \mathbb{R}_{>0} \ltimes \left(\bigoplus_{i\in\mathbb{R}_{\geq 1}} \mathbb{R}\right)$, generated by A_t and $B_{i,s}$ for $t \in \mathbb{R}_{>0}$, $i \in \mathbb{R}_{\geq 1} = [1,\infty)$ and $s \in \mathbb{R}$ satisfying*

$$A_t B_{i,s} A_t^{-1} = B_{i,st^{\frac{i}{i+1}}}, \qquad B_{i,s_1} B_{i,s_2} = B_{i,s_1+s_2},$$

$$A_{t_1} A_{t_2} = A_{t_1 t_2}, \qquad B_{i,s_1} B_{j,s_2} = B_{j,s_2} B_{i,s_1}$$

for any $t_1, t_2 \in \mathbb{R}_{>0}$, $i, j \in \mathbb{R}_{\geq 1}$ and $s_1, s_2 \in \mathbb{R}$. Then G cannot act effectively on the real line $\mathbb{R}$ by homeomorphisms when the induced action of $\langle A_t : t \in \mathbb{R}_{>0}\rangle$ is a topologically diagonal embedding of the translation subgroup $(\mathbb{R},+) \hookrightarrow \mathrm{Homeo}(\mathbb{R})$.

Proof Suppose that G acts effectively on $\mathbb{R}$ with the induced action of $\langle A_t : t \in \mathbb{R}_{>0} \rangle$, a topologically diagonal embedding of the translation subgroup $(\mathbb{R}, +) \hookrightarrow \mathrm{Homeo}(\mathbb{R})$. Let I be a connected component of $\mathbb{R} \setminus \mathrm{Fix}(\langle A_t, B_{i,s} : t \in \mathbb{R}_{>0}, i = 1, s \in \mathbb{R} \rangle)$.

Suppose that there is an element $B_{1,s}$ having a fixed point $x \in I$ for some $s > 0$. Since $A_4 B_{1,s} A_4^{-1} = B_{1,s}^2$, we know that $A_4 x \in \mathrm{Fix}(B_{1,s}) = \mathrm{Fix}(B_{1,s}^2)$. Since there are no fixed points in I for $\langle A_t, B_{1,s} : t \in \mathbb{R}_{>0}, s \in \mathbb{R} \rangle$, we know that $\lim_{n\to\infty} A_4^n x \notin I$.[1] This implies that A_4 has no fixed point in I. Since the group homomorphism

$$\langle A_t : t \in \mathbb{R}_{>0} \rangle \to \mathrm{Homeo}(\mathbb{R})$$

is a diagonal embedding, we see that each A_t has no fixed point in I and the action of $\langle A_t : t \in \mathbb{R}_{>0} \rangle$ on I is conjugate to a group of translations. By Lemma 4.1, the affine group $\langle A_t, B_{1,s} : t \in \mathbb{R}_{>0}, s \in \mathbb{R} \rangle$ cannot act effectively on I. Suppose that $A_t B_{1,s'}$ acts trivially on I for some $t > 0$ and $s' > 0$. We have that $A_t B_{1,s} = A_{s^2 s'^{-2}} (A_t B_{1,s'}) A_{s^2 s'^{-2}}^{-1}$ acts trivially on I. But $A_t B_{1,s}(x) = A_t(x) = x$ implies that $t = 1$. Therefore, the element $B_{1,s}$ (and any $B_{1,t} = A_{t^2 s^{-2}} B_{1,s} A_{t^2 s^{-2}}^{-1}$ for $t \in \mathbb{R}_{>0}$) acts trivially on I. This means that the action of $\langle B_{1,s} : s \in \mathbb{R} \rangle$ on the connected component I is either trivial or free. Since the action of G is effective, there is a connected component I_1 on which $B_{1,s}$ acts freely. A similar argument shows that $B_{i,s'}$ acts freely on a component I_i for each $i \in \mathbb{R}_{\geq 1}$ and any $s' \in \mathbb{R} \setminus \{0\}$.

Since $B_{i,s'}$ commutes with $B_{j,s}$, we have $B_{i,s'}(I_1) \subset \mathbb{R} \setminus \mathrm{Fix}(\langle B_{j,s} : s \in \mathbb{R} \rangle)$. Moreover, $B_{i,s'}(I_j) \cap I_j$ is either I_j or the empty set. Suppose that $I_i \cap I_j \neq \varnothing$ and the right end b_i of I_i lies in I_j. Choose $x \in I_i \cap I_j$. Note that $B_{j,s}([x, b_i)) \cap [x, b_i) = \varnothing$ for any $s > 0$. This is impossible as $B_{j,s/n}(x) \to x$ as $n \to \infty$. Therefore, $I_i \cap I_j = I_i$ or is empty for distinct $i, j \in \mathbb{R}_{\geq 1}$. Since we have uncountably many $i \in \mathbb{R}_{>0}$, there must be some distinct $i, j \in \mathbb{R}_{\geq 1}$ such that $I_i = I_j$. This means that the subgroup $\mathbb{R} \oplus \mathbb{R}$ spanned by the i, j–components acts freely on I_i. Hölder's theorem (Lemma 4.6) gives an injective group homomorphism $\phi \colon \mathbb{R} \oplus \mathbb{R} \to (\mathbb{R}, +)$ and a continuous nondecreasing map $\varphi \colon \mathbb{R} \to \mathbb{R}$ such that

$$\varphi(h(x)) = \varphi(x) + \phi(h)$$

for any $x \in \mathbb{R}$. Since $\langle A_t : t \in \mathbb{R}_{>0} \rangle \to \mathrm{Homeo}(\mathbb{R})$ is a topological embedding, we have that for any fixed $x \in \mathbb{R}$, $A_t(x)$ is continuous with respect to $t \in \mathbb{R}_{>0}$. By Corollary 4.7,

[1] Otherwise, $\lim_{n\to\infty} A_4^n x \in I$. But $A_t(\lim_{n\to\infty} A_4^n x) = \lim_{n\to\infty} A_4^n x$ for any $t > 0$ by the topologically diagonal embedding. For any s', we have $B_{1,s'} = A_{s'^2 s^{-2}} B_{1,s} A_{s'^2 s^{-2}}^{-1}$ and $B_{1,s'}(\lim_{n\to\infty} A_4^n x) = \lim_{n\to\infty} A_4^n x$. This would imply that $\lim_{n\to\infty} A_4^n x$ is a global fixed point of $\langle A_t, B_{1,s} : t \in \mathbb{R}_{>0}, s \in \mathbb{R} \rangle$.

the restriction map $\phi|_{\mathbb{R}}$ is $\mathbb{R}$–linear for each direct summand $\mathbb{R}$. This is a contradiction to the fact that ϕ is injective. Therefore, the group G cannot act effectively. $\square$

Proof of Theorem 1.3 Suppose that $\mathrm{QI}^+(\mathbb{R})$ acts on the real line by an injective group homomorphism $\phi\colon \mathrm{QI}^+(\mathbb{R}) \to \mathrm{Homeo}(\mathbb{R})$. The group $\mathrm{QI}^+(\mathbb{R})$ contains the semidirect product $\mathbb{R}_{>0} \ltimes \bigl(\bigoplus_{i\in\mathbb{R}_{\geq 1}} \mathbb{R}\bigr)$, by Lemma 2.6. The subgroup $\mathbb{R}_{>0}$ (as the image of the exponential map) is a homomorphic image of the subgroup $\mathbb{R} < \widetilde{\Gamma}$, which is the lift of $\mathrm{SO}(2)/\{\pm I_2\} < \mathrm{PSL}_2(\mathbb{R})$ to $\mathrm{Homeo}(\mathbb{R})$. Note that $\widetilde{\Gamma}$ is embedded into $\mathrm{QI}^+(\mathbb{R})$ (see Corollary 2.4 and its proof). By Lemma 4.3, any effective action of $\widetilde{\Gamma}$ on the real line $\mathbb{R}$ is a topological diagonal embedding. This means that the action of $\mathbb{R}_{>0}$ is a topological diagonal embedding (Corollary 4.4). Theorem 4.8 shows that the action of $\mathbb{R}_{>0} \ltimes \bigl(\bigoplus_{i\in\mathbb{R}_{\geq 1}} \mathbb{R}\bigr)$ is not effective. $\square$

Acknowledgements The authors would like to thank Li Cai and Xiaolei Wu for their helpful discussions. This work is supported by NSFC (grant 11971389).

References

[1] **L Chen**, **K Mann**, *There are no exotic actions of diffeomorphism groups on 1–manifolds*, Groups Geom. Dyn. 17 (2023) 91–99 MR Zbl

[2] **B Deroin**, **A Navas**, **C Rivas**, *Groups*, *orders*, *and dynamics*, preprint (2014) arXiv 1408.5805

[3] **M Gromov**, **P Pansu**, *Rigidity of lattices: An introduction*, from "Geometric topology: Recent developments" (P De Bartolomeis, F Tricerri, editors), Lecture Notes in Math. 1504, Springer (1991) 39–137 MR Zbl

[4] **K Mann**, *Left-orderable groups that don't act on the line*, Math. Z. 280 (2015) 905–918 MR Zbl

[5] **S Matsumoto**, *Numerical invariants for semiconjugacy of homeomorphisms of the circle*, Proc. Amer. Math. Soc. 98 (1986) 163–168 MR Zbl

[6] **E Militon**, *Actions of groups of homeomorphisms on one-manifolds*, Groups Geom. Dyn. 10 (2016) 45–63 MR Zbl

[7] **A Navas**, *On the dynamics of (left) orderable groups*, Ann. Inst. Fourier (Grenoble) 60 (2010) 1685–1740 MR Zbl

[8] **A Navas**, *Groups of circle diffeomorphisms*, Univ. of Chicago Press (2011) MR Zbl

[9] **P Sankaran**, *On homeomorphisms and quasi-isometries of the real line*, Proc. Amer. Math. Soc. 134 (2006) 1875–1880 MR Zbl

NYU–ECNU Institute of Mathematical Sciences, NYU Shanghai
Shanghai, China

Department of Pure Mathematics, Xi'an Jiaotong–Liverpool University
Suzhou, Jiangsu, China

sy55@nyu.edu, yanxin.zhao19@student.xjtlu.edu.cn

Received: 26 March 2022 Revised: 28 May 2022

Algebraic & Geometric Topology **23**:8 (2023) 3849–3908
DOI: 10.2140/agt.2023.23.3849
Published: 5 November 2023

Simplicial model structures on pro-categories

THOMAS BLOM
IEKE MOERDIJK

We describe a method for constructing simplicial model structures on ind- and pro-categories. Our method is particularly useful for constructing "profinite" analogues of known model categories. Our construction quickly recovers Morel's model structure for pro-p spaces and Quick's model structure for profinite spaces, but we will show that it can also be applied to construct many interesting new model structures. In addition, we study some general properties of our method, such as its functorial behavior and its relation to Bousfield localization. We compare our construction to the ∞–categorical approach to ind- and pro-categories in an appendix.

55U35; 18N40

1 Introduction

In [31; 32], Quick constructed a fibrantly generated Quillen model structure on the category of simplicial profinite sets that models the homotopy theory of "profinite spaces". This can be seen as a continuation of Morel's work in [30], where, for a given prime p, he presented a model structure on the same category that models the homotopy theory of "pro-p spaces".

The purpose of this paper is to present a new and uniform method that immediately gives these two model structures, as well as many others. For example, while Quick's model structure is in a sense derived from the classical homotopy theory of simplicial sets, our method also applies to the Joyal model structure, thus providing a homotopy theory of profinite ∞–categories. Our construction can also be used to obtain a model category of profinite P–stratified spaces, where P is a finite poset, whose underlying ∞–category is the ∞–category of profinite P–stratified spaces defined in Barwick, Glasman and Haine [8].

One general form that our results take is the following version of pro-completion of model categories:

Theorem 1.1 *Let $\mathcal{E}$ be a simplicial model category in which every object is cofibrant and let $\mathbf{C}$ be an (essentially) small full subcategory of $\mathcal{E}$ closed under finite limits and cotensors by finite simplicial sets. Then for any collection $\mathbf{T}$ of fibrant objects in $\mathbf{C}$, the pro-completion $\mathrm{Pro}(\mathbf{C})$ carries a fibrantly generated simplicial model structure with the following properties:*

(i) *The weak equivalences are the $\mathbf{T}$–local equivalences; that is, $f: C \to D$ is a weak equivalence if and only if*

$$f^*: \mathrm{Map}(D, t) \to \mathrm{Map}(C, t)$$

is a weak equivalence of simplicial sets for any $t \in \mathbf{T}$.

(ii) *Every object in $\mathrm{Pro}(\mathbf{C})$ is again cofibrant.*

(iii) *The inclusion $\mathbf{C} \hookrightarrow \mathcal{E}$ induces a simplicial Quillen adjunction $\mathcal{E} \rightleftarrows \mathrm{Pro}(\mathbf{C})$.*

(iv) *If $\mathbf{T} \subseteq \mathbf{C}$ is closed under pullbacks along fibrations (as in Definition 7.10) and cotensors by finite simplicial sets, then the underlying ∞–category of this model structure on $\mathrm{Pro}(\mathbf{C})$ is equivalent to $\mathrm{Pro}(N(\mathbf{T}))$, where $N(\mathbf{T})$ denotes the homotopy coherent nerve of the full simplicial subcategory of $\mathcal{E}$ spanned by the objects of $\mathbf{T}$.*

The model structures of Quick and Morel mentioned above can be obtained from this theorem by appropriately choosing a full subcategory $\mathbf{C}$ of $\mathbf{sSet}$ and a collection $\mathbf{T}$ of fibrant objects. Another known model structure that can be recovered from the above theorem is the model structure for "profinite groupoids" constructed by Horel in [18, Section 4].

The new model category $\mathrm{Pro}(\mathbf{C})$ is a kind of pro-completion of $\mathcal{E}$ with respect to the pair $(\mathbf{C}, \mathbf{T})$, and could be denoted by $\widehat{\mathcal{E}}$ or $\mathcal{E}^{\wedge}_{(\mathbf{C},\mathbf{T})}$. The left adjoint $\mathcal{E} \to \mathrm{Pro}(\mathbf{C})$ of the Quillen adjunction mentioned in item (iii) can be seen as a "pro-$\mathbf{C}$ completion" functor. For the model structures of Morel, Quick and Horel mentioned above, this functor agrees with the profinite completion functor.

We would like to point out that the above formulation is slightly incomplete since there are multiple ways of choosing sets of generating (trivial) fibrations, which theoretically could lead to different model structures on $\mathrm{Pro}(\mathbf{C})$, though always with the weak equivalences as described above. A noteworthy fact is that the above theorem also holds for model categories enriched over the Joyal model structure on simplicial sets, so in particular it applies to the Joyal model structure itself. In this case, the model structure obtained on $\mathrm{Pro}(\mathbf{C})$ is enriched over the Joyal model structure, but not necessarily over the classical Kan–Quillen model structure on $\mathbf{sSet}$. Another fact worth mentioning is that there exist many simplicial model categories satisfying the hypotheses of the above theorem, that is, all objects being cofibrant. Indeed, by a result of Dugger [11, Corollary 1.2], any combinatorial model category is Quillen equivalent to such a simplicial model category.

Even though we are mostly interested in model structures on pro-categories, we will first describe our construction in the context of ind-categories, and then dualize those results. We have chosen this approach since in the case of ind-categories our construction produces cofibrantly generated model categories, which to most readers will be more familiar territory than that of fibrantly generated model categories. In addition, this will make it clear that the core of our argument, which is contained in Section 3, only takes a few pages. Another reason for describing our construction in the context of ind-categories is that an interesting example occurs there: if we apply our construction to a well-chosen full subcategory of the category of topological spaces, then we obtain a model category that is Quillen equivalent to the usual Quillen model structure on $\mathbf{Top}$, but that has many favorable properties, such as being combinatorial.

Our original motivation partly came from the desire to have a full-fledged Quillen-style homotopy theory of profinite ∞–operads, by using the category of dendroidal Stone spaces (ie dendroidal profinite sets). However, not every object is cofibrant in the operadic model structure for dendroidal sets, so the methods from the current paper do not apply directly to this case. The extra work needed to deal with objects that are not cofibrant is of a technical nature, and very specific to the example of dendroidal sets. For this reason, we have decided to present this case separately; see Blom and Moerdijk [9].

Relation to the construction by Barnea and Schlank There are several results in the literature that describe general methods for constructing model structures on ind- or pro-categories. The construction in the current paper is quite close in spirit to that by Barnea and Schlank in [7]. They show that if $\mathbf{C}$ is a category endowed with the structure of a "weak fibration category", then there exists an "induced" model structure on $\mathrm{Pro}(\mathbf{C})$ provided some additional technical requirements are satisfied. However, there are important examples of model structures on pro-categories that are not of this form. For example, Quick's model structure is not of this kind, as explained just above Proposition 7.4.2 in [4]. In the present paper, we prove the existence of a certain model structure on the pro-category of a simplicial category endowed with the extra structure of a so-called "fibration test category" (defined in Definition 5.1). While the definition of a fibration test category given here seems less general than that of a weak fibration category, there are many interesting examples where it is easy to prove that a category is a fibration test category while it is not clear whether this category is a weak fibration category in the sense of [7]. In particular, Quick's model structure can be obtained through our construction; see Example 5.5 and Corollary 6.6. Another advantage is that we do not have to check the technical requirement of "pro-admissibility" (see [7, Definition 4.4]) to obtain a model structure on $\mathrm{Pro}(\mathbf{C})$, which is generally not an easy task. We also believe that our description of the weak equivalences in $\mathrm{Pro}(\mathbf{C})$, namely as the $\mathbf{T}$–local equivalences for some collection of objects $\mathbf{T}$, is often more natural and flexible than the one given in [7]. It is worth pointing out that if both our model structure and that of [7] on $\mathrm{Pro}(\mathbf{C})$ exist, then they agree by Remark 5.12 below.

Overview of the paper In Section 2, we will establish some terminology and mention a few facts on simplicial model categories and ind- and pro-categories. We will then describe our general construction of the model structure for ind-categories in Section 3. We illustrate our construction with an example in Section 4, where we construct a convenient model category of spaces. In Section 5, we dualize our results to the context of pro-categories, and illustrate this dual construction with many examples. We show that some of these examples coincide with model structures that are already known to exist in Section 6, such as Quick's and Morel's model structures. We then continue the study of our construction in Section 7, where we discuss its functorial behavior, and in Section 8, where we prove results about the existence of certain Bousfield localizations. The latter section also contains the proof of Theorem 1.1, except for item (iv). We then give a detailed discussion of two examples in Section 9; namely the model structure for complete Segal profinite spaces and the model structure for profinite quasicategories. In the appendix, we compare our construction to the ∞–categorical approach to ind- and pro-categories.

Acknowledgements We would like to thank the referees for numerous comments that helped improve the exposition.

2 Preliminaries

In this section, we will briefly review some basic definitions concerning simplicial objects, and then discuss ind- and pro-categories.

2.1 Simplicial conventions

We assume the reader to be familiar with the basic theory of simplicial sets, as in [28; 23; 13; 14]. We will say that a simplicial set X is *skeletal* if it is n–skeletal for some natural number n, ie if the map $\operatorname{sk}_n X \to X$ is an isomorphism. Dually, X is *coskeletal* if $X \to \operatorname{cosk}_n X$ is an isomorphism for some n. Recall that a simplicial set X is *degreewise finite* if each X_n is a finite set, and *finite* if it has finitely many nondegenerate simplices. Note that the latter is equivalent to X being degreewise finite and skeletal. We will say that a simplicial set is *lean* if it is degreewise finite and coskeletal, and write $\mathbf{L}$ for the full subcategory of $\mathbf{sSet}$ on the lean simplicial sets. One can show that if X is a lean simplicial set and if Y is a degreewise finite simplicial set, then the cotensor $X^Y = \operatorname{Map}(Y, X)$ is again a lean simplicial set.

Most categories we deal with are *simplicial categories*, ie categories enriched over simplicial sets. Moreover, they will generally be required to have tensors or cotensors by finite simplicial sets. For objects c and d in a simplicial category $\mathbf{C}$, we will write $\operatorname{Map}(c, d)$ for the simplicial hom set. Recall that for a morphism $c \to d$ in $\mathbf{C}$ and a morphism $U \to V$ of simplicial sets, the *pushout-product map* is the map

$$d \otimes U \cup_{c \otimes U} c \otimes V \to d \otimes V,$$

which makes sense in $\mathbf{C}$ if the necessary pushouts and tensors exist. Dually, we refer to

$$c^V \to c^U \times_{d^U} d^V$$

as the *pullback-power map* (if it exists). If given another morphism $a \to b$ in $\mathbf{C}$, we refer to

$$\mathrm{Map}(b,c) \to \mathrm{Map}(a,c) \times_{\mathrm{Map}(a,d)} \mathrm{Map}(b,d)$$

as a *pullback-power map* as well. Note that this map always exists.

We assume the reader to be familiar with the basic theory of Quillen model categories, as in [19; 17]. Basic examples include the classical Kan–Quillen model structure on simplicial sets, which we denote by $\mathbf{sSet}_{\mathrm{KQ}}$, and the Joyal model structure $\mathbf{sSet}_{\mathrm{J}}$ modeling the homotopy theory of ∞–categories [24]. A *simplicial model category* is a model category $\mathcal{E}$ that is enriched, tensored and cotensored over simplicial sets, and that satisfies the additional axiom SM7 phrased in terms of pullback-power maps, or dually in terms of pushout-product maps; see eg [17, Definition 9.1.6 and Proposition 9.3.7] or [33, Definition II.2.2]. We emphasize that we will use this terminology in a somewhat nonstandard way. Namely, by a simplicial model category, we will either mean that the axiom SM7 holds with respect to the Kan–Quillen model structure or the Joyal model structure. Whenever it is necessary to emphasize the distinction, we will call a simplicial model category of the former kind a $\mathbf{sSet}_{\mathrm{KQ}}$*–enriched model category* and the latter a $\mathbf{sSet}_{\mathrm{J}}$*–enriched model category*. Note that any $\mathbf{sSet}_{\mathrm{KQ}}$–enriched model category is $\mathbf{sSet}_{\mathrm{J}}$–enriched, since $\mathbf{sSet}_{\mathrm{KQ}}$ is a left Bousfield localization of $\mathbf{sSet}_{\mathrm{J}}$.

We will make use of the following fact about the (categorical) fibrations in $\mathbf{sSet}_{\mathrm{J}}$.

Lemma 2.1 *There exists a set M of maps between finite simplicial sets such that a map between quasicategories $X \to Y$ is a fibration in* $\mathbf{sSet}_{\mathrm{J}}$ *if and only if it has the right lifting property with respect to all maps in M.*

Proof Let H denote the simplicial set obtained by gluing two 2–simplices to each other along the edges opposite to the 0^{th} and 2^{nd} vertex, respectively, and then collapsing the edges opposite to the 1st vertex to a point in both of these 2–simplices. This means that H looks as follows, where the dashed lines represent the collapsed edges:

$$H = \begin{tikzcd} & \bullet & \\ \bullet \arrow[ur, dash, dashed] \arrow[rr] & & \bullet \arrow[ul] \\ & \bullet \arrow[ul] \arrow[ur, dash, dashed] & \end{tikzcd}$$

A map from H into a quasicategory X consists of an arrow $f \in X_1$, a left and right homotopy inverse $g, h \in X_1$ and homotopies $gf \sim \mathrm{id}$ and $fh \sim \mathrm{id}$. Let $\{0\} \hookrightarrow H$ denote the inclusion of the leftmost vertex into H. It follows from [24, Corollary 2.4.6.5] that if $X \to Y$ is an inner fibration between quasicategories that has the right lifting property with respect to $\{0\} \hookrightarrow H$, then it is a categorical fibration. The converse is also true. To see this, note that for any quasicategory Z, a map $H \to Z$ lands in the largest Kan complex $k(Z)$ contained in Z. Since $\{0\} \hookrightarrow H$ is a weak homotopy equivalence, we see that $\mathrm{Map}(H, Z) = \mathrm{Map}(H, k(Z)) \simeq \mathrm{Map}(\{0\}, k(Z)) = \mathrm{Map}(\{0\}, Z)$, so the inclusion $\{0\} \hookrightarrow H$ is a categorical equivalence. In particular, any categorical fibration has the right lifting property with respect to $\{0\} \to H$. We conclude that the set $M = \{\Lambda^n_k \hookrightarrow \Delta^n \mid 0 < k < n\} \cup \{\{0\} \hookrightarrow H\}$ has the desired properties. ◻

2.2 Ind- and pro-categories

In this section we recall some basic definitions concerning ind- and pro-categories. Most of these will be familiar to the reader, with the possible exception of Theorem 2.3 below. For details, we refer the reader to [15; 12, Section 2.1; 2, Appendix; 20]. In the discussion below, all (co)limits are assumed to be *small*.

For a category $\mathbf{C}$, its ind-completion $\mathrm{Ind}(\mathbf{C})$ is obtained by freely adjoining filtered (or directed) colimits to $\mathbf{C}$. Dually, the free completion under cofiltered limits is denoted by $\mathrm{Pro}(\mathbf{C})$. This in particular means that $\mathrm{Pro}(\mathbf{C})^{\mathrm{op}} = \mathrm{Ind}(\mathbf{C}^{\mathrm{op}})$, so any statement about ind-categories dualizes to a statement about pro-categories and vice versa. We will therefore mainly discuss ind-categories here and leave it to the reader to dualize the discussion.

One way to make the above precise is to define the objects in $\mathrm{Ind}(\mathbf{C})$ to be all diagrams $I \to \mathbf{C}$ for all filtered categories I. Such objects are called *ind-objects* and denoted by $C = \{c_i\}_{i \in I}$. The morphisms between two such objects $C = \{c_i\}_{i \in I}$ and $D = \{d_j\}_{j \in J}$ are defined by

(1) $$\mathrm{Hom}_{\mathrm{Ind}(\mathbf{C})}(C, D) = \lim_i \operatorname*{colim}_j \mathrm{Hom}_{\mathbf{C}}(c_i, d_j).$$

If $\mathbf{C}$ is a simplicial category, then $\mathrm{Ind}(\mathbf{C})$ can be seen as a simplicial category as well. The enrichment is expressed by a formula similar to (1), namely

$$\mathrm{Map}(\{c_i\}, \{d_j\}) = \lim_i \operatorname*{colim}_j \mathrm{Map}(c_i, d_j).$$

One can define the pro-category $\mathrm{Pro}(\mathbf{C})$ of a (simplicial) category $\mathbf{C}$ as the category of all diagrams $I \to \mathbf{C}$ for all cofiltered I, and with (simplicial) hom sets dual to the ones above. An object in $\mathrm{Pro}(\mathbf{C})$ is called a *pro-object*. One could also simply define $\mathrm{Pro}(\mathbf{C})$ as $\mathrm{Ind}(\mathbf{C}^{\mathrm{op}})^{\mathrm{op}}$.

It can be shown that any object in $\mathrm{Ind}(\mathbf{C})$ is isomorphic to one where the indexing category I is a directed poset, and dually that any object in $\mathrm{Pro}(\mathbf{C})$ is isomorphic to one that is indexed by a codirected poset; see [15, Proposition 8.1.6], or [12, Theorem 2.1.6] with a correction just after Corollary 3.11 of [5].

There is a fully faithful embedding $\mathbf{C} \hookrightarrow \mathrm{Ind}(\mathbf{C})$ sending an object c to the constant diagram with value c, again denoted by c. We will generally identify $\mathbf{C}$ with its image in $\mathrm{Ind}(\mathbf{C})$ under this embedding. This embedding preserves all limits and all finite colimits that exist in $\mathbf{C}$. The universal property of $\mathrm{Ind}(\mathbf{C})$ states that $\mathrm{Ind}(\mathbf{C})$ has all filtered colimits and that any functor $F\colon \mathbf{C} \to \mathcal{E}$, where $\mathcal{E}$ is a category that has all filtered colimits, has an essentially unique extension to a functor $\widetilde{F}\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ that preserves filtered colimits. This extension can be defined explicitly by $\widetilde{F}(\{c_i\}) = \mathrm{colim}_i\, F(c_i)$.

Recall that if $\mathcal{E}$ is a category that has all filtered colimits, then an object c in $\mathcal{E}$ is called *compact* if $\mathrm{Hom}_{\mathcal{E}}(c, -)$ commutes with filtered colimits. The dual notion is called *cocompact*. One can deduce from the definition of the morphisms in $\mathrm{Ind}(\mathbf{C})$ that any object in the image of $\mathbf{C} \hookrightarrow \mathrm{Ind}(\mathbf{C})$ is compact. Dually, the objects of $\mathbf{C}$ are cocompact in $\mathrm{Pro}(\mathbf{C})$.

There is the following recognition principle for ind-completions, whose proof we leave to the reader.

Lemma 2.2 (recognition principle) *Let $\mathcal{E}$ be a category closed under filtered colimits and let $\mathbf{C} \hookrightarrow \mathcal{E}$ be a full subcategory. If*

(i) *any object in $\mathbf{C}$ is compact in $\mathcal{E}$, and*

(ii) *any object in $\mathcal{E}$ is a filtered colimit of objects in $\mathbf{C}$,*

then the canonical extension $\mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$, coming from the universal property of $\mathrm{Ind}(\mathbf{C})$, is an equivalence of categories.

To avoid size issues, we assume from now on that $\mathbf{C}$ is an (essentially) small category. The fact that the presheaf category $\mathbf{Set}^{\mathbf{C}^{\mathrm{op}}}$ is the free cocompletion of $\mathbf{C}$ leads to an alternative description of $\mathrm{Ind}(\mathbf{C})$ that is sometimes easier to work with. Namely, we can think of $\mathrm{Ind}(\mathbf{C})$ as the full subcategory of $\mathbf{Set}^{\mathbf{C}^{\mathrm{op}}}$ consisting of those presheaves which are filtered colimits of representables. If $\mathbf{C}$ is small and has finite colimits, as will be the case in all of our examples, then these are exactly the functors $\mathbf{C}^{\mathrm{op}} \to \mathbf{Set}$ that send the finite colimits of $\mathbf{C}$ to limits in $\mathbf{Set}$ (see [15, Théorème 8.3.3(v)]), that is,

$$\mathrm{Ind}(\mathbf{C}) \simeq \mathrm{lex}(\mathbf{C}^{\mathrm{op}}, \mathbf{Set}),$$

where the right-hand side stands for the category of left exact functors. From this description, one sees immediately that $\mathrm{Ind}(\mathbf{C})$ has all small limits and that the inclusion $\mathrm{Ind}(\mathbf{C}) \to \mathbf{Set}^{\mathbf{C}^{\mathrm{op}}}$ preserves these. The category $\mathrm{Ind}(\mathbf{C})$ also has all colimits in this case. Namely, finite coproducts and pushouts can be computed "levelwise" in $\mathbf{C}$ as described in [2, Appendix 4], while filtered colimits exist as mentioned above. Note, however, that the inclusion $\mathrm{Ind}(\mathbf{C}) \to \mathbf{Set}^{\mathbf{C}^{\mathrm{op}}}$ does not preserve all colimits, but only filtered ones.

One sees dually that if $\mathbf{C}$ is small and has all finite limits, then

$$\mathrm{Pro}(\mathbf{C}) \simeq \mathrm{lex}(\mathbf{C}, \mathbf{Set})^{\mathrm{op}}.$$

As above, it follows that $\mathrm{Pro}(\mathbf{C})$ is complete and cocomplete in this case.

Another consequence of the fact that finite coproducts and pushouts in $\mathrm{Ind}(\mathbf{C})$ are computed "levelwise" is the following: if $F\colon \mathbf{C} \to \mathcal{E}$, with $\mathcal{E}$ cocomplete, preserves finite colimits, then its extension $\widetilde{F}\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ given by the universal property also preserves finite colimits. Since it also preserves filtered colimits, we conclude that it preserves all colimits. In fact, more is true. The above description of $\mathrm{Ind}(\mathbf{C})$ as $\mathrm{lex}(\mathbf{C}^{\mathrm{op}}, \mathbf{Set})$ allows us to construct a right adjoint R of $\widetilde{F}$. Namely, if we define $R(E)(c) := \mathrm{Hom}(Fc, E)$, then $R(E)\colon \mathbf{C}^{\mathrm{op}} \to \mathbf{Set}$ is left exact, hence R defines a functor $\mathcal{E} \to \mathrm{Ind}(\mathbf{C})$. Adjointness follows from the Yoneda lemma. We therefore see that, up to unique natural isomorphism, there is a one-to-one correspondence between finite colimit-preserving functors $\mathbf{C} \to \mathcal{E}$ and functors $\mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ that have a right adjoint.

There are two important examples of adjunctions obtained in this way that we would like to mention here. The first one is the ind-completion functor. If $\mathcal{E}$ is a cocomplete category and $\mathbf{C}$ a full subcategory closed under finite colimits, then the inclusion $\mathbf{C} \subseteq \mathcal{E}$ induces an adjunction

$$U : \mathrm{Ind}(\mathbf{C}) \rightleftarrows \mathcal{E} : \widehat{(\,\cdot\,)}_{\mathrm{Ind}},$$

whose right adjoint we call *ind-completion* (*relative to* $\mathbf{C}$) or *ind*-$\mathbf{C}$ *completion*. Dually, if $\mathcal{E}$ is complete and $\mathbf{C}$ is a full subcategory closed under finite limits, then we obtain an adjunction

$$\widehat{(\,\cdot\,)}_{\mathrm{Pro}} : \mathcal{E} \rightleftarrows \mathrm{Pro}(\mathbf{C}) : U,$$

whose left adjoint we call *pro-completion* (*relative to* $\mathbf{C}$) or *pro-*$\mathbf{C}$ *completion*. In many examples, $\mathbf{C}$ is the full subcategory of $\mathcal{E}$ consisting of objects that are "finite" in some sense, and this left adjoint is better known as the profinite completion functor. For instance, in the case of groups, this functor $\widehat{(\cdot)}_{\mathrm{Pro}} \colon \mathbf{Grp} \to \mathrm{Pro}(\mathbf{FinGrp})$ is the well-known profinite completion functor for groups.

The other important example is about cotensors in ind-categories. Suppose $\mathbf{C}$ is a small simplicial category that has all finite colimits and tensors with finite simplicial sets, and that furthermore these tensors commute with these finite colimits. We will call $\mathbf{C}$ *finitely tensored* if this is the case; see Definition 3.1 for a precise definition. If X is a simplicial set, then we can write it as $\mathrm{colim}_i\, X_i$, where i ranges over all finite simplicial subsets $X_i \subseteq X$. Define

$$- \otimes X \colon \mathbf{C} \to \mathrm{Ind}(\mathbf{C}) \quad \text{by } c \otimes X = \{c \otimes X_i\}_i.$$

This functor preserves finite colimits since these are computed "levelwise" in $\mathrm{Ind}(\mathbf{C})$, hence it extends to a functor $- \otimes X \colon \mathrm{Ind}(\mathbf{C}) \to \mathrm{Ind}(\mathbf{C})$ that has a right adjoint $(-)^X$. These define tensors and cotensors by arbitrary simplicial sets on $\mathrm{Ind}(\mathbf{C})$. In particular, $\mathrm{Ind}(\mathbf{C})$ is a simplicial category that is complete, cocomplete, tensored and cotensored; note the similarity with [6, Proposition 4.10]. The dual of this statement says that for any small simplicial category $\mathbf{C}$ that has finite limits and cotensors with finite simplicial sets, and in which these finite cotensors commute with finite limits in $\mathbf{C}$, the pro-category $\mathrm{Pro}(\mathbf{C})$ is a simplicial category that is complete, cocomplete, tensored and cotensored. We call $\mathbf{C}$ *finitely cotensored* in this case.

Let us return to the basic definition (1) of morphisms in $\mathrm{Ind}(\mathbf{C})$. If $C = \{c_i\}$ and $D = \{d_i\}$ are objects indexed by the same filtered category I, then any natural transformation with components $f_i \colon c_i \to d_i$ represents a morphism in $\mathrm{Ind}(\mathbf{C})$. Morphisms of this type (or more precisely, morphisms represented in this way) will be called *level maps* or *strict maps*. Up to isomorphism, any morphism in $\mathrm{Ind}(\mathbf{C})$ has such a strict representation; see Corollary 3.2 of [2, Appendix]. One can define the notion of a "level" diagram or "strict" diagram in a similar way. Given an indexing category K, a conceptual way of thinking about these is through the canonical functor

$$\mathrm{Ind}(\mathbf{C}^K) \to \mathrm{Ind}(\mathbf{C})^K.$$

A strict diagram can be thought of as an object in the image of this functor. If K is a finite category and $\mathbf{C}$ has all finite colimits, then the above functor is an equivalence of categories [29, Section 4]. This shows in particular that, up to isomorphism, any finite diagram in $\mathrm{Ind}(\mathbf{C})$ is a strict diagram if $\mathbf{C}$ is small and has finite colimits.

In our context, the following extension of Meyer's result is important. Suppose that K is a category which can be written as a union of a sequence of finite full subcategories

$$K_0 \subseteq K_1 \subseteq K_2 \subseteq \cdots \subseteq K = \bigcup_{n \in \mathbb{N}} K_n.$$

Let $\mathbf{C}$ be a small category that has finite colimits. Then any functor $f\colon K_n \to \mathbf{C}$ has a left Kan extension $g\colon K \to \mathbf{C}$ defined in terms of finite colimits as in (the dual to) Theorem X.3.1 of [26]. For $X\colon K \to \mathbf{C}$, write $\mathrm{sk}_n X$ for the left Kan extension of the restriction of X to K_n. We call X *n–skeletal* if the canonical map $\mathrm{sk}_n X \to X$ is an isomorphism, and *skeletal* if this is the case for some n. The full subcategory $\mathrm{sk}(\mathbf{C}^K) \subseteq \mathbf{C}^K$ spanned by the skeletal functors $K \to C$ can be viewed as a full subcategory of $\mathrm{Ind}(\mathbf{C})^K$ via the inclusion $\mathbf{C} \hookrightarrow \mathrm{Ind}(\mathbf{C})$. Note that for any X in $\mathrm{Ind}(\mathbf{C})^K$, we have $X = \mathrm{colim}_n \, \mathrm{sk}_n X$. Exactly as in (the dual of) the proof of [4, Proposition 7.4.1], the result of [29, Section 4] mentioned above can be used to show that the hypotheses of the recognition principle for ind-categories are satisfied, hence that the induced functor $\mathrm{Ind}(\mathrm{sk}(\mathbf{C}^K)) \to \mathrm{Ind}(\mathbf{C})^K$ is an equivalence of categories. In fact, the assumption that K is a union of a *sequence* of finite full subcategories is irrelevant, and the following more general result, which we write down for future reference, can be proved by the same argument. Note that a category K can be written as a union of finite full subcategories if and only if for any $k, k' \in K$, the set $\mathrm{Hom}_K(k, k')$ is finite.

Theorem 2.3 *Let $\mathbf{C}$ be a small category that has finite colimits, and let K be a small category that can be written as a union of finite full subcategories. Write $\mathrm{sk}(\mathbf{C}^K)$ for the full subcategory of $\mathbf{C}^K$ of those functors $K \to \mathbf{C}$ that are isomorphic to the left Kan extension of a functor $K' \to \mathbf{C}$ for some finite full subcategory $K' \subseteq K$. Then $\mathrm{Ind}(\mathrm{sk}(\mathbf{C}^K)) \simeq \mathrm{Ind}(\mathbf{C})^K$.*

This theorem recovers the well-known equivalence $\mathrm{Ind}(\mathbf{sSet}_{\mathrm{fin}}) \simeq \mathbf{sSet}$ when applied to $\Delta^{\mathrm{op}} = \bigcup_n \Delta^{\mathrm{op}}_{\leq n}$ and $\mathbf{C} = \mathbf{FinSet}$. Note that we already (implicitly) used this equivalence when we defined tensors by simplicial sets for ind-categories above.

We can also apply the dual of this theorem to the same categories $K = \Delta^{\mathrm{op}}$ and $\mathbf{C} = \mathbf{FinSet}$. Write $\widehat{\mathbf{Set}} = \mathrm{Pro}(\mathbf{FinSet})$ for the category of profinite sets, which is well known to be equivalent to the category of Stone spaces **Stone**. Since we want to apply the dual of Theorem 2.3, we need to work with right Kan extensions instead of left Kan extensions. In particular, we obtain the full subcategory of $\mathbf{FinSet}^{\Delta^{\mathrm{op}}}$ on those simplicial sets that are the right Kan extension of some functor $\Delta^{\mathrm{op}}_{\leq n} \to \mathbf{FinSet}$. These are exactly the coskeletal degreewise finite simplicial sets, ie the lean simplicial sets. In particular, the theorem above recovers the equivalence $\mathrm{Pro}(\mathbf{L}) \simeq \widehat{\mathbf{sSet}}$ proved in Proposition 7.4.1 of [4].

An example that plays an important role in Section 9 is that of bisimplicial (profinite) sets. The dual of the above theorem shows that the category of bisimplicial profinite sets $\widehat{\mathbf{bisSet}} = \widehat{\mathbf{sSet}}^{\Delta^{\mathrm{op}}} \cong \widehat{\mathbf{Set}}^{\Delta^{\mathrm{op}} \times \Delta^{\mathrm{op}}}$ is canonically equivalent to the category $\mathrm{Pro}(\mathbf{L}^{(2)})$ for a certain full subcategory $\mathbf{L}^{(2)}$ of the category of bisimplicial sets $\mathbf{bisSet} = \mathbf{sSet}^{\Delta^{\mathrm{op}}}$. This category $\mathbf{L}^{(2)}$ consists of those bisimplicial sets that are isomorphic to the right Kan extension of a functor $\Delta^{\mathrm{op}}_{\leq t} \times \Delta^{\mathrm{op}}_{\leq n} \to \mathbf{FinSet}$ along the inclusion $\Delta^{\mathrm{op}}_{\leq t} \times \Delta^{\mathrm{op}}_{\leq n} \hookrightarrow \Delta^{\mathrm{op}} \times \Delta^{\mathrm{op}}$ for some $t, n \in \mathbb{N}$. We will refer to such bisimplicial sets as *doubly lean*.

3 The completed model structure on $\mathrm{Ind}(\mathbf{C})$

In this section, we will describe our construction of the model structure on $\mathrm{Ind}(\mathbf{C})$, where $\mathbf{C}$ is what we call a "cofibration test category". In Section 5, we will dualize this construction to the context of pro-categories. After that, we will study the functorial behavior of the construction in Section 7 and discuss Bousfield localizations in Section 8.

Throughout these sections, the terms "weak equivalence" and "fibration" of simplicial sets refer to either the classical Kan–Quillen model structure or to the Joyal model structure. When we say that a model category is simplicial, this can mean either that it is enriched over the Kan–Quillen model structure or over the Joyal model structure.

We wish to single out the definition of being finitely tensored, since it occurs many times throughout this paper.

Definition 3.1 Let $\mathbf{C}$ be a simplicial category. Then $\mathbf{C}$ is called *finitely tensored* if

(i) it admits finite colimits,

(ii) it admits tensors by finite simplicial sets, and

(iii) these commute with each other, meaning that the canonical map

$$\operatorname*{colim}_i (c_i \otimes X) \to \big(\operatorname*{colim}_i c_i\big) \otimes X$$

is an isomorphism for any finite diagram $\{c_i\}$ in $\mathbf{C}$ and any finite simplicial set X.

Remark 3.2 Condition (iii) is equivalent to asking that the finite colimits of (i) are *enriched colimits*; that is, for any finite diagram $\{c_i\}$ in $\mathbf{C}$ and any object d in $\mathbf{C}$, the canonical map $\operatorname{colim}_i \operatorname{Map}(c_i, d) \to \operatorname{Map}(\operatorname{colim}_i c_i, d)$ is an isomorphism of simplicial sets.

As explained in Section 2.2, if $\mathbf{C}$ is finitely tensored, then the category $\operatorname{Ind}(\mathbf{C})$ is a tensored and cotensored simplicial category that is both complete and cocomplete. We will endow $\mathbf{C}$ with some additional structure, that of a "cofibration test category", and show that it induces a simplicial model structure on $\operatorname{Ind}(\mathbf{C})$ in Theorem 3.9 below.

Definition 3.3 A *cofibration test category* $(\mathbf{C}, \mathbf{T})$ consists of a small finitely tensored simplicial category $\mathbf{C}$, a full subcategory $\mathbf{T}$ of *test objects* and two classes of maps in $\mathbf{T}$ called *cofibrations*, denoted by $\rightarrowtail$, and *trivial cofibrations*, denoted by $\xrightarrow{\sim}$, both containing all isomorphisms, that satisfy the following properties:

(i) The initial object $\varnothing$ is a test object, and for every test object $t \in \mathbf{T}$, the map $\varnothing \to t$ is a cofibration.

(ii) For every cofibration between test objects $s \rightarrowtail t$ and cofibration between finite simplicial set $U \rightarrowtail V$, the pushout-product map $t \otimes U \cup_{s \otimes U} s \otimes V \to t \otimes V$ is a cofibration between test objects which is trivial if either $s \rightarrowtail t$ or $U \rightarrowtail V$ is.

(iii) A morphism $r \to s$ in $\mathbf{T}$ is a trivial cofibration if and only if it is a cofibration and $\operatorname{Map}(t, r) \to \operatorname{Map}(t, s)$ is a weak equivalence of simplicial sets for every $t \in \mathbf{T}$.

(iv) Any object $c \in \mathbf{C}$ has the right lifting property with respect to trivial cofibrations.

Remark 3.4 Property (iv) implies that $\operatorname{Map}(t, C)$ is fibrant for every $t \in \mathbf{T}$ and $C \in \operatorname{Ind}(\mathbf{C})$. Namely, writing C as a filtered colimit $\operatorname{colim}_i c_i$ with $c_i \in \mathbf{C}$ for every i, we see that $\operatorname{Map}(t, C) = \operatorname{colim}_i \operatorname{Map}(t, c_i)$. Hence it suffices to show that $\operatorname{Map}(t, c)$ is fibrant for every object c in $\mathbf{C}$. This is equivalent to c having the right lifting property with respect to certain maps of the form $t \otimes \Lambda^n_k \to t \otimes \Delta^n$, which is indeed the case by items (i), (ii) and (iv).

Remark 3.5 The definition of a cofibration test category depends on whether we work with the Kan–Quillen model structure $\mathbf{sSet}_{\mathrm{KQ}}$ or the Joyal model structure $\mathbf{sSet}_{\mathrm{J}}$. However, since $\mathbf{sSet}_{\mathrm{KQ}}$ is a left Bousfield localization of $\mathbf{sSet}_{\mathrm{J}}$, any cofibration test category with respect to $\mathbf{sSet}_{\mathrm{KQ}}$ is also a cofibration test category with respect to $\mathbf{sSet}_{\mathrm{J}}$.

To see this, suppose that $(\mathbf{C}, \mathbf{T})$ is a cofibration test category with respect to $\mathbf{sSet}_{\mathrm{KQ}}$. It is clear that items (i), (ii) and (iv) also hold with respect to $\mathbf{sSet}_{\mathrm{J}}$. For item (iii), note that the map $\mathrm{Map}(t,r) \to \mathrm{Map}(t,s)$ is a map between Kan complexes by Remark 3.4, hence it is a weak equivalence in $\mathbf{sSet}_{\mathrm{J}}$ if and only if it is in $\mathbf{sSet}_{\mathrm{KQ}}$.

We will often write $\mathbf{C}$ for a cofibration test category $(\mathbf{C}, \mathbf{T})$, omitting the full subcategory of test objects $\mathbf{T}$ from the notation. We will write $\mathrm{cof}(\mathbf{C})$ for the set of cofibrations. Note that this is a subset of the morphisms of $\mathbf{T}$.

The role of the test objects $t \in \mathbf{T}$ is to detect the weak equivalences in $\mathrm{Ind}(\mathbf{C})$ "from the left". More precisely, the weak equivalences in $\mathrm{Ind}(\mathbf{C})$ will be those arrows $C \to D$ for which $\mathrm{Map}(t, C) \to \mathrm{Map}(t, D)$ is a weak equivalence for every $t \in \mathbf{T}$. For this reason, we will call an arrow $c \to d$ in $\mathbf{C}$ for which $\mathrm{Map}(t,c) \to \mathrm{Map}(t,d)$ is a weak equivalence for every $t \in \mathbf{T}$ a *weak equivalence*, and denote such arrows by $\xrightarrow{\sim}$. We write $\mathrm{we}(\mathbf{C})$ for the set of weak equivalences in $\mathbf{C}$. Using this terminology, item (iii) of the above definition can be rephrased as saying that the trivial cofibrations are precisely the cofibrations that are weak equivalences. In particular, the set of trivial cofibrations in a cofibration test category $\mathbf{C}$ is fully determined by the full subcategory $\mathbf{T}$ and the set $\mathrm{cof}(\mathbf{C})$.

Let us look at a few examples. We will discuss more interesting examples in Section 5, where we consider fibration test categories, the dual of cofibration test categories.

Example 3.6 Suppose $\mathcal{E}$ is a simplicial model category in which every object is fibrant, and let $\mathbf{C} \subseteq \mathcal{E}$ be a (small) full subcategory closed under finite colimits and finite tensors. If we define $\mathbf{T}$ to be the full subcategory on the cofibrant objects, then $(\mathbf{C}, \mathbf{T})$ forms a cofibration test category where the (trivial) cofibrations are the (trivial) cofibrations of $\mathcal{E}$ between objects of $\mathbf{T}$. We say that $\mathbf{C}$ *inherits* this structure of a cofibration test category from $\mathcal{E}$. Properties (i), (ii) and (iv) of Definition 3.3 follow directly from the fact that $\mathcal{E}$ is a (simplicial) model category and the fact that any object in $\mathcal{E}$ is fibrant. For one direction of property (iii), note that since all objects in $\mathcal{E}$ are fibrant, the functor $\mathrm{Map}(t, -)$ preserves weak equivalences for any cofibrant object t. For the converse direction, note that a cofibration $r \rightarrowtail s$ is trivial if and only if it is mapped to an isomorphism in the homotopy category $\mathrm{Ho}(\mathcal{E})$. By the Yoneda lemma applied to the full subcategory $\mathrm{Ho}(\mathbf{T}) \subseteq \mathrm{Ho}(\mathcal{E})$ spanned by the objects of $\mathbf{T}$, this is equivalent to $\mathrm{Hom}_{\mathrm{Ho}(\mathcal{E})}(t,r) \to \mathrm{Hom}_{\mathrm{Ho}(\mathcal{E})}(t,s)$ being an isomorphism for every t. Since $\mathrm{Map}(t,r) \to \mathrm{Map}(t,s)$ is a weak equivalence by assumption and $\mathrm{Hom}_{\mathrm{Ho}(\mathcal{E})}(t,-)$ equals the set of path components of (the maximal Kan complex contained in) $\mathrm{Map}(t,-)$, this is indeed the case.

Example 3.7 Suppose that a cofibration test category $(\mathbf{C}, \mathbf{T})$ is given, and let $\mathbf{T}' \subseteq \mathbf{T}$ be a full subcategory such that $\varnothing \in \mathbf{T}'$ and such that for any cofibration $s \rightarrowtail t$ between objects of $\mathbf{T}'$ and any cofibration $U \rightarrowtail V$ in $\mathbf{sSet}_{\mathrm{fin}}$, the object $t \otimes U \cup_{s \otimes U} s \otimes V$ is again in $\mathbf{T}'$. We will call such a full subcategory $\mathbf{T}' \subseteq \mathbf{T}$ *closed under finite pushout-products*. Then $(\mathbf{C}, \mathbf{T}')$ is again a cofibration test category if we define the (trivial) cofibrations to be those of $(\mathbf{C}, \mathbf{T})$ between objects of $\mathbf{T}'$. All items of Definition 3.3 are straightforward to show except possibly property (iii). The "only if" direction follows immediately. For the "if" direction of (iii), suppose $r \rightarrowtail s$ is a map in $\mathbf{T}'$ that is a cofibration with the property that $\mathrm{Map}(t, r) \to \mathrm{Map}(t, s)$ is a weak equivalence for any $t \in \mathbf{T}'$. Applying this to $t = r$ and $t = s$ and using that these mapping spaces are fibrant, we obtain left and right homotopy inverses of $r \rightarrowtail s$, where homotopies in $\mathbf{T}'$ are defined using the tensor $- \otimes \Delta^1$ (in the case of $\mathbf{sSet}_{\mathrm{KQ}}$) or $- \otimes H$ (in the case $\mathbf{sSet}_{\mathrm{J}}$, where H is as in the proof of Lemma 2.1). Since $\mathrm{Map}(t, -)$ is a simplicial functor it preserves these homotopies, showing that $\mathrm{Map}(t, r) \to \mathrm{Map}(t, s)$ is a homotopy equivalence for every $t \in \mathbf{T}$. We conclude that $r \rightarrowtail s$ is a trivial cofibration in $\mathbf{T}$ and hence a trivial cofibration in $\mathbf{T}'$ by definition.

Example 3.8 Let **Top** be a convenient category of topological spaces, such as k–spaces or compactly generated (weak) Hausdorff spaces. The Quillen model structure on **Top** is a simplicial model structure, in which tensors are given by $C \otimes X = C \times |X|$ for any $C \in \mathbf{Top}$ and $X \in \mathbf{sSet}$. Let $\mathbf{C} \subseteq \mathbf{Top}$ be any small full subcategory of **Top** that is closed under finite colimits and finite tensors, and moreover contains the space $|X|$ for any finite simplicial set X. Define $\mathbf{T} \subseteq \mathbf{C}$ to be the full subcategory consisting of the objects $|X|$ for any finite simplicial set X, and define a map $|X| \to |Y|$ in $\mathbf{T}$ to be a (trivial) cofibration if it is the geometric realization of a (trivial) cofibration $X \rightarrowtail Y$ in the Kan–Quillen model structure on **sSet**. Using that there are natural isomorphisms $|Y| \otimes V \cong |Y \times V|$ and $|X| \otimes V \cup_{|X| \otimes U} |Y| \otimes U \cong |X \times V \cup_{X \times U} Y \times V|$ for any pair of maps $X \to Y$ and $U \to V$ in **sSet**, it is straightforward to verify that $(\mathbf{C}, \mathbf{T})$ is a cofibration test category in the sense of Definition 3.3 (with respect to $\mathbf{sSet}_{\mathrm{KQ}}$). This example will be studied further in Section 4.

For a cofibration test category $\mathbf{C}$, we will write I for the image of the set of cofibrations of $\mathbf{C}$ in $\mathrm{Ind}(\mathbf{C})$, and J for the image of the set of trivial cofibrations of $\mathbf{C}$ in $\mathrm{Ind}(\mathbf{C})$. Identifying $\mathbf{C}$ with its image in $\mathrm{Ind}(\mathbf{C})$, we can write

$$I = \{f : s \to t \mid f \text{ is a cofibration in } \mathbf{C}\} = \mathrm{cof}(\mathbf{C}),$$
$$J = \{f : s \to t \mid f \text{ is a trivial cofibration in } \mathbf{C}\} = \mathrm{cof}(\mathbf{C}) \cap \mathrm{we}(\mathbf{C}).$$

Recall that the sets of (trivial) cofibrations $\mathrm{cof}(\mathbf{C})$ and $\mathrm{cof}(\mathbf{C}) \cap \mathrm{we}(\mathbf{C})$ in $(\mathbf{C}, \mathbf{T})$ are both contained in $\mathbf{T}$; that is, any (trivial) cofibration is a map between test objects.

The sets I and J are generating (trivial) cofibrations for a model structure on $\mathrm{Ind}(\mathbf{C})$ in which the weak equivalences are as above.

Theorem 3.9 *Let* $\mathbf{C}$ *be a cofibration test category. Then* $\mathrm{Ind}(\mathbf{C})$ *carries a cofibrantly generated (hence combinatorial) simplicial model structure, the* ***completed model structure****, where a map* $C \to D$ *is a weak equivalence if and only if for every* $t \in \mathbf{T}$, $\mathrm{Map}(t, C) \to \mathrm{Map}(t, D)$ *is a weak equivalence. A set of generating cofibrations (resp. generating trivial cofibrations) is given by* I *(resp.* J*). Every object is fibrant in this model structure.*

Remark 3.10 As mentioned in Remark 3.5, the definition of a cofibration test category depends on whether we work with the Joyal model structure or the Kan–Quillen model structure on $\mathbf{sSet}$. In the first case, the model structure on $\mathrm{Ind}(\mathbf{C})$ will be $\mathbf{sSet}_{\mathrm{J}}$–enriched, while in the latter case, it will be $\mathbf{sSet}_{\mathrm{KQ}}$–enriched.

The proof uses the following lemmas.

Lemma 3.11 *Let* $\mathbf{C}$ *be a cofibration test category. The weak equivalences of* $\mathrm{Ind}(\mathbf{C})$ *as defined in Theorem 3.9 are stable under filtered colimits.*

Proof Let $\{C_i \xrightarrow{\sim} D_i\}$ be a levelwise weak equivalence between filtered diagrams in $\mathrm{Ind}(\mathbf{C})$ and let $t \in \mathbf{T}$. Then $\mathrm{Map}(t, \mathrm{colim}_i\, C_i) \to \mathrm{Map}(t, \mathrm{colim}_i\, D_i)$ is the filtered colimit of the maps $\mathrm{Map}(t, C_i) \to \mathrm{Map}(t, D_i)$ since t is compact in $\mathrm{Ind}(\mathbf{C})$, which are weak equivalences by assumption. The proof therefore reduces to the statement in $\mathbf{sSet}$ that a filtered colimit of weak equivalences, indexed by some filtered category I, is again a weak equivalence. This can be proved for the Kan–Quillen and Joyal model structure in exactly the same way. Namely, this is equivalent to the statement that the functor $\mathrm{colim}\colon \mathbf{sSet}^{I} \to \mathbf{sSet}$, where $\mathbf{sSet}^{I}$ is endowed with the projective model structure, preserves weak equivalences. To see that this is the case, factor $\{X_i\} \xrightarrow{\sim} \{Y_i\}$ in $\mathbf{sSet}^{I}$ as a projective trivial cofibration $\{X_i\} \overset{\sim}{\rightarrowtail} \{Z_i\}$ followed by a pointwise trivial fibration $\{Z_i\} \overset{\sim}{\twoheadrightarrow} \{Y_i\}$. Then $\mathrm{colim}\, X_i \to \mathrm{colim}\, Z_i$ is again a trivial cofibration, so in particular a weak equivalence. Furthermore, since the generating cofibrations $\partial\Delta^n \to \Delta^n$ in $\mathbf{sSet}$ are maps between compact objects, we see that $\mathrm{colim}\, Z_i \to \mathrm{colim}\, Y_i$ must have the right lifting property with respect to these maps, ie it is a trivial fibration. We conclude that $\mathrm{colim}\colon \mathbf{sSet}^{I} \to \mathbf{sSet}$ preserves weak equivalences. □

Lemma 3.12 *Let $\mathbf{C}$ be a cofibration test category, let $s \rightarrowtail t$ be a cofibration in $\mathbf{C}$, ie a map in I, and let $C \to D$ be an arrow in $\mathrm{Ind}(\mathbf{C})$ which has the right lifting property with respect to all maps in J. Then $\mathrm{Map}(t, C) \to \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is a fibration, which is trivial if either $s \rightarrowtail t$ is trivial or if $C \to D$ is a weak equivalence in the sense of Theorem 3.9.*

Proof Let M be a set of trivial cofibrations in $\mathbf{sSet}_{\mathrm{fin}}$ such that a map between fibrant objects in $\mathbf{sSet}$ is a fibration if and only if it has the right lifting property with respect to the maps in M. For the Kan–Quillen model structure, one can take the set of horn inclusions, while for $\mathbf{sSet}_{\mathrm{J}}$, the set M from Lemma 2.1 works. By Remark 3.4, for any test object $t \in \mathbf{T}$ and any $C \in \mathrm{Ind}(\mathbf{C})$, the simplicial set $\mathrm{Map}(t, C)$ is fibrant. For any $t \in \mathbf{T}$ and any map $U \xrightarrow{\sim} V$ in M, the map $t \otimes U \to t \otimes V$ is in J by items (i) and (ii) of Definition 3.3. By adjunction, we conclude that for any $C \to D$ that has the right lifting property with respect to maps in J, the map $\mathrm{Map}(t, C) \to \mathrm{Map}(t, D)$ is a fibration. If we are given a map $s \rightarrowtail t$ in I, then $\mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is fibrant because the map to $\mathrm{Map}(t, D)$ is the pullback of the fibration $\mathrm{Map}(s, C) \twoheadrightarrow \mathrm{Map}(s, D)$. By a similar argument as above, $\mathrm{Map}(t, C) \to \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is a fibration. The same argument with the set of boundary inclusions $\{\partial\Delta^n \to \Delta^n\}$ instead of M shows that $\mathrm{Map}(t, C) \to \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is a trivial fibration if $s \rightarrowtail t$ is in J. If $C \to D$ is a weak equivalence, then the maps $\mathrm{Map}(s, C) \to \mathrm{Map}(s, D)$ and $\mathrm{Map}(t, C) \to \mathrm{Map}(t, D)$ are weak equivalences by definition, hence trivial fibrations by the above. As indicated in the diagram

$$
\begin{tikzcd}
\mathrm{Map}(t, C) \arrow[dr, two heads] \arrow[drr, bend left] \arrow[ddr, bend right, two heads, "\sim"'] & & \\
& \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D) \arrow[r] \arrow[d, two heads] \arrow[dr, phantom, "\lrcorner", very near start] & \mathrm{Map}(s, C) \arrow[d, two heads, "\wr"] \\
& \mathrm{Map}(t, D) \arrow[r] & \mathrm{Map}(s, D)
\end{tikzcd}
$$

the map $\mathrm{Map}(t, C) \to \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is a trivial fibration by the two-out-of-three property. □

Lemma 3.13 *Let $\mathbf{C}$ be a cofibration test category and let $s \xrightarrow{\sim} t$ be a trivial cofibration in $\mathbf{C}$. Then any pushout of $s \xrightarrow{\sim} t$ in $\mathrm{Ind}(\mathbf{C})$ is a weak equivalence in the sense of Theorem 3.9.*

Proof The following proof works if $\mathbf{C}$ is a cofibration test category with respect to the Kan–Quillen model structure on **sSet**. The same proof works in the case that $\mathbf{C}$ is a cofibration test category with respect to $\mathbf{sSet}_{\mathrm{J}}$ if one replaces every instance of Δ^1 by H, where H is as in the proof of Lemma 2.1.

We will first show that $i: s \xrightarrow{\sim} t$ is a deformation retract. By item (iv) of Definition 3.3, there exists a lift in

$$\begin{tikzcd} s \arrow[r, equal] \arrow[d, tail, "i", "\wr"'] & s \\ t \arrow[ur, dashed, "r"'] & \end{tikzcd}$$

ie a retract r of i. By (ii) and (iv) of Definition 3.3, there exists a lift F in

$$\begin{tikzcd} s \otimes \Delta^1 \cup_{s \otimes \partial\Delta^1} t \otimes \partial\Delta^1 \arrow[r, "i \cup (\mathrm{id}_t, ir)"] \arrow[d, tail, "\wr"'] & t \\ t \otimes \Delta^1 \arrow[ur, dashed, "F"'] & \end{tikzcd}$$

This lift $F: t \otimes \Delta^1 \to t$ is a deformation retract.

Now let a pushout square

$$\begin{tikzcd} s \arrow[r, "f"] \arrow[d, tail, "i", "\wr"'] & C \arrow[d, "j"] \\ t \arrow[r, "g"] \arrow[ur, phantom, "\ulcorner"] & D \end{tikzcd} \tag{2}$$

be given, where $s \to C$ is any map in $\mathrm{Ind}(\mathbf{C})$. The maps $fr: t \to C$ and $\mathrm{id}_C: C \to C$ give, by the universal property of the pushout, a retract r' of $j: C \to D$. Since tensors preserve colimits, we see that $D \otimes \Delta^1$ is the pushout of $t \otimes \Delta^1$ and $C \otimes \Delta^1$ along $s \otimes \Delta^1$. Then $g \circ F: t \otimes \Delta^1 \to D$ and $C \otimes \Delta^1 \to C \otimes * \cong C \xrightarrow{j} D$ give, by the universal property of the pushout, a map $G: D \otimes \Delta^1 \to D$. Write $\iota_0, \iota_1: D \to D \otimes \Delta^1$ for the endpoint inclusions. It follows from the universal property of the pushout (2) that $\iota_0 G = \mathrm{id}_D$ while $\iota_1 G = jr'$, ie G is a deformation retract.

Now let $u \in \mathbf{T}$ be any test object. We deduce from the existence of the deformation retract G that $\mathrm{Map}(u, C) \to \mathrm{Map}(u, D)$ is the inclusion of a deformation retract, hence a weak equivalence. □

Proof of Theorem 3.9 We check all four assumptions of Kan's recognition theorem as spelled out in [17, Theorem 11.3.1]. The weak equivalences satisfy the two-out-of-three property and are closed under retracts since this holds for the weak equivalences in **sSet**.

(1) Since all objects of $\mathbf{C}$ are compact in $\mathrm{Ind}(\mathbf{C})$, the sets I and J permit the small object argument.

(2) It suffices to prove that any transfinite composition of pushouts of maps in J is a weak equivalence. This follows immediately from Lemmas 3.11 and 3.13.

(3) We need to show that any map having the right lifting property with respect to maps in I has the right lifting property with respect to maps in J and is a weak equivalence. The first of these follows since $J \subseteq I$. To see that any map that has the right lifting property with respect to maps in I is a weak equivalence, let such a map $C \to D$ be given. Note that $t \otimes \partial\Delta[n] \to t \otimes \Delta[n]$ is in I for any $t \in \mathbf{T}$ and $n \geq 0$, by items (i) and (ii) of Definition 3.3. This implies that $\mathrm{Map}(t, C) \to \mathrm{Map}(t, D)$ is a trivial fibration for any $t \in \mathbf{T}$ and in particular that $C \to D$ is a weak equivalence.

(4) We need to show that if $C \to D$ has the right lifting property with respect to maps in J and is a weak equivalence, then it has the right lifting property with respect to maps in I. Let $s \rightarrowtail t$ in I be given. Then $\mathrm{Map}(t, C) \xrightarrow{\sim\!\!\!\!\to} \mathrm{Map}(s, C) \times_{\mathrm{Map}(s,D)} \mathrm{Map}(t, D)$ is a trivial fibration by Lemma 3.12, and in particular surjective on 0–simplices. In particular, $C \to D$ has the right lifting property with respect to $s \rightarrowtail t$.

The fact that this model structure is simplicial follows from Lemma 3.12. By (iv), all objects in $\mathbf{C} \subseteq \mathrm{Ind}(\mathbf{C})$ are fibrant. Since the generating trivial cofibrations are maps between compact objects and any $C \in \mathrm{Ind}(\mathbf{C})$ is a filtered colimit of objects in $\mathbf{C}$, it follows that all objects in $\mathrm{Ind}(\mathbf{C})$ are fibrant. □

Example 3.14 Let $(\mathbf{C}, \mathbf{T})$ be the cofibration test category from Example 3.8. The model structure on $\mathrm{Ind}(\mathbf{C})$ obtained by applying Theorem 3.9 turns out to be Quillen equivalent to the Kan–Quillen model structure on **sSet** and the Quillen model structure on **Top**. More precisely, there is a canonical way to factor the geometric realization functor for simplicial sets $|\cdot|\colon \mathbf{sSet} \to \mathbf{Top}$ as a composite $\mathbf{sSet} \to \mathrm{Ind}(\mathbf{C}) \to \mathbf{Top}$, where both of these functors are left Quillen equivalences. This will be proved in Proposition 4.2.

Example 3.15 For this example, **Top** is again a convenient category of spaces as in Example 3.8. If $\mathcal{P}$ is a topological operad, then the category $\mathcal{P}$–**Alg** of $\mathcal{P}$–algebras admits a model structure, obtained through transfer along the free-forgetful adjunction $F\colon \mathbf{Top} \rightleftarrows \mathcal{P}\text{–}\mathbf{Alg} :U$. In particular, any object is fibrant in this model structure. This category is **Top**–enriched, since one can view $\mathrm{Hom}_{\mathcal{P}\text{–}\mathbf{Alg}}(S, T)$ as a subspace of $\underline{\mathrm{Hom}}_{\mathbf{Top}}(US, UT)$. For any topological space X and any $\mathcal{P}$–algebra S, one can endow the space S^X with the "pointwise" structure of a $\mathcal{P}$–algebra. By restricting the usual homeomorphism coming from the cartesian closed structure on **Top**, we obtain a natural homeomorphism $\underline{\mathrm{Hom}}_{\mathcal{P}\text{–}\mathbf{Alg}}(S, T^X) \cong \underline{\mathrm{Hom}}_{\mathbf{Top}}(X, \underline{\mathrm{Hom}}_{\mathcal{P}\text{–}\mathbf{Alg}}(S, T))$. One can furthermore show that $-^X : \mathcal{P}\text{–}\mathbf{Alg} \to \mathcal{P}\text{–}\mathbf{Alg}$ has a left adjoint that makes $\mathcal{P}$–**Alg** into a tensored and cotensored topological category. In particular, it can be viewed as a tensored and cotensored simplicial category. Since the cotensors, fibrations and weak equivalences are defined underlying in **Top**, we see that $\mathcal{P}$–**Alg** is an $\mathbf{sSet}_{\mathrm{KQ}}$–enriched model category with respect to this enrichment. By Example 3.6, any small full subcategory closed under finite colimits and tensors with finite simplicial sets inherits the structure of a cofibration test category.

Example 3.16 One can modify the previous example in a way that is similar to Example 3.8. Namely, suppose that $\mathbf{C} \subseteq \mathcal{P}\text{–}\mathbf{Alg}$ is a small full subcategory which is closed under finite colimits and tensors by finite simplicial sets, and suppose that $F|X|$ is contained in $\mathbf{C}$ for any finite simplicial set X, where $F\colon \mathbf{Top} \to \mathcal{P}\text{–}\mathbf{Alg}$ is the left adjoint of the free-forgetful adjunction. Define the full subcategory of test objects $\mathbf{T} \subseteq \mathbf{C}$ to be the category of objects of the form $F|X|$ for X a finite simplicial set, and define the (trivial) cofibrations to be the maps of the form $F|i|\colon F|X| \to F|Y|$, where i is a (trivial) cofibration between finite simplicial sets in $\mathbf{sSet}_{\mathrm{KQ}}$. Then $(\mathbf{C}, \mathbf{T})$ is a cofibration test category; hence we obtain a model structure on $\mathrm{Ind}(\mathbf{C})$ by Theorem 3.9. Since the inclusion $\mathbf{C} \hookrightarrow \mathcal{P}\text{–}\mathbf{Alg}$ preserves finite colimits, it induces an adjunction $\mathrm{Ind}(\mathbf{C}) \rightleftarrows \mathcal{P}\text{–}\mathbf{Alg}$. One can show that this adjunction is a Quillen equivalence.

4 Example: a convenient model category of topological spaces

Throughout this section, let **Top** be a convenient category of spaces, such as k–spaces, compactly generated weak Hausdorff spaces or compactly generated Hausdorff spaces. Suppose that a small full subcategory $\mathbf{C} \subseteq \mathbf{Top}$ is given that is closed under finite colimits and tensors with finite simplicial sets, and that contains the space $|X|$ for any finite simplicial set X. As explained in Example 3.8, if we define $\mathbf{T}$ to be the collection of spaces of the form $|X|$, where X is any finite simplicial set, and if we define a map to be a (trivial) cofibration if and only if it is the geometric realization of a (trivial) cofibration in $\mathbf{sSet}_{\mathrm{KQ}}$ between finite simplicial sets, then $(\mathbf{C}, \mathbf{T})$ is a cofibration test category. In this section, we will study this example in more detail.

We begin by characterizing the weak equivalences of $\mathrm{Ind}(\mathbf{C})$. Note that the geometric realization functor $|\cdot|\colon \mathbf{sSet}_{\mathrm{fin}} \to \mathbf{C}$ extends uniquely to a filtered colimit-preserving functor $|\cdot|\colon \mathbf{sSet} \to \mathrm{Ind}(\mathbf{C})$ that has a right adjoint Sing defined by $(\operatorname{Sing} C)_n = \operatorname{Hom}(|\Delta^n|, C)$ for any $C \in \mathrm{Ind}(\mathbf{C})$.

Lemma 4.1 *Let* $(\mathbf{C}, \mathbf{T})$ *be a cofibration test category as above. Then a map* $C \to D$ *in* $\mathrm{Ind}(\mathbf{C})$ *is a weak equivalence if and only if* $\operatorname{Map}(*, C) \to \operatorname{Map}(*, D)$ *is a weak equivalence, where* $*$ *is the terminal object. In particular,* $C \to D$ *is a weak equivalence if and only if* $\operatorname{Sing} C \to \operatorname{Sing} D$ *is a weak equivalence in* $\mathbf{sSet}_{\mathrm{KQ}}$.

Proof If $C \to D$ is a weak equivalence in $\mathrm{Ind}(\mathbf{C})$, then $\operatorname{Map}(*, C) \to \operatorname{Map}(*, D)$ is a weak equivalence by definition. Conversely, suppose that $\operatorname{Map}(*, C) \to \operatorname{Map}(*, D)$ is a weak equivalence and let X be a finite simplicial set. It follows by adjunction that $\operatorname{Map}(|X|, C) \to \operatorname{Map}(|X|, D)$ agrees with $\operatorname{Map}(*, C)^X \to \operatorname{Map}(*, D)^X$, hence this map is a weak equivalence.

For the second statement, note that $\operatorname{Map}(*, E) \cong \operatorname{Hom}(* \otimes \Delta^\bullet, E) \cong \operatorname{Sing} E$. $\square$

The inclusion $\mathbf{C} \hookrightarrow \mathbf{Top}$ induces an adjunction $L\colon \mathrm{Ind}(\mathbf{C}) \rightleftarrows \mathbf{Top} : \widehat{(\cdot)}_{\mathrm{Ind}}$ as explained in Section 2.2, where L is defined by $L(\{c_i\}) = \operatorname{colim}_i c_i$ for any $\{c_i\}$ in $\mathrm{Ind}(\mathbf{C})$. Since geometric realization commutes with colimits, we see that the geometric realization functor $|\cdot|\colon \mathbf{sSet} \to \mathbf{Top}$ factors as $\mathbf{sSet} \xrightarrow{|\cdot|} \mathrm{Ind}(\mathbf{C}) \xrightarrow{L} \mathbf{Top}$.

Proposition 4.2 *Let* $(\mathbf{C}, \mathbf{T})$ *be a cofibration test category as above. The adjunctions* $|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \rightleftarrows \mathrm{Ind}(\mathbf{C}) : \operatorname{Sing}$ *and* $L\colon \mathrm{Ind}(\mathbf{C}) \rightleftarrows \mathbf{Top} : \widehat{(\cdot)}_{\mathrm{Ind}}$ *are Quillen equivalences.*

Proof It is clear from the definition of the (trivial) cofibrations in $(\mathbf{C}, \mathbf{T})$ that

$$|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \to \mathrm{Ind}(\mathbf{C}) \quad \text{and} \quad L\colon \mathrm{Ind}(\mathbf{C}) \to \mathbf{Top}$$

send generating (trivial) cofibrations to (trivial) cofibrations. In particular, they are left Quillen functors. Since the composition of these adjunctions is the well-known Quillen equivalence $|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \rightleftarrows \mathbf{Top} : \operatorname{Sing}$, it suffices to show by the two-out-of-three property for Quillen equivalences that $|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \rightleftarrows \mathrm{Ind}(\mathbf{C}) : \operatorname{Sing}$ is a Quillen equivalence. By Lemma 4.1, a map $C \to D$ in $\mathrm{Ind}(\mathbf{C})$ is a weak equivalence if and only if $\operatorname{Sing} C \to \operatorname{Sing} D$ is. In particular, this adjunction is a Quillen equivalence if and only if the unit $X \to \operatorname{Sing}|X|$ is a weak equivalence for any simplicial set X. If X is a finite simplicial set, then $X \to \operatorname{Sing}|X|$ agrees by definition with the unit of the adjunction $|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \rightleftarrows \mathbf{Top} : \operatorname{Sing}$, which is always a weak equivalence. Since weak equivalences are stable under filtered colimits in $\mathbf{sSet}_{\mathrm{KQ}}$, it follows that the unit $X \to \operatorname{Sing}|X|$ of $|\cdot|\colon \mathbf{sSet}_{\mathrm{KQ}} \rightleftarrows \mathrm{Ind}(\mathbf{C}) : \operatorname{Sing}$ is a weak equivalence for any simplicial set X. $\square$

One can show that the model category $\operatorname{Ind}(\mathbf{C})$, with $(\mathbf{C}, \mathbf{T})$ a cofibration test category of the type considered above, is very similar to **Top**. We mention a few similarities. We first note that it is possible to define homotopy groups for objects of $\operatorname{Ind}(\mathbf{C})$, and that they detect weak equivalences. If C is an object in $\operatorname{Ind}(\mathbf{C})$, then by a basepoint of C we mean a map $* \to C$.

Definition 4.3 The n^{th} homotopy group $\pi_n(C, c_0)$ of an object $C \in \operatorname{Ind}(\mathbf{C})$ and a basepoint $c_0 \colon * \to C$ is defined as the set of pointed maps $|\Delta^n/\partial\Delta^n| \to C$ modulo pointed homotopy.

It follows from this definition that $\pi_n(C, c_0) = \pi_n(\operatorname{Sing} C, c_0)$ for any $C \in \operatorname{Ind}(\mathbf{C})$ and $c_0 \in C$. We conclude the following:

Proposition 4.4 *A map $f \colon C \to D$ in $\operatorname{Ind}(\mathbf{C})$ is a weak equivalence if and only if $\pi_n(C, c_0) \to \pi_n(D, f(c_0))$ is a bijection for any $c_0 \in C$ and $n \geq 0$. Moreover, the homotopy groups for objects in $\operatorname{Ind}(\mathbf{C})$ commute with filtered colimits.*

Proof The first statement follows since f is a weak equivalence if and only if $\operatorname{Sing} f$ is. The second part follows since both the functor Sing and the homotopy groups of simplicial sets commute with filtered colimits. □

It can also be shown that one can take the same generating (trivial) cofibrations in $\operatorname{Ind}(\mathbf{C})$ as in the usual Quillen model structure on **Top**. Define

$$\begin{aligned} I &= \{\partial D^n \hookrightarrow D^n \mid n \geq 0\}, \\ J &= \{D^n \times \{0\} \hookrightarrow D^n \times [0,1] \mid n \geq 0\}. \end{aligned}$$

Proposition 4.5 *The sets I and J are sets of generating cofibrations and generating trivial cofibrations for $\operatorname{Ind}(\mathbf{C})$, respectively.*

Proof We need to show that the geometric realization of any cofibration (resp. trivial cofibration) between finite simplicial sets lies in the saturation of I (resp. J). This follows from the fact that $|\cdot| \colon \mathbf{sSet} \to \operatorname{Ind}(\mathbf{C})$ preserves colimits and that each map of the form $|\partial\Delta^n| \to |\Delta^n|$ (resp. $|\Lambda^n_k| \to |\Delta^n|$) is isomorphic to a map in I (resp. J). □

For any two objects $C = \{c_i\}_i$ and $D = \{d_j\}_j$ in $\operatorname{Ind}(\mathbf{C})$, one can compute their product levelwise by $C \times D = \{c_i \times d_j\}_{(i,j) \in I \times J}$. Since the finite colimits of $\mathbf{C}$ are computed in **Top**, and finite colimits in $\operatorname{Ind}(\mathbf{C})$ can be computed levelwise, we see that the functor $- \times D \colon \mathbf{C} \to \operatorname{Ind}(\mathbf{C})$ preserves finite colimits for any D in $\operatorname{Ind}(\mathbf{C})$. As explained in

Section 2.2, it follows from this that the product functor $-\times D\colon \mathrm{Ind}(\mathbf{C}) \to \mathrm{Ind}(\mathbf{C})$ has a right adjoint. In particular, $\mathrm{Ind}(\mathbf{C})$ is cartesian closed. This cartesian closed structure interacts well with the model structure defined above.

Proposition 4.6 $\mathrm{Ind}(\mathbf{C})$ *is a cartesian closed model category.*

Proof It suffices to show that for any pair of generating cofibrations $C \rightarrowtail D$ and $C' \rightarrowtail D'$, the pushout-product

$$C \times D' \cup_{C \times C'} D \times C' \to D \times D'$$

is a cofibration that is trivial if either $C \rightarrowtail D$ or $C' \rightarrowtail D'$ is. This is clearly true. $\square$

One can furthermore show that the full subcategory of **Top** on the CW–complexes embeds fully faithfully into $\mathrm{Ind}(\mathbf{C})$, by using that any finite CW–complex X is (homeomorphic to) an object in $\mathbf{C}$.

Proposition 4.7 *There is a fully faithful functor from the category of CW–complexes into* $\mathrm{Ind}(\mathbf{C})$ *that preserves and detects weak equivalences.*

Proof If X is a CW–complex, then one can always choose a CW–decomposition. The finite CW–subcomplexes in this decomposition together with their inclusions form a directed diagram $\{X_i\}$ for which $\operatorname{colim}_i X_i \cong X$. Suppose that we have chosen a CW–decomposition for any CW–complex X, and denote the associated directed diagram of finite CW–subcomplexes by $\{X_{i_X}\}$. Since a map from a compact space into a CW–complex (with a given CW–decomposition) always lands in a finite CW–subcomplex, we see that the canonical map

$$\lim_{i_X} \operatorname*{colim}_{i_Y} \mathrm{Hom}(X_{i_X}, Y_{i_Y}) \to \mathrm{Hom}(X, Y)$$

is an isomorphism for any pair of CW–complexes. By definition of the morphisms in $\mathrm{Ind}(\mathbf{C})$, this implies that the functor that sends a CW–complex X to the ind-object $\{X_{i_X}\}$ in $\mathrm{Ind}(\mathbf{C})$ is well-defined and fully faithful. Preservation and detection of weak equivalences follows directly from the fact that Sing detects weak equivalences and that $\operatorname{colim}_{i_X} \mathrm{Sing}(X_{i_X}) \cong \mathrm{Sing}(X)$ for any CW–complex X. $\square$

We end this section by discussing a specific example of such a full subcategory $\mathbf{C}$, namely the category **CM** of compact metrizable spaces. Under Gelfand–Naimark duality, this category corresponds to the category of separable commutative unital C^*–algebras. If we let **Top** be the category of compactly generated Hausdorff spaces, then $\mathbf{CM}$ as a full subcategory is closed under all finite colimits and tensors by finite simplicial sets. In particular, by the above we obtain a model structure on $\mathrm{Ind}(\mathbf{CM})$ that is equivalent to the Quillen model structure on $\mathbf{Top}$. In [3], Barnea also proposes a model structure on $\mathrm{Ind}(\mathbf{CM})$. However, this model structure does not agree with the one constructed above, so we will briefly describe his model structure and the difference with ours. Let us denote our model structure by $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}}$.

Barnea shows in [3] that $\mathbf{CM}$ is a "special weak cofibration category", and hence that there exists an induced model structure on $\mathrm{Ind}(\mathbf{CM})$, which we will denote by $\mathrm{Ind}(\mathbf{CM})_B$. This model structure is cofibrantly generated and one can take the set of Hurewicz cofibrations in $\mathbf{CM}$ as a set of generating cofibrations, while one can take the Hurewicz cofibrations that are also homotopy equivalences as a set of generating trivial cofibrations. If we define $\mathbf{T} = \mathbf{CM}$ and if we define a map in $\mathbf{T}$ to be a (trivial) cofibration if it is in the set of generating (trivial) cofibrations just mentioned, then $(\mathbf{CM}, \mathbf{T})$ is a cofibration test category and the completed model structure on $\mathrm{Ind}(\mathbf{CM})$ coincides with the one that Barnea constructed. Since Barnea's model structure $\mathrm{Ind}(\mathbf{CM})_{\mathrm{B}}$ has strictly more generating (trivial) cofibrations than our model structure $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}}$, we see that the identity functor is a left Quillen functor $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}} \to \mathrm{Ind}(\mathbf{CM})_{\mathrm{B}}$. To see that the model structures do not coincide, we will show that $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}}$ has strictly more weak equivalences than $\mathrm{Ind}(\mathbf{CM})_{\mathrm{B}}$. Let C be any metrizable infinite Stone space, such as a Cantor space. Then, for Sing and $|\cdot|$ as defined just above Lemma 4.1, the counit $|\mathrm{Sing}\, C| \to C$ is a weak equivalence in $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}}$. However, this map is not a weak equivalence in $\mathrm{Ind}(\mathbf{CM})_{\mathrm{B}}$, since $\mathrm{Map}(C, |\mathrm{Sing}\, C|) \to \mathrm{Map}(C, C)$ is not a weak equivalence of simplicial sets: since these mapping spaces are discrete, this would imply that the map is an isomorphism. However, it is not surjective since there is no map $C \to |\mathrm{Sing}\, C|$ that gets mapped to id_C. The model structure $\mathrm{Ind}(\mathbf{CM})_{\mathrm{Q}}$ defined here is similar to the Quillen model structure on $\mathbf{Top}$, while Barnea's model structure $\mathrm{Ind}(\mathbf{CM})_{\mathrm{B}}$ bears some similarity to the Strøm model structure on $\mathbf{Top}$.

5 The dual model structure on Pro(C)

A model structure on $\mathcal{E}$ also gives rise to a model structure on $\mathcal{E}^{\mathrm{op}}$, where the fibrations (resp. cofibrations) of $\mathcal{E}^{\mathrm{op}}$ are the cofibrations (resp. fibrations) of $\mathcal{E}$. In particular, $\mathcal{E}$ is cofibrantly generated if and only if $\mathcal{E}^{\mathrm{op}}$ is fibrantly generated. Since $\mathrm{Pro}(\mathbf{C}) \simeq \mathrm{Ind}(\mathbf{C}^{\mathrm{op}})^{\mathrm{op}}$, this implies that if $\mathbf{C}$ is the dual of a cofibration test category, then $\mathrm{Pro}(\mathbf{C})$ admits a fibrantly generated simplicial model structure. We explicitly dualize the main definition and result of Section 3 in this section, and then discuss a few examples of such fibrantly generated simplicial model structures on pro-categories. Again, we work with **sSet** endowed with either the Joyal or the Kan–Quillen model structure.

We say that a simplicial category $\mathbf{C}$ is *finitely cotensored* if $\mathbf{C}^{\mathrm{op}}$ is finitely tensored in the sense of Definition 3.1. Explicitly, this means that $\mathbf{C}$ admits finite limits and cotensors by finite simplicial sets, and that these commute with each other. As explained in Section 2.2, if $\mathbf{C}$ is a small simplicial category that is finitely cotensored, then the simplicial category $\mathrm{Pro}(\mathbf{C})$ is tensored, cotensored, complete and cocomplete.

Definition 5.1 A *fibration test category* $(\mathbf{C}, \mathbf{T})$ consists of a small finitely cotensored simplicial category $\mathbf{C}$, a full subcategory $\mathbf{T} \subseteq \mathbf{C}$ of *test objects* and two classes of maps in $\mathbf{T}$ called *fibrations*, denoted by $\twoheadrightarrow$, and *trivial fibrations*, denoted by $\xrightarrow{\sim}\!\!\!\!\twoheadrightarrow$, both containing all isomorphisms, that satisfy the following properties:

- (i) The terminal object $*$ is a test object, and for every test object $t \in \mathbf{T}$, the map $t \to *$ is a fibration.
- (ii) For every fibration between test objects $s \twoheadrightarrow t$ and cofibration between finite simplicial sets $U \rightarrowtail V$, the pullback-power map $s^V \to s^U \times_{t^U} t^V$ is a fibration between test objects, which is trivial if either $s \twoheadrightarrow t$ or $U \rightarrowtail V$ is.
- (iii) A morphism $c \to d$ in $\mathbf{T}$ is a trivial fibration if and only if it is a fibration and $\mathrm{Map}(d, t) \to \mathrm{Map}(c, t)$ is a weak equivalence of simplicial sets for every $t \in \mathbf{T}$.
- (iv) Any object $c \in \mathbf{C}$ has the left lifting property with respect to trivial fibrations.

For a fibration test category $\mathbf{C}$, we write $\mathrm{fib}(\mathbf{C})$ for the set of fibrations and $\mathrm{we}(\mathbf{C})$ for the set of maps $c \to d$ that induce a weak equivalence $\mathrm{Map}(d, t) \to \mathrm{Map}(c, t)$ for every $t \in \mathbf{T}$. By property (iii), the set of trivial fibrations is $\mathrm{fib}(\mathbf{C}) \cap \mathrm{we}(\mathbf{C})$. Note that the definition of a fibration test category is formally dual to that of a cofibration test category. More precisely, $(\mathbf{C}, \mathbf{T})$ is a fibration test category if and only if $(\mathbf{C}^{\mathrm{op}}, \mathbf{T}^{\mathrm{op}})$ is a cofibration test category in the sense of Definition 3.3, where the (trivial) cofibrations of $(\mathbf{C}^{\mathrm{op}}, \mathbf{T}^{\mathrm{op}})$ are defined as the (trivial) fibrations of $(\mathbf{C}, \mathbf{T})$.

Let $P \subseteq \mathrm{Ar}(\mathrm{Pro}(\mathbf{C}))$ denote the image of the set $\mathrm{fib}(\mathbf{C})$ along the inclusion $\mathbf{C} \hookrightarrow \mathrm{Pro}(\mathbf{C})$, and $Q \subseteq \mathrm{Ar}(\mathrm{Pro}(\mathbf{C}))$ the image of the set of trivial fibrations. The sets P and Q are the generating (trivial) fibrations of the completed model structure on $\mathrm{Pro}(\mathbf{C})$. The following theorem is formally dual to Theorem 3.9.

Theorem 5.2 *Let* $(\mathbf{C}, \mathbf{T})$ *be a fibration test category. Then* $\mathrm{Pro}(\mathbf{C})$ *carries a fibrantly generated (hence cocombinatorial) simplicial model structure, the* ***completed model structure****, where a map* $C \to D$ *is a weak equivalence if and only if* $\mathrm{Map}(D, t) \to \mathrm{Map}(C, t)$ *is a weak equivalence for every* $t \in \mathbf{T}$*. A set of generating fibrations (resp. generating trivial fibrations) is given by* P *(resp.* Q*). Every object is cofibrant in this model structure.*

Example 5.3 Dualizing Example 3.6, we see that if $\mathcal{E}$ is a simplicial model category in which every object is cofibrant, then any small full subcategory $\mathbf{C} \subseteq \mathcal{E}$ which is closed under finite limits and finite cotensors admits the structure of a fibration test category. Namely, defining $\mathbf{T}$ to be the full subcategory of fibrant objects of $\mathbf{C}$, and defining the (trivial) fibrations to be those of $\mathcal{E}$ between objects in $\mathbf{T}$, then $(\mathbf{C}, \mathbf{T})$ is a fibration test category. As in Example 3.6, will say that $\mathbf{C}$ *inherits* the structure of a fibration test category from $\mathcal{E}$.

Remark 5.4 For the fibration test category $(\mathbf{C}, \mathbf{T})$ from the previous example, the completed model structure on $\mathrm{Pro}(\mathbf{C})$ is a special case of Theorem 1.1, namely the case where $\mathbf{T}$ is the collection of all fibrant objects in $\mathbf{C}$. By (the dual of) Example 3.7, it follows that we can take $\mathbf{T}$ to be any collection of fibrant objects in $\mathbf{C}$ that is closed under "finite pullback-powers". The general case, where we let $\mathbf{T}$ be any collection of fibrant objects in $\mathbf{C}$, is discussed in Section 8.

Example 5.5 Recall that we call a simplicial set *lean* if it is degreewise finite and coskeletal. The full subcategory of $\mathbf{sSet}$ spanned by all lean simplicial sets $\mathbf{L}$ is closed under finite limits and finite cotensors. By Example 5.3, it inherits the structure of a fibration test category from $\mathbf{sSet}_{\mathrm{KQ}}$, which we will denote by $\mathbf{L}_{\mathrm{KQ}}$. By Theorem 5.2 we obtain a model structure on $\mathrm{Pro}(\mathbf{L})$. Since this category is equivalent to the category of simplicial profinite sets $\mathbf{s}\widehat{\mathbf{Set}}$ by (the dual of) Theorem 2.3, we in particular obtain a simplicial model structure on $\mathbf{s}\widehat{\mathbf{Set}}$. This model structure coincides with Quick's model structure for profinite spaces [31], as explained in Corollary 6.6 below. We denote it by $\mathbf{s}\widehat{\mathbf{Set}}_{\mathrm{Q}}$.

Example 5.6 Consider the full simplicial subcategory $\mathbf{T}_p$ of $\mathbf{sSet}$ whose objects are those lean Kan complexes that have finite p–groups as homotopy groups. One can show that $\mathbf{T}_p$ is closed under "finite pullback-powers", so by the previous example and the dual of Example 3.7, we obtain a fibration test category $\mathbf{L}_p = (\mathbf{L}, \mathbf{T}_p)$ in which the (trivial) fibrations are the (trivial) Kan fibrations between objects of $\mathbf{T}_p$. It is proved in Corollary 6.7 that the completed model structure on $\mathrm{Pro}(\mathbf{L}_p)$ agrees with Morel's model structure for pro-p spaces [30].

Example 5.7 The category of lean simplicial sets also inherits the structure of a fibration test category from the Joyal model structure $\mathbf{sSet}_{\mathrm{J}}$, which we will denote by $\mathbf{L}_{\mathrm{J}}$. The corresponding model structure on $\widehat{\mathbf{sSet}}$ obtained from Theorem 5.2 will be called the *profinite Joyal model structure*, and its fibrant objects will be called *profinite quasicategories*. We will come back to this model category in Section 9, and we will describe its underlying ∞–category in Remark A.11.

Example 5.8 In [16], Haine defines the Joyal–Kan model structure on $\mathbf{sSet}_{/P}$, where P is (the nerve of) a poset. This model category describes the homotopy theory of P–stratified spaces. Since it is a left Bousfield localization of the Joyal model structure on $\mathbf{sSet}_{/P}$, any object is cofibrant and it is an $\mathbf{sSet}_{\mathrm{J}}$–enriched model category. Actually, this model structure can be shown to be $\mathbf{sSet}_{\mathrm{KQ}}$–enriched [16, Sections 2.4–2.5]. In particular, any small full subcategory $\mathbf{C}$ closed under finite limits and cotensors by finite simplicial sets inherits the structure of a fibration test category. If P is a finite poset and $\mathbf{C} = \mathbf{L}_{/P}$ is the full subcategory of lean simplicial sets over (the nerve of) P, then one can show that $\mathrm{Pro}(\mathbf{L}_{/P}) \cong \widehat{\mathbf{sSet}}_{/P}$. In particular, by Theorem 5.2, we obtain a model structure on $\widehat{\mathbf{sSet}}_{/P}$ that is $\mathbf{sSet}_{\mathrm{KQ}}$–enriched. It is shown in Example A.8 that the underlying ∞–category of this model category is the ∞–category of profinite P–stratified spaces defined in [8].

Example 5.9 We call a groupoid finite if it has finitely many arrows (including the identity arrows). The category of finite groupoids **FinGrpd** inherits the structure of a fibration test category from the canonical model structure on **Grpd** [1, Section 5]. (Note that **Grpd** can be viewed as an $\mathbf{sSet}_{\mathrm{KQ}}$–enriched model structure by defining $\mathrm{Map}(A, B) = N(\mathrm{Fun}(A, B))$ for any $A, B \in \mathbf{Grpd}$.) The completed model structure on the category of profinite groupoids $\widehat{\mathbf{Grpd}} = \mathrm{Pro}(\mathbf{FinGrpd})$ obtained from Theorem 5.2 coincides with the model structure for profinite groupoids defined by Horel in [18, Section 4]. To see this, note that Horel shows in [18, Section 4] that the Barnea–Schlank model structure on $\widehat{\mathbf{Grpd}}$ exists and coincides with his model structure. By Remark 5.12 below, the Barnea–Schlank model structure on $\widehat{\mathbf{Grpd}}$ must coincide with our model structure. In particular, Horel's model structure agrees with the one that we construct in this example.

Example 5.10 Similarly, we call a category finite if it has finitely many arrows. The category of all small categories admits the canonical model structure, defined for example in [34]. Since this model structure is $\mathbf{sSet}_{\mathrm{J}}$–enriched, the category of finite categories **FinCat** inherits the structure of a fibration test category. By Theorem 5.2, we obtain an $\mathbf{sSet}_{\mathrm{J}}$–enriched model structure on $\widehat{\mathbf{Cat}} = \mathrm{Pro}(\mathbf{FinCat})$, which we will call the *model structure for profinite categories*.

Example 5.11 Let **bisSet** be endowed with the Reedy model structure with respect to the Kan–Quillen model structure on **sSet**. Recall that the category of bisimplicial profinite sets $\mathbf{bis\widehat{Set}}$ is equivalent to $\mathrm{Pro}(\mathbf{L}^{(2)})$, where $\mathbf{L}^{(2)}$ denotes the category of doubly lean bisimplicial sets defined at the end of Section 2.2. Since any object in **bisSet** is cofibrant, $\mathbf{L}^{(2)}$ inherits the structure of a fibration test category from the Reedy model structure on **bisSet**. By applying Theorem 5.2, we obtain a model structure on $\mathrm{Pro}(\mathbf{L}^{(2)}) \simeq \mathbf{bis\widehat{Set}}$. This model structure coincides with the Reedy model structure on $\mathbf{bis\widehat{Set}}$ with respect to the Quick model structure on $\mathbf{s\widehat{Set}}$, as will be shown in Proposition 6.9.

Remark 5.12 As discussed in the introduction, there are similarities between our construction of a model structure on $\mathrm{Pro}(\mathbf{C})$ and the construction of Barnea–Schlank in [7]. Suppose $\mathbf{C}$ is a fibration test category in the sense of Definition 5.1. Then $\mathbf{C}$ comes with a set $\mathrm{fib}(\mathbf{C})$ of fibrations and a set of $\mathrm{we}(\mathbf{C})$ of weak equivalences. It is very unlikely that the triple $(\mathbf{C}, \mathrm{fib}(\mathbf{C}), \mathrm{we}(\mathbf{C}))$ is a "weak fibration category" in the sense of Definition 1.2 of [7]. Namely, that definition asks that $\mathrm{fib}(\mathbf{C})$ contain all isomorphisms of $\mathbf{C}$, that it be closed under composition, and that a pushout of a map in $\mathrm{fib}(\mathbf{C})$ be again in $\mathrm{fib}(\mathbf{C})$. However, if we define $\mathrm{fib}'(\mathbf{C})$ to be the smallest set that contains $\mathrm{fib}(\mathbf{C})$ and that satisfies these properties, then $(\mathbf{C}, \mathrm{fib}'(\mathbf{C}), \mathrm{we}(\mathbf{C}))$ might be a weak fibration category. If this is the case, then the "induced" model structure on $\mathrm{Pro}(\mathbf{C})$, in the sense of Theorem 1.8 of [7], could exist. The cofibrations of this model structure are defined as the maps that have the left lifting property with respect to $\mathrm{fib}'(\mathbf{C}) \cap \mathrm{we}(\mathbf{C})$, while the trivial cofibrations are the maps that have the left lifting property with respect to $\mathrm{fib}'(\mathbf{C})$. Since the maps in $\mathrm{fib}'(\mathbf{C})$ are clearly fibrations in our construction of the "completed model structure" on $\mathrm{Pro}(\mathbf{C})$ (see Theorem 5.2), we conclude that the (trivial) cofibrations for both model structures must agree. In particular, if both our model structure and the Barnea–Schlank model structure of [7] exist on $\mathrm{Pro}(\mathbf{C})$, then they must coincide. An example where this happens is when $\mathbf{C} = \mathbf{FinGrpd}$. (See Example 5.9 above.)

6 Comparison to some known model structures

As stated in Theorem 5.2, for any fibration test category $(\mathbf{C}, \mathbf{T})$, all objects in the completed model structure on $\mathrm{Pro}(\mathbf{C})$ are cofibrant. We will now show that, in the case that $\mathbf{C}$ is the category $\mathbf{L}$ of lean simplicial sets, for many choices of $\mathbf{T}$ this statement can be strengthened to say that the cofibrations are exactly the monomorphisms. We show how this can be used to prove that the model structures on $\mathbf{s}\widehat{\mathbf{Set}}$ obtained in Examples 5.5 and 5.6 agree with Quick's model structure and Morel's model structure, respectively. It will also follow that the cofibrations in the profinite Joyal model structure from Example 5.7 are exactly the monomorphisms. We conclude this section by showing that the model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ from Example 5.11 agrees with the Reedy model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ with respect to Quick's model structure on $\mathbf{s}\widehat{\mathbf{Set}}$. The main result about cofibrations in $\mathbf{s}\widehat{\mathbf{Set}}$ is the following:

Proposition 6.1 *Let $\mathbf{L}$ be the category of lean simplicial sets endowed with the structure of a fibration test category. Suppose that for any contractible lean Kan complex K, the map $K \to *$ is a trivial fibration in $\mathbf{L}$, and further that any trivial fibration $L \xrightarrow{\sim} K$ in $\mathbf{L}$ is a trivial Kan fibration. Then the cofibrations in the completed model structure on $\mathrm{Pro}(\mathbf{L}) \simeq \mathbf{s}\widehat{\mathbf{Set}}$ are the monomorphisms.*

This proposition clearly applies to the fibration test categories $\mathbf{L}_{\mathrm{KQ}}$, $\mathbf{L}_p$ and $\mathbf{L}_{\mathrm{J}}$ of Examples 5.5, 5.6 and 5.7. The following lemmas will be used to prove this result. Recall that the category $\widehat{\mathbf{Set}}$ of profinite sets is equivalent to the category $\mathbf{Stone}$ of Stone spaces.

Lemma 6.2 *A map of profinite sets (resp. simplicial profinite sets) $S \to T$ is a monomorphism if and only if it is (isomorphic to) the limit of a cofiltered diagram $\{S_i \rightarrowtail T_i\}_{i \in I}$ consisting of monomorphisms between finite sets (resp. degreewise finite simplicial sets).*

Proof In the category of Stone spaces, the monomorphisms are precisely the injective continuous maps. Since a cofiltered limit of injective maps is again injective, we see that if $S \to T$ is an inverse limit of monomorphisms $S_i \rightarrowtail T_i$, then $S \to T$ is itself a monomorphism.

Conversely, suppose that $S \to T$ is a monomorphism of profinite sets (resp. simplicial profinite sets). Write $T = \lim_i T_i$ as a cofiltered limit of finite sets (resp. lean simplicial sets), and, for every i, write S'_i for the image of the composition $S \to T \to T_i$. Then

$\{S'_i\}_{i\in I}$ is a cofiltered diagram since the structure maps $T_i \to T_j$ restrict to maps $S'_i \to S'_j$ for any $i \to j$ in I. Since $\{S'_i \to T_i\}_{i\in I}$ is levelwise a monomorphism, the proof is complete if we can show that $S \to \lim_i S'_i$ is an isomorphism. Since isomorphisms of Stone spaces are detected on the underlying sets, it suffices to show that this map is both injective and surjective. It is injective since the composition $S \to \lim_i S'_i \to T$ is, while it is surjective by [36, Corollary 1.1.6]. $\square$

We will denote the two-element set $\{0, 1\}$ by $\mathbf{2}$.

Lemma 6.3 *A map of (profinite) sets $S \to T$ is a monomorphism if and only if it has the left lifting property with respect to $\mathbf{2} \to *$.*

Proof We leave the case where $S \to T$ is a map of sets to the reader. For the "if" direction in the profinite case, suppose that $f : S \to T$ has the left lifting property with respect to $\mathbf{2} \to *$, but is not a monomorphism. Regarding S and T as Stone spaces, there must exist distinct $s, s' \in S$ such that $f(s) = f(s')$. Choose some clopen $U \subseteq S$ such that $s \in U$ and $s' \notin U$. Then the indicator function $\mathbb{1}_U : S \to \mathbf{2}$ is continuous but does not extend to a map $T \to \mathbf{2}$. We conclude that $S \to T$ must be a monomorphism.

For the converse, note that by Lemma 6.2 we may assume without loss of generality that $S \to T$ can be represented by levelwise monomorphisms $\{S_i \to T_i\}$. Since $\mathbf{2}$ is cocompact in $\widehat{\mathbf{Set}}$, any map $S \to \mathbf{2}$ factors through S_i for some i. Since $S_i \to T_i$ is a monomorphism of sets, the result follows. $\square$

Consider the diagram

$$\begin{tikzcd} * \arrow[r, "\mathbf{2}"] \arrow[d, "{[n]}"'] & \mathbf{FinSet} \\ \Delta^{\mathrm{op}} \arrow[ur, dashed, "R_n\mathbf{2}"'] & \end{tikzcd}$$

where $[n]$ denotes the inclusion of the terminal category $*$ into Δ^{op} at $[n]$, and $\mathbf{2}$ denotes the inclusion of $*$ into $\mathbf{FinSet}$ at the two-element set $\mathbf{2}$. Since $\mathbf{FinSet}$ has all finite limits, the right Kan extension $R_n\mathbf{2}$ exists. Since the inclusion $* \hookrightarrow \Delta^{\mathrm{op}}$ factors through $\Delta^{\mathrm{op}}_{\le n}$, the simplicial set $R_n\mathbf{2}$ is n–coskeletal. In particular, it is a lean simplicial set.

Lemma 6.4 *A map of simplicial (profinite) sets is a monomorphism if and only if it has the left lifting property with respect to $R_n\mathbf{2} \to *$ for every $n \in \mathbb{N}$.*

Proof Since the inclusions of **FinSet** into **Set** and $\widehat{\mathbf{Set}}$ both preserve limits, we see that the lean simplicial set $R_n\mathbf{2}$ constructed above is also the right Kan extension of $* \xrightarrow{[n]} \Delta^{\mathrm{op}}$ along $* \xrightarrow{\mathbf{2}} \mathbf{Set}$ and along $* \xrightarrow{\mathbf{2}} \widehat{\mathbf{Set}}$. In particular, a map of simplicial (profinite) sets $X \to Y$ has the left lifting property with respect to $R_n\mathbf{2} \to *$ if and only if $X_n \to Y_n$ has the left lifting property with respect to $\mathbf{2} \to *$, hence the result follows from Lemma 6.3. □

Proof of Proposition 6.1 We first show that any cofibration in the model category $\mathrm{Pro}(\mathbf{L})$ is a monomorphism. Since $R_n\mathbf{2} \to *$ has the right lifting property with respect to all monomorphisms in **sSet**, we see that it is a trivial Kan fibration, hence by assumption a trivial fibration in the fibration test category $\mathbf{L}$ and a generating trivial fibration in $\mathrm{Pro}(\mathbf{L})$. By Lemma 6.4, any cofibration in $\mathrm{Pro}(\mathbf{L}) \simeq \mathbf{s}\widehat{\mathbf{Set}}$ is a monomorphism.

For the converse, suppose $X \to Y$ is a monomorphism in $\mathrm{Pro}(\mathbf{L})$. By Lemma 6.2, we may assume that $X \to Y$ is a cofiltered limit of monomorphisms between degreewise finite simplicial sets $\{X_i \to Y_i\}_{i\in I}$. We see that for every i, the map $X_i \to Y_i$ has the left lifting property with respect to the generating trivial fibrations of $\mathrm{Pro}(\mathbf{L})$ since these are trivial Kan fibrations between lean simplicial sets. Since any generating trivial fibration is a map between cocompact objects, it follows that $X \to Y$ also has the left lifting property with respect to the generating trivial fibrations. □

Proposition 6.1 shows that, for $\mathbf{L}_{\mathrm{KQ}}$ and $\mathbf{L}_p$ the fibration test categories of Examples 5.5 and 5.6, the cofibrations of the model categories $\mathrm{Pro}(\mathbf{L}_{\mathrm{KQ}})$ and $\mathrm{Pro}(\mathbf{L}_p)$ are the monomorphisms. This means that the cofibrations coincide with those of Quick's model structure [31] and Morel's model structure [30], respectively. The same is true for the weak equivalences. This follows from the results in Section 7 of [4] (most notably Lemmas 7.4.7 and 7.4.10), using that Quick's and Morel's model structures on $\mathbf{s}\widehat{\mathbf{Set}}$ are simplicial. We state this explicitly as follows:

Proposition 6.5 [4] *A map $X \to Y$ of simplicial profinite sets is a weak equivalence in Quick's model structure if and only if* $\mathrm{Map}(Y, K) \to \mathrm{Map}(X, K)$ *is a weak equivalence for any lean Kan complex K. It is a weak equivalence in Morel's model structure if and only if* $\mathrm{Map}(Y, K) \to \mathrm{Map}(X, K)$ *is a weak equivalence for any lean Kan complex K whose homotopy groups are finite p–groups.*

From this proposition and the definition of the completed model structure (Theorem 5.2), we see that the weak equivalences of $\mathrm{Pro}(\mathbf{L}_{\mathrm{KQ}})$ (resp. $\mathrm{Pro}(\mathbf{L}_p)$) agree with the weak equivalences in Quick's model structure (resp. Morel's model structure) on $\mathbf{s}\widehat{\mathbf{Set}}$.

Corollary 6.6 *The completed model structure on* $\mathrm{Pro}(\mathbf{L}_{\mathrm{KQ}})$ *coincides with Quick's model structure.*

Corollary 6.7 *For any prime number* p*, the completed model structure on* $\mathrm{Pro}(\mathbf{L}_p)$ *coincides with Morel's model structure.*

The proof of Proposition 6.1 admits an analogue for bisimplicial sets (in fact, for the category of presheaves on K for any small category K that can be written as a union of finite full subcategories), which we leave as an exercise to the reader.

Proposition 6.8 *Let* **bisSet** *be endowed with a simplicial model structure in which the cofibrations are the monomorphisms, and let* $\mathbf{L}^{(2)}$ *be the full subcategory of doubly lean bisimplicial sets, which inherits the structure of a fibration test category in the sense of Example 5.3. Then the cofibrations in* $\mathrm{Pro}(\mathbf{L}^{(2)}) \simeq \mathbf{bis}\widehat{\mathbf{Set}}$ *are the monomorphisms.*

Note that this proposition implies that the cofibrations in the model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ from Example 5.11 are exactly the monomorphisms. We will show that, in fact, this model structure coincides with the Reedy model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ with respect to $\mathbf{s}\widehat{\mathbf{Set}}_{\mathrm{Q}}$. We do this by inspecting the generating (trivial) fibrations of the Reedy model structure. For the following proof, note that Quick's model structure coincides with the completed model structure on $\mathrm{Pro}(\mathbf{L}_{\mathrm{KQ}})$ by Corollary 6.6, hence that any (trivial) Kan fibration between lean Kan complexes is a (trivial) fibration in Quick's model structure.

Proposition 6.9 *The model structure on* $\mathrm{Pro}(\mathbf{L}^{(2)})$ *of Example 5.11 coincides with the Reedy model structure on* $\mathbf{bis}\widehat{\mathbf{Set}}$ *(with respect to* $\mathbf{s}\widehat{\mathbf{Set}}_{\mathrm{Q}}$*).*

Proof Note that if $L \to K$ is a (trivial) Reedy fibration between Reedy fibrant doubly lean bisimplicial sets, then L_n and $M_n L \times_{M_n K} K_n$ are lean Kan complexes for every n. In particular, the map $L_n \to M_n L \times_{M_n K} K_n$ is a (trivial) fibration between lean Kan complexes for every n. This shows that any generating (trivial) fibration in $\mathrm{Pro}(\mathbf{L}^{(2)})$ is a (trivial) fibration in the Reedy model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$.

For the converse, note that the Reedy model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ is fibrantly generated. Its generating (trivial) fibrations are maps of the form

$$\mathbf{G}_n L \to \partial \mathbf{G}_n L \times_{\partial \mathbf{G}_n K} \mathbf{G}_n K \tag{3}$$

for any $n \geq 0$, where $\mathbf{G}_n$ is the right adjoint to the functor $X \mapsto X_n$, while $\partial \mathbf{G}_n$ is the right adjoint to the latching object functor $X \mapsto L_n X$, and $L \to K$ is a generating (trivial) fibration in $\mathbf{s}\widehat{\mathbf{Set}}$. It can be shown using the right adjointness of $\mathbf{G}_n$ and $\partial \mathbf{G}_n$

that these functors restrict to functors $\mathbf{L} \to \mathbf{L}^{(2)}$. One can furthermore deduce from the adjointness that if L and K are fibrant in **sSet**, then both the domain and codomain of the map (3) are Reedy fibrant in $\mathbf{bis}\widehat{\mathbf{Set}}$ and hence in **bisSet**. This shows that any map of the form (3), with $L \to K$ a (trivial) fibration in $\mathbf{L}$, is a (trivial) fibration in $\mathbf{L}^{(2)}$. In particular, any generating (trivial) fibration in the Reedy model structure on $\mathbf{bis}\widehat{\mathbf{Set}}$ is a (trivial) fibration in $\mathrm{Pro}(\mathbf{L}^{(2)})$. We conclude that both model structures coincide. $\square$

7 Quillen pairs

As explained in Section 2.2, there is an easy criterion for constructing adjunctions between ind-categories: if $\mathbf{C}$ is a small category that admits finite colimits and if $\mathcal{E}$ is any cocomplete category, then a functor $F\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ has a right adjoint if and only if it preserves all colimits. Furthermore, these functors correspond to functors $\mathbf{C} \to \mathcal{E}$ that preserve all finite colimits. There is a dual criterion for pro-categories. In the simplicial case, this can be strengthened as in the following lemma.

If $\mathcal{E}$ is a tensored cocomplete simplicial category, then we say that colimits and tensors commute in $\mathcal{E}$ if the analogue of item (iii) of Definition 3.1 holds for all diagrams in $\mathcal{E}$ and all simplicial sets.

Lemma 7.1 *Let $\mathbf{C}$ be a small finitely tensored simplicial category and let $\mathcal{E}$ be a tensored cocomplete simplicial category in which colimits and tensors commute. Then any simplicial functor $F\colon \mathbf{C} \to \mathcal{E}$ that preserves finite colimits and tensors with finite simplicial sets extends to a functor $\widetilde{F}\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ that admits a right adjoint. Moreover, this adjunction is an enriched adjunction.*

Proof The simplicial functor $\widetilde{F}\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ is defined on objects by $\widetilde{F}(\{c_i\}) = \mathrm{colim}_i\, F(c_i)$ and on the internal homs by

$$\mathrm{Map}(\{c_i\},\{d_j\}) = \lim_i \operatorname*{colim}_j \mathrm{Map}(c_i,d_j) \to \lim_i \operatorname*{colim}_j \mathrm{Map}(F(c_i),F(d_j)) \\ \to \mathrm{Map}\big(\operatorname*{colim}_i F(c_i), \operatorname*{colim}_i F(d_j)\big).$$

We saw in the preliminaries that $\widetilde{F}$ preserves all colimits and has a right adjoint (as functor of unenriched categories). In particular, it is part of an enriched adjunction if and only if it preserves tensors. To see that this is the case, let $X = \mathrm{colim}_j\, X_j$ be a simplicial set written as a filtered colimit of finite simplicial sets. Then $\{c_i\}_i \otimes X \cong \{c_i \otimes X_j\}_{(i,j)}$, hence $F(\{c_i\}_i \otimes X) \cong \mathrm{colim}_{(i,j)}\, F(c_i) \otimes X_j \cong \widetilde{F}(\{c_i\}) \otimes X$, using the hypothesis that F preserves tensors with finite simplicial sets. $\square$

In this section we give some assumptions under which an adjunction of the type above is a Quillen adjunction, and give a further criterion for this adjunction to be a Quillen equivalence. This gives a straightforward way of constructing "profinite" versions of certain classical Quillen adjunctions, as illustrated in Example 7.7. At the end of this section, we show that if $\mathbf{C} \subseteq \mathcal{E}$ inherits the structure of a (co)fibration test category in the sense of Example 3.6, then the ind- or pro-completion functor (relative to $\mathbf{C}$) is a Quillen functor.

Definition 7.2 A *morphism of cofibration test categories* $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$ is a simplicial functor $\phi\colon \mathbf{C}_1 \to \mathbf{C}_2$ that preserves finite colimits, finite tensors and (trivial) cofibrations, and in particular maps the full subcategory $\mathbf{T}_1$ into $\mathbf{T}_2$. Dually, a *morphism of fibration test categories* $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$ is a simplicial functor $\phi\colon \mathbf{C}_1 \to \mathbf{C}_2$ that preserves finite limits, finite cotensors and (trivial) fibrations, and in particular maps the full subcategory $\mathbf{T}_1$ into $\mathbf{T}_2$.

Example 7.3 The nerve functor $N\colon \mathbf{FinGrpd} \to \mathbf{L}_{\mathrm{KQ}}$ is a morphism of fibration test categories. Similarly, taking the nerve of a category gives a morphism of fibration test categories $N\colon \mathbf{FinCat} \to \mathbf{L}_{\mathrm{J}}$.

Remark 7.4 If $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$ is a morphism of cofibration test categories, then its canonical filtered colimit-preserving extension $\phi_!\colon \mathrm{Ind}(\mathbf{C}_1) \to \mathrm{Ind}(\mathbf{C}_2)$ has a right adjoint $\phi^*\colon \mathrm{Ind}(\mathbf{C}_2) \to \mathrm{Ind}(\mathbf{C}_1)$ by Lemma 7.1. Since $\phi_!$ is an extension of $\phi\colon \mathbf{C}_1 \to \mathbf{C}_2$, it sends all objects in the image of $\mathbf{C}_1 \hookrightarrow \mathrm{Ind}(\mathbf{C}_1)$ to compact objects in $\mathrm{Ind}(\mathbf{C}_2)$, hence its right adjoint ϕ^* must preserve filtered colimits. Dually, if ϕ is a morphism of fibration test categories, then it canonically extends to a functor $\phi_*\colon \mathrm{Pro}(\mathbf{C}_1) \to \mathrm{Pro}(\mathbf{C}_2)$ that admits a left adjoint $\phi^*\colon \mathrm{Pro}(\mathbf{C}_2) \to \mathrm{Pro}(\mathbf{C}_1)$ which preserves cofiltered limits.

Proposition 7.5 *Let $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$ be a morphism of cofibration test categories. Then the induced adjunction from Remark 7.4*

$$\phi_!\colon \mathrm{Ind}(\mathbf{C}_1) \rightleftarrows \mathrm{Ind}(\mathbf{C}_2) :\phi^*$$

is a simplicial Quillen adjunction. Dually, for a morphism of fibration test categories $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$, the induced adjunction from Remark 7.4

$$\phi^*\colon \mathrm{Pro}(\mathbf{C}_2) \rightleftarrows \mathrm{Pro}(\mathbf{C}_1) :\phi_*$$

is a simplicial Quillen adjunction.

Proof Suppose $\phi\colon (\mathbf{C}_1,\mathbf{T}_1)\to(\mathbf{C}_2,\mathbf{T}_2)$ is a morphism of cofibration test categories. By Lemma 7.1, the adjunction $\phi_! \dashv \phi^*$ is an enriched adjunction of simplicial functors. Since $\phi_!$ extends ϕ and $\phi\colon \mathbf{C}_1\to\mathbf{C}_2$ preserves all (trivial) cofibrations, we conclude that $\phi_!\colon \mathrm{Ind}(\mathbf{C}_1)\to\mathrm{Ind}(\mathbf{C}_2)$ preserves all generating (trivial) cofibrations. We conclude that $\phi_! \dashv \phi^*$ is a simplicial Quillen adjunction. The case of fibration test categories is dual. □

Remark 7.6 One could weaken the definition of a morphism of (co)fibration test categories $\phi\colon \mathbf{C}_1\to\mathbf{C}_2$ by only asking it to be an (unenriched) functor of underlying categories and not asking it to preserve (co)tensors. In this case, one would still obtain a Quillen adjunction between the completed model structures, but it would merely be a Quillen adjunction between the underlying model categories, and not a simplicial one. Moreover, the proof of Proposition 7.8 below would not go through in this case.

Example 7.7 The nerve functors from Example 7.3 induce simplicial Quillen adjunctions $\widehat{\Pi}_1\colon \widehat{\mathbf{sSet}}_{\mathrm{Q}} \rightleftarrows \widehat{\mathbf{Grpd}}:\widehat{N}$ and $\widehat{h}\colon \widehat{\mathbf{sSet}}_{\mathrm{J}} \rightleftarrows \widehat{\mathbf{Cat}}:\widehat{N}$. These left adjoints are profinite versions of the fundamental groupoid and the homotopy category, respectively.

We call the restriction $\phi\colon \mathbf{T}_1\to\mathbf{T}_2$ of a morphism of cofibration test categories *homotopically essentially surjective* if for any $t'\in\mathbf{T}_2$, there exists a $t\in\mathbf{T}_1$ together with a weak equivalence $\phi(t)\xrightarrow{\sim} t'$ in $\mathbf{T}_2$.

Proposition 7.8 *Let $\phi\colon (\mathbf{C}_1,\mathbf{T}_1)\to(\mathbf{C}_2,\mathbf{T}_2)$ be a morphism of (co)fibration test categories.*

(a) *If the restriction $\mathbf{T}_1\to\mathbf{T}_2$ of ϕ is homotopically essentially surjective, then ϕ^* detects weak equivalences.*

(b) *In the case of a morphism of cofibration test categories, if moreover for any $t\in\mathbf{T}_1$ and $c\in\mathbf{C}_1$ the map*

$$\mathrm{Map}(t,c)\to\mathrm{Map}(\phi(t),\phi(c))$$

is a weak equivalence, then the induced Quillen adjunction of Proposition 7.5 is a Quillen equivalence.

(b′) *In the case of fibration test categories, if ϕ is homotopically essentially surjective and for any $t\in\mathbf{T}_1$ and $c\in\mathbf{C}_1$, the map*

$$\mathrm{Map}(c,t)\to\mathrm{Map}(\phi(c),\phi(t))$$

is a weak equivalence, then the induced Quillen adjunction of Proposition 7.5 is a Quillen equivalence.

Proof We again only include a proof for cofibration test categories, as the case of a morphism of fibration test categories is dual. For item (a), let $f: C \to D$ be a map in $\mathrm{Ind}(\mathbf{C}_2)$ and suppose that $\phi^*(f)$ is a weak equivalence in $\mathrm{Ind}(\mathbf{C}_1)$. If $t' \in \mathbf{T}_2$, then since $\phi: \mathbf{T}_1 \to \mathbf{T}_2$ is homotopically essentially surjective, there is a $t \in \mathbf{T}_1$ together with an equivalence $\phi(t) \xrightarrow{\sim} t'$. Since C and D are fibrant in $\mathrm{Ind}(\mathbf{C}_2)$, the map $\mathrm{Map}(t', C) \to \mathrm{Map}(t', D)$ is a weak equivalence if and only if $\mathrm{Map}(\phi(t), C) \to \mathrm{Map}(\phi(t), D)$ is so. Since $\phi_!$ extends ϕ and the adjunction $\phi_! \dashv \phi^*$ is enriched, we see that

$$\begin{tikzcd}
\mathrm{Map}(\phi(t), C) \arrow[r] \arrow[d, "\cong"] & \mathrm{Map}(\phi(t), D) \arrow[d, "\cong"] \\
\mathrm{Map}(t, \phi^*(C)) \arrow[r, "\sim"] & \mathrm{Map}(t, \phi^*(D))
\end{tikzcd}$$

commutes, hence $\mathrm{Map}(t', C) \to \mathrm{Map}(t', D)$ is a weak equivalence.

For item (b), since the right adjoint ϕ^* detects weak equivalences by part (a), it suffices to show that the unit $C \to \phi^*\phi_! C$ is a weak equivalence for every cofibrant C in $\mathrm{Ind}(\mathbf{C}_1)$. Since C is a cofiltered limit of objects in $\mathbf{C}_1$, by Remark 7.4 it is enough to show that $c \to \phi^*\phi_! c$ is a weak equivalence for every c in $\mathbf{C}_1$. By definition of the weak equivalences in $\mathrm{Ind}(\mathbf{C}_1)$ and by the simplicial adjunction $\phi_! \dashv \phi^*$, this is equivalent to

$$\mathrm{Map}(t, c) \to \mathrm{Map}(\phi_!(t), \phi_!(c)) \cong \mathrm{Map}(\phi(t), \phi(c))$$

being a weak equivalence, which holds by assumption. □

An interesting consequence of Proposition 7.8 is that if, for a (co)fibration test category $(\mathbf{C}, \mathbf{T})$, one "enlarges" $\mathbf{C}$ to a bigger category $\mathbf{C}'$ but keeps $\mathbf{T}$ the same, then one obtains Quillen equivalent model structures on $\mathrm{Ind}(\mathbf{C})$ and $\mathrm{Ind}(\mathbf{C}')$ (or $\mathrm{Pro}(\mathbf{C})$ and $\mathrm{Pro}(\mathbf{C}')$). The next example gives an illustration of this.

Example 7.9 Recall from Example 5.5 that the category of lean simplicial sets $\mathbf{L}$ inherits the structure of a fibration test category from $\mathbf{sSet}_{\mathrm{KQ}}$. We could give the category of degreewise finite simplicial sets $\mathbf{sFinSet}$ a similar structure of a fibration test category, namely by defining the test objects to be the lean Kan complexes and the (trivial) fibrations to be those of $\mathbf{L}_{\mathrm{KQ}}$. That is, the test objects and the (trivial) fibrations of $\mathbf{sFinSet}$ and of $\mathbf{L}_{\mathrm{KQ}}$ are identical. It is well known that the pro-categories $\mathrm{Pro}(\mathbf{sFinSet})$ and $\mathrm{Pro}(\mathbf{L}) \simeq \widehat{\mathbf{sSet}}$ are not equivalent. However, the inclusion $\iota: \mathbf{L}_{\mathrm{KQ}} \hookrightarrow \mathbf{sFinSet}$ is a morphism of fibration test categories that satisfies item (b′) of Proposition 7.8, hence the induced adjunction

$$\iota^* : \mathrm{Pro}(\mathbf{sFinSet}) \rightleftarrows \widehat{\mathbf{sSet}}_{\mathrm{Q}} : \iota_*$$

is a Quillen equivalence.

The hypotheses for item (b) of Proposition 7.8 can usually be weakened, namely if $\mathbf{T}$ is "large enough" in the following sense.

Definition 7.10 Let $(\mathbf{C}, \mathbf{T})$ be a cofibration test category. We say that $\mathbf{T}$ is *closed under pushouts along cofibrations* if, for any cofibration $r \rightarrowtail s$ in $\mathbf{T}$ and any map $r \to t$ in $\mathbf{T}$, the pushout $s \cup_r t$ is again contained in $\mathbf{T}$.

Dually, for a fibration test category $(\mathbf{C}, \mathbf{T})$, we say that $\mathbf{T}$ is *closed under pullbacks along fibrations* if, for any fibration $s \twoheadrightarrow r$ and any map $t \to r$ in $\mathbf{T}$, the pullback $s \times_r t$ is again contained in $\mathbf{T}$.

This definition can be seen as ensuring that $\mathbf{T}$ has all finite homotopy (co)limits. If $\mathbf{T}$ is closed under pushouts along cofibrations, then it is enough to assume in item (b) that the restriction $\phi\colon \mathbf{T}_1 \to \mathbf{T}_2$ is *homotopically fully faithful*, ie that $\mathrm{Map}(s,t) \to \mathrm{Map}(\phi(s), \phi(t))$ is a weak equivalence for all $s, t \in \mathbf{T}_1$. The main ingredient is the following useful lemma.

Lemma 7.11 *Let $(\mathbf{C}, \mathbf{T})$ be a cofibration test category and suppose that $\mathbf{T}$ is closed under pushouts along cofibrations. Then any cofibrant object in $\mathrm{Ind}(\mathbf{C})$ is a filtered colimit of objects in $\mathbf{T}$.*

Proof The "fat small object argument" of [27] shows that if C in $\mathrm{Ind}(\mathbf{C})$ is cofibrant, then it is a retract of a colimit $\mathrm{colim}_{i \in I} c_i$ indexed by a directed poset I that has a least element $\perp$, such that $c_\perp$ is the initial object $\varnothing$ and such that $c_\perp \to c_i$ is a (finite) composition of pushouts of generating cofibrations for any i. (This follows from Theorem 4.11 of [27] together with the fact that all objects in $\mathbf{T}$ are compact.) In particular, since $\mathbf{T}$ is closed under pushouts along cofibrations, it follows that $c_i \in \mathbf{T}$ for every $i \in I$. Since ind-categories are idempotent complete, it follows that any retract of such a colimit is an object of $\mathrm{Ind}(\mathbf{T})$ as well. In particular, any cofibrant object of $\mathrm{Ind}(\mathbf{C})$ lies in $\mathrm{Ind}(\mathbf{T})$. □

We leave it to the reader to dualize Lemma 7.11 to the context of fibration test categories.

Proposition 7.12 *Let $\phi\colon (\mathbf{C}_1, \mathbf{T}_1) \to (\mathbf{C}_2, \mathbf{T}_2)$ be a morphism of cofibration test categories (resp. fibration test categories) and suppose that $\mathbf{T}_1$ is closed under pushouts along cofibrations (resp. closed under pullbacks along fibrations). If the restriction $\phi\colon \mathbf{T}_1 \to \mathbf{T}_2$ is homotopically essentially surjective and homotopically fully faithful, then the induced Quillen adjunction of Proposition 7.5 is a Quillen equivalence.*

Proof We prove the statement for ind-categories. As in the proof of Proposition 7.8, it suffices to show that the unit $C \to \phi^*\phi_! C$ is a weak equivalence for every cofibrant C in $\mathrm{Ind}(\mathbf{C}_1)$. By Lemma 7.11 any cofibrant object is a filtered colimit of objects of $\mathbf{T}_1$, so by Remark 7.4 it suffices to show that $t \to \phi^*\phi_! t$ is a weak equivalence for every $t \in \mathbf{T}_1$. This follows exactly as in the proof of Proposition 7.8. □

Recall from Section 2.2 that if $\mathcal{E}$ is a complete category and if $\mathbf{C} \subseteq \mathcal{E}$ is a small full subcategory closed under finite limits, then the functor $U\colon \mathrm{Pro}(\mathbf{C}) \to \mathcal{E}$ that sends a pro-object to its limit in $\mathcal{E}$ has a left adjoint $\widehat{(\cdot)}_{\mathrm{Pro}}$, the pro-$\mathbf{C}$ completion functor. Dually, if $\mathcal{E}$ is cocomplete and $\mathbf{C}$ is closed under finite colimits, then the canonical functor $U\colon \mathrm{Ind}(\mathbf{C}) \to \mathcal{E}$ has a right adjoint $\widehat{(\cdot)}_{\mathrm{Ind}}$. In the situation where $\mathcal{E}$ is a simplicial model category and $\mathbf{C}$ is a (co)fibration test category, these adjunctions are almost by definition Quillen pairs. Note that in the case of pro-categories, this is the Quillen pair mentioned in item (iii) of Theorem 1.1.

Proposition 7.13 *Let $\mathcal{E}$ be a simplicial model category in which every object is fibrant and $\mathbf{C} \subseteq \mathcal{E}$ a full subcategory closed under finite colimits and finite tensors with the inherited structure of a cofibration test category (in the sense of Example 3.6). Then*

$$U\colon \mathrm{Ind}(\mathbf{C}) \rightleftarrows \mathcal{E} : \widehat{(\cdot)}_{\mathrm{Ind}}$$

is a simplicial Quillen adjunction. Dually, if every object in $\mathcal{E}$ is cofibrant and $\mathbf{C} \subseteq \mathcal{E}$ is a full subcategory closed under finite limits and finite cotensors, given the inherited structure of a fibration test category (as in Example 5.3), then

$$\widehat{(\cdot)}_{\mathrm{Pro}}\colon \mathcal{E} \rightleftarrows \mathrm{Pro}(\mathbf{C}) : U$$

is a simplicial Quillen adjunction.

Proof The first adjunction arises by applying Lemma 7.1 to the inclusion $\mathbf{C} \hookrightarrow \mathcal{E}$. We need to show that the left adjoint U preserves the generating (trivial) cofibrations. Note that U agrees with the inclusion $\mathbf{C} \hookrightarrow \mathcal{E}$ when restricted to $\mathbf{C} \subseteq \mathrm{Ind}(\mathbf{C})$. Since the generating (trivial) cofibrations are defined as the (trivial) cofibrations in $\mathbf{C} \subseteq \mathcal{E}$ between cofibrant objects, they are preserved by U.

The case for pro-$\mathbf{C}$ completion follows dually. $\square$

Example 7.14 The proposition above shows that the profinite completion functors for $\mathbf{sSet}_{\mathrm{KQ}}$ and $\mathbf{Grpd}$ are left Quillen. These Quillen adjunctions fit into a commutative diagram

$$\begin{tikzcd}
\mathbf{sSet}_{\mathrm{KQ}} \arrow[r, bend left, "\Pi_1"] \arrow[r, phantom, "\perp"] \arrow[d, bend right, "\widehat{(\cdot)}_{\mathrm{Pro}}"'] \arrow[d, phantom, "\dashv"] & \mathbf{Grpd} \arrow[l, bend left, "N"] \arrow[d, bend right, "\widehat{(\cdot)}_{\mathrm{Pro}}"'] \arrow[d, phantom, "\dashv"] \\
\widehat{\mathbf{sSet}}_{\mathrm{Q}} \arrow[u, bend right, "U"'] \arrow[r, bend left, "\widehat{\Pi}_1"] \arrow[r, phantom, "\perp"] & \widehat{\mathbf{Grpd}} \arrow[l, bend left, "\widehat{N}"] \arrow[u, bend right, "U"']
\end{tikzcd}$$

where $\widehat{N}$ is the nerve adjunction from Example 7.7. There is a similar diagram of Quillen adjunctions for the (profinite) Joyal model structure and the model category of (profinite) categories.

8 Bousfield localizations

Suppose we are given a cofibration test category $(\mathbf{C}, \mathbf{T})$ and that we wish to shrink the full subcategory of test objects $\mathbf{T}$ to a smaller one $\mathbf{T}' \subseteq \mathbf{T}$. If $\mathbf{T}'$ is closed under finite pushout-products, then $(\mathbf{C}, \mathbf{T}')$ is a cofibration test category by Example 3.7, hence we obtain two model structures $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$ and $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$ on the category $\mathrm{Ind}(\mathbf{C})$. Since the (trivial) cofibrations of $(\mathbf{C}, \mathbf{T}')$ are those of $(\mathbf{C}, \mathbf{T})$ between objects of $\mathbf{T}'$, the sets of generating (trivial) cofibrations of $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$ are contained in those of $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$. In particular, the identity functor is right Quillen when viewed as a functor $\mathrm{Ind}(\mathbf{C}, \mathbf{T}) \to \mathrm{Ind}(\mathbf{C}, \mathbf{T}')$. Since there are fewer weak equivalences in $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$ than in $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$, this right Quillen functor is close to being a right Bousfield localization. Recall that a right Bousfield localization of a model category is a model structure on the same category with the same class of fibrations, but with a larger class of weak equivalences. The model category $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$ is not necessarily a right Bousfield localization of $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$ since it has fewer generating trivial cofibrations, and hence it might have more fibrations than $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$. However, it is a general fact about model categories that in such a situation, there exists a model structure on $\mathrm{Ind}(\mathbf{C})$ with the weak equivalences of $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$ and the fibrations of $\mathrm{Ind}(\mathbf{C}, \mathbf{T})$:

Lemma 8.1 *Let $\mathcal{E}_\alpha$ and $\mathcal{E}_\beta$ be cofibrantly generated model structures on the same category $\mathcal{E}$ and suppose that sets of generating cofibrations I_α and I_β and sets of generating trivial cofibration J_α and J_β respectively, are given. If $I_\alpha \subseteq I_\beta$ and $J_\alpha \subseteq J_\beta$, and if $\mathcal{E}_\alpha$ has more weak equivalences than $\mathcal{E}_\beta$, then there exists a cofibrantly generated model structure on $\mathcal{E}$ with the weak equivalences of $\mathcal{E}_\alpha$ and the fibrations of $\mathcal{E}_\beta$.*

Proof It easily follows by checking the hypotheses of Theorem 11.3.1 of [17] that the sets $I_\alpha \cup J_\beta$ and J_β determine a cofibrantly generated model structure on $\mathcal{E}$ in which the weak equivalences agree with those of $\mathcal{E}_\alpha$. This model structure has the desired properties. As an example, we check item (4b) of Theorem 11.3.1 of [17], and leave the other hypotheses to the reader. This comes down to showing that if $E \to F$ has the right lifting property with respect to J_β and is a weak equivalence in $\mathcal{E}_\alpha$, then it must have the right lifting property with respect to $I_\alpha \cup J_\beta$. It suffices to show that $E \to F$ has the right lifting property with respect to I_α. Since $E \to F$ has the right lifting property with respect to J_β, it has so with respect to $J_\alpha \subseteq J_\beta$, hence it is a fibration in $\mathcal{E}_\alpha$. Since it is also a weak equivalence in $\mathcal{E}_\alpha$, it follows that it has the right lifting property with respect to I_α and hence with respect to $I_\alpha \cup J_\beta$. $\square$

If $\mathcal{E}$ is a simplicial model category with a given full subcategory $\mathbf{T} \subseteq \mathcal{E}$, then $R_{\mathbf{T}}\mathcal{E}$ denotes (if it exists) the right Bousfield localization of $\mathcal{E}$ in which a map $E \to E'$ is a weak equivalence if and only if $\mathrm{Map}(t, E) \to \mathrm{Map}(t, E')$ is a weak equivalence for every $t \in \mathbf{T}$. We call such a map a $\mathbf{T}$–colocal weak equivalence. Dually, $L_{\mathbf{T}}\mathcal{E}$ denotes (if it exists) the left Bousfield localization of $\mathcal{E}$ in which $E \to E'$ is a weak equivalence if and only if $\mathrm{Map}(E', t) \to \mathrm{Map}(E, t)$ is a weak equivalence for every $t \in \mathbf{T}$. Such a map is called a $\mathbf{T}$–local weak equivalence.

Proposition 8.2 *Let $(\mathbf{C}, \mathbf{T})$ be a cofibration test category and let $\mathbf{T}' \subseteq \mathbf{T}$ be a full subcategory. Then the right Bousfield localization $R_{\mathbf{T}'} \mathrm{Ind}(\mathbf{C})$ exists and is cofibrantly generated.*

Dually, if $(\mathbf{C}, \mathbf{T})$ is a fibration test category and $\mathbf{T}' \subseteq \mathbf{T}$ a full subcategory, then the left Bousfield localization $L_{\mathbf{T}'} \mathrm{Pro}(\mathbf{C})$ exists and is fibrantly generated.

Proof We first prove the proposition in the special case that $\mathbf{T}'$ is closed under finite pushout-products, and then deduce the general case from this. In this special case,

$(\mathbf{C}, \mathbf{T}')$ is a cofibration test category as in Example 3.7, so we obtain a cofibrantly generated model category on $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$ in which the weak equivalences are the $\mathbf{T}'$–colocal ones. We also have the model structure on $\mathrm{Ind}(\mathbf{C})$ corresponding to the cofibration test category $(\mathbf{C}, \mathbf{T})$, which by construction has more generating (trivial) cofibrations than $\mathrm{Ind}(\mathbf{C}, \mathbf{T}')$. By applying Lemma 8.1, we obtain the desired right Bousfield localization $R_{\mathbf{T}'} \mathrm{Ind}(\mathbf{C})$.

Now suppose that $\mathbf{T}'$ is not necessarily closed under finite pushout-products. Let $\mathbf{T}''$ be the smallest full subcategory of $\mathbf{T}$ that contains $\mathbf{T}'$ and is closed under finite pushout-products and isomorphisms. This category can be obtained by repeatedly enlarging $\mathbf{T}'$ by adding all objects isomorphic to an object of the form $t' \otimes U \cup_{s' \otimes U} s' \otimes V$ to $\mathbf{T}'$, for $s' \rightarrowtail t'$ a cofibration in $\mathbf{T}'$ and $U \rightarrowtail V$ a cofibration of finite simplicial sets. This produces a sequence of full subcategories $\mathbf{T}' \subseteq \mathbf{T}'_1 \subseteq \mathbf{T}'_2 \subseteq \cdots \subseteq \mathbf{T}$ such that $\mathbf{T}'' = \bigcup_{n \in \mathbb{N}} \mathbf{T}'_n$. We claim that the $\mathbf{T}'$–colocal weak equivalences and the $\mathbf{T}''$–colocal weak equivalences in $\mathrm{Ind}(\mathbf{C})$ agree. By the above inductive construction of $\mathbf{T}''$, it suffices to show that for any cofibration $s' \rightarrowtail t'$ of $(\mathbf{C}, \mathbf{T})$ with $s', t' \in \mathbf{T}'$ and any cofibration $U \rightarrowtail V$ in $\mathbf{sSet}_{\mathrm{fin}}$, the map

$$\mathrm{Map}(t' \otimes U \cup_{s' \otimes U} s' \otimes V, C) \to \mathrm{Map}(t' \otimes U \cup_{s' \otimes U} s' \otimes V, D)$$

is a weak equivalence for any $\mathbf{T}'$–colocal weak equivalence $C \to D$. We leave this as an exercise to the reader, noting that these pushouts can be taken out of the mapping spaces to obtain homotopy pullbacks. □

Example 8.3 Let $\mathbf{L}_{\mathrm{KQ}}$ be the category of lean simplicial sets with the structure of a fibration test category as in Example 5.5. The model structure $\mathrm{Pro}(\mathbf{L}_{\mathrm{KQ}})$ then coincides with Quick's model structure $\widehat{\mathbf{sSet}}_Q$ under the equivalence of categories $\mathrm{Pro}(\mathbf{L}) \simeq \widehat{\mathbf{sSet}}_{\mathrm{Q}}$, by Corollary 6.6. In particular, by Proposition 8.2, the left Bousfield localization $L_{\mathbf{T}} \widehat{\mathbf{sSet}}_{\mathrm{Q}}$ exists for any collection of lean Kan complexes $\mathbf{T}$. If one takes $\mathbf{T}$ to consist of the spaces $K(\mathbb{F}_p, n)$ for all $n \in \mathbb{N}$, then one obtains a model structure on $\widehat{\mathbf{sSet}}$ in which the weak equivalences are the maps that induce equivalences in $\mathbb{F}_p$–cohomology and in which the cofibrations are the monomorphisms. This is exactly Morel's model structure on $\widehat{\mathbf{sSet}}$ for pro-p spaces [30]. In particular, this is an alternative to the construction in Example 5.6.

Example 8.4 Recall the Reedy model structure (with respect to Quick's model structure on $\widehat{\mathbf{sSet}}$) on $\mathbf{bis}\widehat{\mathbf{Set}}$ from Example 5.11. By Proposition 6.9, this model structure can be obtained by applying Theorem 5.2 to a certain fibration test category $\mathbf{L}_{\mathrm{R}}^{(2)}$. In particular, Proposition 8.2 ensures that the left Bousfield localization $L_{\mathbf{T}}\mathbf{bis}\widehat{\mathbf{Set}}$ exists for any collection $\mathbf{T}$ of Reedy fibrant doubly lean simplicial sets. For example, one can take $\mathbf{T}$ to be the collection of all doubly lean bisimplicial sets that are complete Segal spaces in the sense of [35]. This model structure will be called the *model structure for complete Segal profinite spaces* and denoted by $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$. We will study this model structure in detail in Section 9. In particular, we will show in Proposition 9.3 that $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$ is equivalent to the model structure for profinite quasicategories $\mathbf{s}\widehat{\mathbf{Set}}_{\mathrm{J}}$ from Example 5.7.

Proposition 8.2 was the last missing piece in the proof of Theorem 1.1 (except for item (iv) of that theorem, which follows from Theorem A.7).

Proof of Theorem 1.1 Let $\mathcal{E}$ be a simplicial model category in which every object is cofibrant, and let $\mathbf{C} \subseteq \mathcal{E}$ be a small full subcategory of $\mathcal{E}$ which is closed under finite limits and cotensors by finite simplicial sets. Then $(\mathbf{C}, \mathbf{T}')$, where $\mathbf{T}' \subseteq \mathbf{C}$ is the full subcategory on the fibrant objects, inherits the structure of a fibration test category from $\mathcal{E}$ in the sense of Example 5.3.

Now suppose $\mathbf{T}$ is any collection of fibrant objects in $\mathbf{C}$. By applying Theorem 5.2 to $(\mathbf{C}, \mathbf{T}')$ and then applying Proposition 8.2 (with $\mathbf{T}$ and $\mathbf{T}'$ interchanged), we obtain a model structure on $\mathrm{Pro}(\mathbf{C})$ together with a (fibrantly generated) left Bousfield localization $\mathrm{Pro}(\mathbf{C}) \rightleftarrows L_{\mathbf{T}}\,\mathrm{Pro}(\mathbf{C})$. The weak equivalences of $L_{\mathbf{T}}\,\mathrm{Pro}(\mathbf{C})$ are by definition the $\mathbf{T}$–local equivalences. By Theorem 5.2, any object in $\mathrm{Pro}(\mathbf{C})$ (and hence in $L_{\mathbf{T}}\,\mathrm{Pro}(\mathbf{C})$) is cofibrant. By Proposition 7.13, we obtain a simplicial Quillen adjunction $\mathcal{E} \rightleftarrows \mathrm{Pro}(\mathbf{C})$ and hence a simplicial Quillen adjunction $\mathcal{E} \rightleftarrows L_{\mathbf{T}}\,\mathrm{Pro}(\mathbf{C})$. We conclude that the model structure $L_{\mathbf{T}}\,\mathrm{Pro}(\mathbf{C})$ satisfies items (i)–(iii) of Theorem 1.1. □

9 Example: complete Segal profinite spaces vs profinite quasicategories

Recall that in Example 5.7, we defined the profinite Joyal model structure. In this section, we will define another candidate for the homotopy theory of profinite ∞–categories, namely a profinite version of Rezk's model category of complete Segal spaces. We then show that there are two Quillen equivalences between the model category of complete Segal profinite spaces and the profinite Joyal model structure. After establishing these Quillen equivalences, we characterize in both these model categories the weak equivalences between the fibrant objects as the essentially surjective and fully faithful maps, where being fully faithful is defined in terms of the Quick model structure. It is worth mentioning that in Remark A.11, we moreover give a precise description of the underlying ∞–category of these model categories.

Let us start with a short review of the theory of complete Segal spaces, originally defined by Rezk in [35]. Consider the category $\mathbf{bisSet} = \mathbf{sSet}^{\Delta^{\mathrm{op}}}$ of *bisimplicial sets*, or *simplicial spaces*, equipped with the Reedy model structure (with respect to the Kan–Quillen model structure on $\mathbf{sSet}$). We denote this model category by $\mathbf{bisSet}_{\mathrm{R}}$. Objects of $\mathbf{bisSet}$ have two simplicial parameters. We denote the "inner" one by $n, m, \dots$ and refer to it as the *space parameter*, and we denote the "outer" one (corresponding to the Δ^{op} in $\mathbf{sSet}^{\Delta^{\mathrm{op}}}$) by $s, t, r, \dots$. For any pair of simplicial sets X and Y, one can define the *external product* $X \underline{\times} Y$ by $(X \underline{\times} Y)_{t,n} = X_t \times Y_n$. Note that the external product $\Delta^t \underline{\times} \Delta^n$ is the functor $\Delta^{\mathrm{op}} \times \Delta^{\mathrm{op}} \to \mathbf{Set}$ represented by $([t],[n])$. In particular, the internal hom of $\mathbf{bisSet}$ can be defined by $(Y^X)_{t,n} = \mathrm{Hom}((\Delta^t \underline{\times} \Delta^n) \times X, Y)$. This internal hom allows one to regard $\mathbf{bisSet}$ as a simplicial category in multiple ways; the two simplicial enrichments that we will use are given by

$$\mathrm{Map}_1(X,Y) := (Y^X)_{\bullet,0} \quad \text{and} \quad \mathrm{Map}_2(X,Y) := (Y^X)_{0,\bullet}.$$

The category $\mathbf{bisSet}$ is tensored and cotensored with respect to both of these enrichments.

As described in [35, Sections 10 and 12], one can localize the Reedy model structure on $\mathbf{bisSet}$ by the Segal maps

$$\mathrm{Sp}\,\Delta^t \underline{\times} \Delta^0 \rightarrowtail \Delta^t \underline{\times} \Delta^0,$$

where $\mathrm{Sp}\,\Delta^t = \Delta[0,1] \cup \dots \cup \Delta[t-1,t]$ is the spine of the t–simplex. This gives the model category $\mathbf{bisSet}_{\mathrm{SS}}$ for *Segal spaces*. Localizing one step further by the map

$$\{0\} \underline{\times} \Delta^0 \rightarrowtail J \underline{\times} \Delta^0$$

gives the model category $\mathbf{bisSet}_{\mathrm{CSS}}$ for *complete Segal spaces*. Here J is the nerve of the groupoid with two objects and exactly one isomorphism between any ordered pair of objects. It is part of a cosimplicial object $J^\bullet$ in $\mathbf{sSet}$, J^t being the nerve of the groupoid with $t+1$ objects and exactly one isomorphism between any ordered pair of objects.

All three of the model structures $\mathbf{bisSet}_{\mathrm{R}}$, $\mathbf{bisSet}_{\mathrm{SS}}$ and $\mathbf{bisSet}_{\mathrm{CSS}}$ are $\mathbf{sSet}_{\mathrm{KQ}}$–enriched model structures with respect to the enrichment Map_2 mentioned above.

The model category $\mathbf{bisSet}_{\mathrm{CSS}}$ is Quillen equivalent to $\mathbf{sSet}_{\mathrm{J}}$. In fact, there are Quillen pairs in both directions, whose right Quillen functors are the evaluation at the inner coordinate $n=0$,

$$\mathrm{ev}_0\colon \mathbf{bisSet} \to \mathbf{sSet}; \quad (\mathrm{ev}_0 X)_t = X_{t,0},$$

and the singular complex functor with respect to $J^\bullet$,

$$\mathrm{Sing}_J\colon \mathbf{sSet} \to \mathbf{bisSet}; \quad \mathrm{Sing}_J(X)_{t,n} = \mathrm{Map}(J^n, X)_t = \mathrm{Hom}(\Delta^t \times J^n, X).$$

These Quillen equivalences are described in detail in [22]. One can prove, using the Quillen equivalence ev_0 together with the fact that $\mathbf{bisSet}_{\mathrm{CSS}}$ is a cartesian closed model category, that $\mathbf{bisSet}_{\mathrm{CSS}}$ is an $\mathbf{sSet}_{\mathrm{J}}$–enriched model category with respect to the simplicial enrichment Map_1 mentioned above. Both of the above right Quillen functors are simplicial functors that preserve cotensors with respect to this simplicial enrichment. This is explained in detail in the proof of Proposition E.2.2 of [37].

Now let $\mathbf{L}^{(2)}$ be the category of *doubly lean* bisimplicial sets, ie those bisimplicial sets X for which $X_{t,n}$ is finite for each t and n, and such that $X \cong \mathrm{cosk}_{t,n}(X)$ for some t and n. Here $\mathrm{cosk}_{t,n}\colon \mathbf{bisSet} \to \mathbf{bisSet}$ is the functor that restricts $X \in \mathbf{bisSet}$ to a functor $\Delta^{\mathrm{op}}_{\le t} \times \Delta^{\mathrm{op}}_{\le n} \to \mathbf{Set}$ and then right Kan extends along $\Delta^{\mathrm{op}}_{\le t} \times \Delta^{\mathrm{op}}_{\le n} \hookrightarrow \Delta^{\mathrm{op}} \times \Delta^{\mathrm{op}}$. This agrees with the notion of doubly lean as defined at the end of Section 2.2, and it follows from (the dual of) Theorem 2.3 that the inclusion $\mathbf{L}^{(2)} \hookrightarrow \widehat{\mathbf{bisSet}}$ extends to an equivalence $\mathrm{Pro}(\mathbf{L}^{(2)}) \simeq \widehat{\mathbf{bisSet}}$.

Each of the three model structures $\mathbf{bisSet}_{\mathrm{R}}$, $\mathbf{bisSet}_{\mathrm{SS}}$ and $\mathbf{bisSet}_{\mathrm{CSS}}$ gives rise to the structure of a fibration test category on $\mathbf{L}^{(2)}$ by the general scheme of Example 5.3. We will mainly be interested in the Reedy and the complete Segal model structures, so denote the corresponding fibration test categories by $\mathbf{L}^{(2)}_{\mathrm{R}}$ and $\mathbf{L}^{(2)}_{\mathrm{CSS}}$, respectively.

Definition 9.1 The model structures on $\widehat{\mathbf{bisSet}}$ obtained by applying Theorem 5.2 to the fibration test categories $\mathbf{L}^{(2)}_{\mathrm{R}}$ and $\mathbf{L}^{(2)}_{\mathrm{CSS}}$ will be called the *Reedy model structure for profinite spaces* and *model structure for complete Segal profinite spaces*, and denoted by $\widehat{\mathbf{bisSet}}_{\mathrm{R}}$ and $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$, respectively. A fibrant object in $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ will be called a *complete Segal profinite space*.

Since we can view $\mathbf{bisSet}_{\mathrm{CSS}}$ as a simplicial model category in two ways, the full subcategory $\mathbf{L}^{(2)}_{\mathrm{CSS}}$ inherits two different structures of a fibration test category, namely one with respect to the enrichment Map_1 and one with respect to Map_2. The (trivial) fibrations of both fibration test category structures agree, so they will induce the same model structures on $\mathrm{Pro}(\mathbf{L}^{(2)}) \cong \widehat{\mathbf{bisSet}}$. This shows that we can view $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ as an $\mathbf{sSet}_{\mathrm{J}}$–enriched model category through the enrichment Map_1, and as an $\mathbf{sSet}_{\mathrm{KQ}}$–enriched model category through Map_2.[1] In what follows, we will consider the simplicial enrichment Map_1, since this one is compatible with the right Quillen functors ev_0 and Sing_J discussed above.

By Proposition 7.13, the profinite completion functor $\mathbf{bisSet} \to \widehat{\mathbf{bisSet}}$ is a left Quillen functor, whose right adjoint is given by the functor $U\colon \widehat{\mathbf{bisSet}} \to \mathbf{bisSet}$ that sends a bisimplicial profinite set to its underlying bisimplicial set. Levelwise, this is the functor that sends a profinite set to its underlying set.

Since $\mathbf{L}^{(2)}_{\mathrm{CSS}}$ has fewer test objects than $\mathbf{L}^{(2)}_{\mathrm{R}}$, we see that $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ has more weak equivalences than $\widehat{\mathbf{bisSet}}_{\mathrm{R}}$. By Proposition 6.8, the cofibrations are the monomorphisms in both model structures, hence $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ is a left Bousfield localization of $\widehat{\mathbf{bisSet}}_{\mathrm{R}}$. In particular, the construction of the model structure $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ given in Example 8.4 agrees with the one given here.

The right Quillen functors ev_0 and Sing_J mentioned above restrict to morphisms of fibration test categories between $\mathbf{L}_{\mathrm{J}}$ and $\mathbf{L}^{(2)}_{\mathrm{CSS}}$, where $\mathbf{L}_{\mathrm{J}}$ is the category of lean simplicial sets (with the fibration test category structure from Example 5.7). This amounts to showing that ev_0 maps doubly lean bisimplicial sets to lean simplicial sets, and that Sing_J maps lean simplicial sets to doubly lean bisimplicial sets. In the case of ev_0, this follows directly from the definition, while the case of Sing_J requires some work.

Lemma 9.2 *The functor* $\mathrm{Sing}_J\colon \mathbf{sSet} \to \mathbf{bisSet}$ *takes lean simplicial sets to doubly lean bisimplicial sets.*

Proof Let X be a lean simplicial set and suppose that X is n–coskeletal. It suffices to show that $\mathrm{Sing}_J(X)_{\bullet,m}$ and $\mathrm{Sing}_J(X)_{t,\bullet}$ are both n–coskeletal and degreewise finite simplicial sets for any $t, m \in \mathbb{N}$. Since J^m is a degreewise finite simplicial set for every m, we see that $\mathrm{Sing}_J(X)_{\bullet,m} = \mathrm{Map}(J^m, X)$ is an n–coskeletal degreewise finite simplicial set for every m. This automatically shows that $\mathrm{Sing}_J(X)_{t,\bullet}$ is a degreewise finite simplicial set as well. It therefore remains to show that, for every n–coskeletal simplicial set X and every t, the simplicial set $\mathrm{Sing}_J(X)_{t,\bullet} \cong \mathrm{Hom}(J^\bullet \times \Delta^t, X) \cong \mathrm{Hom}(J^\bullet, X^{\Delta^t})$ is n–coskeletal. Since any cotensor X^Y of an n–coskeletal simplicial set X is again n–coskeletal, it suffices to prove the case $t = 0$. To this end, let ∂J^{k+1} denote the simplicial subset

$$\partial J^{k+1} = \bigcup_{x \in (\Delta^{k+1})_k} J^k \subseteq J^{k+1},$$

[1] In fact, one can show that $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ is a $\mathbf{bisSet}_{\mathrm{CSS}}$–enriched model category, strengthening this statement.

or equivalently, the left Kan extension of $J^\bullet\colon \Delta \to \mathbf{sSet}$ along the Yoneda embedding $\Delta \to \mathbf{sSet}$, evaluated at $\partial\Delta^{k+1} \in \mathbf{sSet}$. The inclusion $\partial J^{k+1} \hookrightarrow J^{k+1}$ restricts to an isomorphism $\mathrm{sk}_n\, \partial J^{k+1} \to \mathrm{sk}_n\, J^{k+1}$ for any $k \geq n$. Combining this with the canonical isomorphism $\mathrm{Hom}(\partial\Delta^{k+1}, \mathrm{Hom}(J^\bullet, X)) \cong \mathrm{Hom}(\partial J^{k+1}, X)$, it follows that $\mathrm{Hom}(J^\bullet, X)$ is n–coskeletal. $\square$

Denote the profinite Joyal model structure by $\widehat{\mathbf{sSet}}_{\mathrm{J}}$. We can apply Proposition 7.12 to $\mathrm{ev}_0\colon \mathbf{L}^{(2)}_{\mathrm{CSS}} \to \mathbf{L}_{\mathrm{J}}$ and $\mathrm{Sing}_J\colon \mathbf{L}_{\mathrm{J}} \to \mathbf{L}^{(2)}_{\mathrm{CSS}}$ to show that the induced functors between $\widehat{\mathbf{sSet}}_{\mathrm{J}}$ and $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$ are right Quillen equivalences. We will denote these functors by ev_0 and Sing_J as well.

Proposition 9.3 *The functors* $\mathrm{ev}_0\colon \mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}} \to \widehat{\mathbf{sSet}}_{\mathrm{J}}$ *and* $\mathrm{Sing}_J\colon \widehat{\mathbf{sSet}}_{\mathrm{J}} \to \mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$ *are right Quillen equivalences.*

Proof Since there is a natural isomorphism $\mathrm{ev}_0\,\mathrm{Sing}_J(X) \cong X$, it suffices to show that $\mathrm{ev}_0\colon \mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}} \to \widehat{\mathbf{sSet}}_{\mathrm{J}}$ is a right Quillen equivalence. The same then follows for Sing_J by the two-out-of-three property. Since $\mathrm{ev}_0\colon \mathbf{bisSet}_{\mathrm{CSS}} \to \mathbf{sSet}_{\mathrm{J}}$ is a (simplicial) right Quillen equivalence, its restriction $\mathrm{ev}_0\colon \mathbf{L}^{(2)}_{\mathrm{CSS}} \to \mathbf{L}_{\mathrm{J}}$ is a morphism of fibration test categories that is homotopically fully faithful when restricted to test objects. Furthermore, it is homotopically essentially surjective since $X \cong \mathrm{ev}_0(\mathrm{Sing}_J\, X)$ for any lean quasicategory X. By Proposition 7.12, we conclude that induced functor $\mathrm{ev}_0\colon \mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}} \to \widehat{\mathbf{sSet}}_{\mathrm{J}}$ (and hence $\mathrm{Sing}_J\colon \widehat{\mathbf{sSet}}_{\mathrm{J}} \to \mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$) is a right Quillen equivalence. $\square$

One can prove "profinite versions" of many of the properties that complete Segal spaces enjoy. The general strategy for proving such a profinite version of a given property is to reduce it to its classical counterpart. We will illustrate this by showing that the weak equivalences between complete Segal profinite spaces coincide with (a profinite version of) the Dwyer–Kan equivalences. This is done by exploiting two facts: that $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$ is a left Bousfield localization of the Reedy model structure $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{R}}$ (with respect to $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$), and that the weak equivalences between fibrant objects in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$ can be detected underlying in $\mathbf{sSet}_{\mathrm{KQ}}$. To state this explicitly, denote the functor that sends a simplicial profinite set to its underlying simplicial set by $U\colon \widehat{\mathbf{sSet}} \to \mathbf{sSet}$. Note that this functor is right Quillen as a functor from Quick's model structure to the Kan–Quillen model structure, and that its left adjoint is the profinite completion functor.

Proposition 9.4 *A map $X \to Y$ between fibrant objects in* $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$ *is a weak equivalence if and only if $UX \to UY$ is a weak equivalence in* $\mathbf{sSet}_{\mathrm{KQ}}$.

Proof This follows from Theorem E.3.1.6 of [25], which states that the functor between the underlying ∞–categories of $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$ and $\mathbf{sSet}_{\mathrm{KQ}}$ induced by U (which is called "Mat" by Lurie) is conservative. Another way to deduce this proposition is to show that the weak equivalences between fibrant objects in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$ are the π_*–isomorphisms (as in the proof of Proposition 3.9 of [10]) and that the underlying group/set $U\pi_n(X,x)$ of the profinite group/set $\pi_n(X,x)$ agrees with $\pi_n(UX,x)$ for any fibrant $X \in \widehat{\mathbf{sSet}}_{\mathrm{Q}}$ and any $x \in X_0$. □

Since $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$ is a left Bousfield localization of the model category $\widehat{\mathbf{bisSet}}_{\mathrm{R}}$, which by Proposition 6.9 coincides with the Reedy model structure on $\widehat{\mathbf{bisSet}}$ with respect to $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$, we see that a map between complete Segal profinite spaces is a weak equivalence if and only if it is levelwise a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$. In particular, we obtain the following result:

Proposition 9.5 *A map $X \to Y$ between complete Segal profinite spaces is a weak equivalence if and only if for every t, the map $X_{t,\bullet} \to Y_{t,\bullet}$ is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$. In particular, $X \to Y$ is a weak equivalence between complete Segal profinite spaces if and only if $UX \to UY$ is a weak equivalence between complete Segal spaces.*

For a complete Segal profinite space X and two objects $x, y \in X_{0,0}$, ie two maps $\Delta^0 \underline{\times} \Delta^0 \to X$, we can mimic the classical definition of the mapping space by defining a profinite space $\mathrm{map}_X(x,y)$ as the pullback

$$\begin{tikzcd}
\mathrm{map}_X(x,y) \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & X_{1,\bullet} \arrow[d, "{(d_1,d_0)}"] \\
\Delta^0 \arrow[r, "{(x,y)}"] & X_{0,\bullet} \times X_{0,\bullet}
\end{tikzcd}$$

Since X is Reedy fibrant, the map $(d_1,d_0)\colon X_{1,\bullet} \to X_{0,\bullet} \times X_{0,\bullet}$ is a fibration in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$, and hence $\mathrm{map}_X(x,y)$ is a fibrant object in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$. Since $U\colon \widehat{\mathbf{bisSet}}_{\mathrm{CSS}} \to \mathbf{bisSet}_{\mathrm{CSS}}$ preserves limits, we see that $U(\mathrm{map}_X(x,y)) \cong \mathrm{map}_{UX}(x,y)$ for any complete Segal profinite space X. If $f\colon X \to Y$ is a map between complete Segal profinite spaces, then for any $x, y \in X_{0,0}$, we obtain a map $\mathrm{map}_X(x,y) \to \mathrm{map}_Y(fx, fy)$ from the universal property of the pullback. We call a map between complete Segal profinite spaces $f\colon X \to Y$ *fully faithful* if, for any $x, y \in X_{0,0}$, the map $\mathrm{map}_X(x,y) \to \mathrm{map}_Y(fx, fy)$ is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$. It follows from Proposition 9.4 that $X \to Y$ is fully faithful if and only if $UX \to UY$ is a fully faithful map of complete Segal spaces.

One can also mimic the classical definitions of a homotopy and of homotopy equivalences in a complete Segal space, and use this to define what it means for a map of complete Segal profinite spaces to be essentially surjective. An equivalent, but easier, way is to say that $X \to Y$ is *essentially surjective* if and only if the induced map $\pi_0 X_{0,\bullet} \to \pi_0 Y_{0,\bullet}$ is an epimorphism of profinite sets. Since $U\pi_0 Z \cong \pi_0 UZ$ for any fibrant object Z in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$, and since epimorphisms of profinite sets are detected underlying, we see that a map of complete Segal profinite spaces $X \to Y$ is essentially surjective if and only if $UX \to UY$ is.

Definition 9.6 A map between complete Segal profinite spaces is called a *Dwyer–Kan equivalence* or *DK–equivalence* if it is essentially surjective and fully faithful.

Theorem 9.7 *A map between complete Segal profinite spaces is a Dwyer–Kan equivalence if and only if it is a weak equivalence in* $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$.

Proof As explained above Definition 9.6, $f : X \to Y$ is essentially surjective and fully faithful if and only if $UX \to UY$ is so. By Proposition 7.6 of [35], this is the case if and only if $UX \to UY$ is a weak equivalence in $\mathbf{bisSet}_{\mathrm{CSS}}$. By Proposition 9.5, this is equivalent to $X \to Y$ being a weak equivalence in $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$. □

One can lift Proposition 9.5 and Theorem 9.7 to analogous results about weak equivalences between profinite quasicategories using the Quillen equivalences ev_0 and Sing_J between $\widehat{\mathbf{sSet}}_{\mathrm{J}}$ and $\mathbf{bis}\widehat{\mathbf{Set}}_{\mathrm{CSS}}$.

Proposition 9.8 *A map $X \to Y$ between profinite quasicategories is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$ if and only if $UX \to UY$ is a weak equivalence in $\mathbf{sSet}_{\mathrm{J}}$.*

Proof Let $f : X \to Y$ be a map between profinite quasicategories. If f is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$, then $Uf : UX \to UY$ is a weak equivalence of quasicategories since $U : \widehat{\mathbf{sSet}}_{\mathrm{J}} \to \mathbf{sSet}_{\mathrm{J}}$ is right Quillen. Conversely, suppose Uf is a weak equivalence of quasicategories. Then $\mathrm{Sing}_J\, Uf : \mathrm{Sing}_J(UX) \to \mathrm{Sing}_J(UY)$ is a weak equivalence between complete Segal spaces. Note that $\mathrm{Sing}_J \circ U \simeq U \circ \mathrm{Sing}_J$, since both functors preserve cofiltered limits and they agree on lean simplicial sets. By Proposition 9.5, $\mathrm{Sing}_J\, X \to \mathrm{Sing}_J\, Y$ is a weak equivalence between complete Segal profinite spaces. Since ev_0 is right Quillen, the original map $\mathrm{ev}_0 \,\mathrm{Sing}_J\, X \cong X \to Y \cong \mathrm{ev}_0\, \mathrm{Sing}_J\, Y$ is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$. □

For a profinite quasicategory X and two 0–simplices $x, y \in X_0$ (ie maps $\Delta^0 \to X$), we define $\mathrm{map}_X(x, y)$ as the pullback

$$\begin{tikzcd} \mathrm{map}_X(x,y) \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner", very near start] & X^{\Delta^1} \arrow[d, "{(\mathrm{ev}_0, \mathrm{ev}_1)}"] \\ \Delta^0 \arrow[r, "{(x,y)}"] & X \times X \end{tikzcd}$$

Since the right-hand vertical map is obtained by cotensoring with the cofibration $\partial\Delta^1 \hookrightarrow \Delta^1$, it must be a fibration in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$. In particular, $\mathrm{map}_X(x, y)$ is fibrant in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$. One can show that, analogously to the classical case, $\mathrm{map}_X(x, y)$ is actually fibrant in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$. However, the proof of this is technical and not necessary for what follows, so it is not included.

A map $f\colon X \to Y$ of profinite quasicategories induces a morphism $\mathrm{map}_X(x, y) \to \mathrm{map}_Y(fx, fy)$ for any $x, y \in X_0$ by the universal property of the pullback. We say that f is *fully faithful* if $\mathrm{map}_X(x, y) \to \mathrm{map}_Y(fx, fy)$ is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{J}}$ for any $x, y \in X_0$.[2] For a 1–simplex $\alpha \in X_1$ with $d_1\alpha = x$ and $d_0\alpha = y$, ie a 0–simplex in $\mathrm{map}_X(x, y)$, we say that α is a *homotopy equivalence* if $\Delta^1 \xrightarrow{\alpha} X$ extends to a map $J^1 \to X$. Here J^1 is viewed as a simplicial profinite set through the inclusion $\mathbf{sFinSet} \hookrightarrow \widehat{\mathbf{sSet}}$. We say that a map of profinite quasicategories $f\colon X \to Y$ is *essentially surjective* if for any $y \in Y_0$, there exists an $x \in X_0$ and an $\alpha \in \mathrm{map}_Y(fx, y)$ such that α is a homotopy equivalence.

Since $U\colon \widehat{\mathbf{sSet}} \to \mathbf{sSet}$ preserves pullbacks, we see that $U\,\mathrm{map}_X(x, y) \cong \mathrm{map}_{UX}(x, y)$. By Proposition 9.8, a map $X \to Y$ of profinite quasicategories is fully faithful if and only if $UX \to UY$ is. Since $\mathrm{Hom}(J^1, X) \cong \mathrm{Hom}(J^1, UX)$ for any $X \in \widehat{\mathbf{sSet}}$, we also see that $X \to Y$ is essentially surjective if and only if $UX \to UY$ is.

Definition 9.9 A map between profinite quasicategories is called a *Dwyer–Kan equivalence* or *DK–equivalence* if it is essentially surjective and fully faithful.

Theorem 9.10 *A map between profinite quasicategories is a Dwyer–Kan equivalence if and only if it is a weak equivalence in* $\widehat{\mathbf{sSet}}_{\mathrm{J}}$.

Proof A map $X \to Y$ of profinite quasicategories is a DK–equivalence if and only if $UX \to UY$ is. Since the weak equivalences between fibrant objects in $\mathbf{sSet}_{\mathrm{J}}$ are

[2]Since the simplicial profinite sets $\mathrm{map}_X(x, y)$ and $\mathrm{map}_Y(fx, fy)$ are actually fibrant in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$, this is equivalent to asking that $\mathrm{map}_X(x, y) \to \mathrm{map}_Y(fx, fy)$ is a weak equivalence in $\widehat{\mathbf{sSet}}_{\mathrm{Q}}$.

exactly the DK–equivalences, we conclude from Proposition 9.8 that a map of profinite quasicategories $X \to Y$ is a DK–equivalence if and only if it is a weak equivalence. □

Appendix Comparison to the ∞–categorical approach

The goal of this appendix is to compare the model structures on $\mathrm{Ind}(\mathbf{C})$ and $\mathrm{Pro}(\mathbf{C})$ constructed in this paper to the ∞–categorical approach to ind- and pro-categories. Since the cases of ind- and pro-categories are dual, we only treat the case of ind-categories and dualize the main result at the end of this appendix.

Given a cofibration test category $\mathbf{C}$, the underlying ∞–category of the completed model structure on $\mathrm{Ind}(\mathbf{C})$ will be denoted by $\mathrm{Ind}(\mathbf{C})_\infty$. Recall that this ∞–category is defined as the homotopy-coherent nerve of the full simplicial subcategory spanned by the fibrant-cofibrant objects. We will show that if $(\mathbf{C}, \mathbf{T})$ is a cofibration test category with a suitable assumption on $\mathbf{T}$, then the ∞–category $\mathrm{Ind}(\mathbf{C})_\infty$ is equivalent to $\mathrm{Ind}(N(\mathbf{T}))$. Here $N(\mathbf{T})$ is the homotopy-coherent nerve of the simplicial category $\mathbf{T}$, and Ind denotes the ∞–categorical version of the ind-completion as defined in [24, Definition 5.3.5.1].

Warning A.1 There is a subtlety here that we should point out: if $(\mathbf{C}, \mathbf{T})$ is a cofibration test category with respect to the Joyal model structure on **sSet**, meaning that items (ii) and (iii) of Definition 3.3 hold with respect to the trivial cofibrations and weak equivalences of $\mathbf{sSet}_\mathrm{J}$, then the "mapping spaces" of $\mathbf{T}$ are quasicategories but not necessarily Kan complexes. Recall that any quasicategory X contains a maximal Kan complex, which we will denote by $k(X)$. Since this functor k preserves cartesian products, any category enriched in quasicategories can be replaced by a category enriched in Kan complexes by applying the functor k to the simplicial hom. If $(\mathbf{C}, \mathbf{T})$ is a cofibration test category with respect to $\mathbf{sSet}_\mathrm{J}$, then we will abusively write $N(\mathbf{T})$ for the simplicial set obtained by first applying the functor k to all the mapping spaces in $\mathbf{T}$, and then applying the homotopy-coherent nerve. Similarly, by the underlying infinity category $\mathrm{Ind}(\mathbf{C})_\infty$ of $\mathrm{Ind}(\mathbf{C})$, we mean the quasicategory obtained by taking the full subcategory on fibrant-cofibrant objects, applying k to all mapping spaces, and then taking the homotopy-coherent nerve.

Since $\mathrm{Ind}(\mathbf{C})_\infty$ is the underlying ∞–category of a combinatorial model category, we see that it is complete and cocomplete. Furthermore, since $\mathbf{T}$ is a full subcategory of the fibrant-cofibrant objects in $\mathrm{Ind}(\mathbf{C})$, we see that the inclusion $\mathbf{T} \hookrightarrow \mathrm{Ind}(\mathbf{C})$ induces a fully faithful inclusion $N(\mathbf{T}) \hookrightarrow \mathrm{Ind}(\mathbf{C})_\infty$. By Proposition 5.3.5.10 of [24], this inclusion extends canonically to a filtered colimit-preserving functor $F\colon \mathrm{Ind}(N(\mathbf{T})) \to \mathrm{Ind}(\mathbf{C})_\infty$. In order for this functor to be an equivalence, any object in $\mathrm{Ind}(\mathbf{C})$ needs to be equivalent to a filtered homotopy colimit of objects in $\mathbf{T}$. This means that $\mathbf{T}$ should be "large enough" for this to hold. It turns out that this is the case if $\mathbf{T}$ is closed under pushouts along cofibrations (in the sense of Definition 7.10).

Theorem A.2 *Let $(\mathbf{C},\mathbf{T})$ be a cofibration test category and suppose that $\mathbf{T}$ is closed under pushouts along cofibrations. Then the canonical functor*

$$F\colon \mathrm{Ind}(N(\mathbf{T})) \to \mathrm{Ind}(\mathbf{C})_\infty$$

is an equivalence of quasicategories.

Remark A.3 In many of the examples discussed in this paper, the category $\mathbf{T}$ of test objects is closed under pushouts along cofibrations. For example, this is the case if $(\mathbf{C},\mathbf{T})$ has inherited the structure of a cofibration test category from some model category $\mathcal{E}$ in the sense of Example 3.6.

Remark A.4 If $(\mathbf{C},\mathbf{T})$ is a cofibration test category, then one can always "enlarge" the full subcategory $\mathbf{T}$ together with the sets of (trivial) cofibrations to obtain a cofibration test category $(\mathbf{C},\mathbf{T}')$ such that $\mathbf{T}'$ is closed under pushouts along cofibrations, and for which the completed model structures $\mathrm{Ind}(\mathbf{C},\mathbf{T})$ and $\mathrm{Ind}(\mathbf{C},\mathbf{T}')$ coincide. To see this, note that we can define $\mathbf{T}'$ to consist of all objects in $\mathbf{C}$ that are cofibrant in $\mathrm{Ind}(\mathbf{C},\mathbf{T})$, and that we can define the (trivial) cofibrations of $(\mathbf{C},\mathbf{T}')$ to be the trivial cofibrations of $\mathrm{Ind}(\mathbf{C},\mathbf{T})$ between objects of $\mathbf{T}'$; that is, we endow $\mathbf{C}$ with the structure of a cofibration test category inherited from $\mathrm{Ind}(\mathbf{C},\mathbf{T})$; see Example 3.6. It is then clear that the model structures $\mathrm{Ind}(\mathbf{C},\mathbf{T})$ and $\mathrm{Ind}(\mathbf{C},\mathbf{T}')$ coincide, and that $\mathbf{T}'$ is closed under pushouts along cofibrations. In particular, we see by Theorem A.2 that the underlying ∞–category of $\mathrm{Ind}(\mathbf{C},\mathbf{T})$ can be described as the ind-category of the small ∞–category $N(\mathbf{T}')$, which contains $N(\mathbf{T})$ as a full subcategory.

Before proving this theorem, we will prove the following rectification result.

Lemma A.5 *Let $(\mathbf{C},\mathbf{T})$ be a cofibration test category such that $\mathbf{T}$ is closed under pushouts along cofibrations, and let I be a poset with the property that $I_{<i}$ is finite for every i. For any diagram $X\colon N(I) \to N(\mathbf{T})$, there exists a strict diagram $Y\colon I \to \mathbf{T}$*

such that $N(Y)\colon N(I) \to N(\mathbf{T})$ is naturally equivalent to X. This diagram Y can be constructed in such a way that for any $i \in I$, the map

$$\operatorname*{colim}_{j<i} Y_j \to Y_i$$

is a composition of two pushouts of cofibrations in $\mathbf{T}$.

The following lemma is needed for the proof.

Lemma A.6 *Let $(\mathbf{C}, \mathbf{T})$ be a cofibration test category and let $\{Y_i\}_{i \in I}$ be a diagram in $\mathbf{T}$ indexed by a finite poset such that for any $i \in I$, the map*

$$\operatorname*{colim}_{j<i} Y_j \to Y_i$$

is a finite composition of pushouts of cofibrations of $(\mathbf{C}, \mathbf{T})$. Then, for any $k \in I$, the map

$$Y_k \to \operatorname*{colim}_{i \in I} Y_i$$

is a finite composition of pushouts of cofibrations. In particular, if $\mathbf{T}$ is closed under pushouts along cofibrations, then $\operatorname{colim}_i Y_i$ is an object of $\mathbf{T}$.

Proof This follows from the dual of [7, Proposition 2.17]. For the convenience of the reader, we spell out their argument in our setting. Throughout this proof, we call a map in $\mathbf{C}$ *good* if it is a finite composition of pushouts of cofibrations. Note that any pushout of a good map is again a good map. A subposet $S \subseteq I$ is called a *sieve* if for any $i \in S$ and any $j \le i$ in I, one has $j \in S$. Write $Y_S = \operatorname{colim}_{j \in S} Y_j$ for any sieve S and $Y_{<i}$ for $Y_{I_{<i}} = \operatorname{colim}_{j<i} Y_j$ for any $i \in I$.

We will prove inductively that for two sieves $S \subseteq T$, the map $Y_S \to Y_T$ is good. This certainly holds if $|T| = 0$, so suppose this holds for $|T| < n$ and let sieves $S \subseteq T$ with $|T| = n$ be given. If $S = T$ then there is nothing to prove, so suppose that $S \subsetneq T$ and choose some maximal $i \in T \setminus S$. We then obtain a diagram

$$\begin{tikzcd}
 & Y_{<i} \arrow[r, "\text{good}"] \arrow[d] & Y_i \arrow[d] \\
Y_S \arrow[r, "\text{good}"] & Y_{T\setminus\{i\}} \arrow[r] \arrow[ur, phantom, "\ulcorner", very near end] & Y_T
\end{tikzcd}$$

where the square is a pushout. The map $Y_{<i} \to Y_i$ is good by assumption while $Y_S \to Y_{T\setminus\{i\}}$ is good by the induction hypothesis, so we conclude that $Y_S \to Y_T$ is good. This completes the induction, and the lemma now follows by considering the sieves $S = I_{\le k}$ and $T = I$. □

Proof of Lemma A.5 To distinguish colimits in quasicategories from homotopy colimits and ordinary colimits in simplicial categories, we will call them ∞–colimits. By a homotopy colimit of a diagram $Z\colon J \to \mathbf{T}$, we mean a cocone $Z_j \to W$ that induces an equivalence

$$\mathrm{Map}(W,t) \xrightarrow{\sim} \operatorname*{holim}_{j\in J} \mathrm{Map}(Z_j,t) \quad \text{for every } t \in \mathbf{T}.$$

The following proof is for the case that $(\mathbf{C},\mathbf{T})$ is a cofibration test category with respect to the Kan–Quillen model structure on $\mathbf{sSet}$. The same proof works if $(\mathbf{C},\mathbf{T})$ is a cofibration test category with respect to $\mathbf{sSet}_\mathrm{J}$; however, one has to replace $\mathrm{Map}(-,-)$ with the maximal Kan complex $k(\mathrm{Map}(-,-))$ contained in it, and one has to replace Δ^1 by the simplicial set H (as defined in Lemma 2.1) in the construction of the mapping cylinder below.

We will construct the diagram $Y : I \to \mathbf{T}$ and the equivalence $N(Y) \simeq X$ inductively. Let $i \in I$ be given and suppose that $Y|_{I_{<i}} : I_{<i} \to \mathbf{T}$ and $N(Y|_{I_{<i}}) \simeq X|_{N(I_{<i})}$ have been constructed and have the desired properties. We need to construct $Y|_{I_{\le i}} : I_{\le i} \to \mathbf{T}$ and an equivalence $N(Y_{I_{\le i}}) \simeq X|_{I_{\le i}}$ extending these. Write $Y_{<i} := \mathrm{colim}_{j<i} Y_j$. If $I_{<i}$ is empty, then $Y_{<i}$ is the initial object of $\mathbf{C}$ and hence an object of $\mathbf{T}$ by definition. If $I_{<i}$ is not empty, then it follows from the assumptions on $Y|_{I_{<i}}$ and Lemma A.6 that $Y_{<i}$ is an object of $\mathbf{T}$. The assumptions on $Y|_{I_{<i}}$ and the fact that $\mathrm{Ind}(\mathbf{C})$ is a simplicial model structure ensure that, for any $t \in \mathbf{T}$, the diagram $j \mapsto \mathrm{Map}(Y_j,t)$ is fibrant in the injective model structure on $\mathbf{sSet}^{(I_{<i})^{\mathrm{op}}}$. In particular, we see that

$$\mathrm{Map}(Y_{<i},t) \cong \lim_{j<i} \mathrm{Map}(Y_j,t) \simeq \operatorname*{holim}_{j<i} \mathrm{Map}(Y_j,t),$$

so $Y_{<i}$ is a homotopy colimit of the diagram $Y|_{I_{<i}}$. By Theorem 4.2.4.1 of [24], it follows that it is also the ∞–colimit of the diagram $N(Y|_{I_{<i}})\colon N(I_{<i}) \to N(\mathbf{T})$. In particular, if we define the diagram $Y'\colon I_{\le i} \to \mathbf{T}$ by $Y'_j = Y_j$ for all $j<i$ and $Y'_i = Y_{<i}$, then the natural equivalence $N(Y|_{I_{<i}}) \simeq X|_{N(I_{<i})}$ extends to a natural map $N(Y') \to X|_{N(I_{\le i})}$. The map $Y_{<i} = Y'_i \to X_i$ factors through the mapping cylinder

$$Y_{<i} \cong Y_{<i} \otimes \{0\} \to Y_{<i} \otimes \Delta^1 \cup_{Y_{<i}\otimes\{1\}} X_i \otimes \{1\} \xrightarrow{\sim} X_i$$

in $\mathbf{T}$, where the second map is a weak equivalence. The first map can be written as a composition of the following two pushouts of cofibrations:

$$\begin{tikzcd}
\varnothing \arrow[r] \arrow[d, tail] & Y_{<i}\otimes\{0\} \arrow[d] \\
X_i\otimes\{1\} \arrow[r] \arrow[ur, phantom, "\ulcorner"] & Y_{<i}\otimes\{0\}\sqcup X_i\otimes\{1\}
\end{tikzcd}
\qquad
\begin{tikzcd}
Y_{<i}\otimes\partial\Delta^1 \arrow[r] \arrow[d, tail] & Y_{<i}\otimes\{0\}\sqcup X_i\otimes\{1\} \arrow[d] \\
Y_{<i}\otimes\Delta^1 \arrow[r] \arrow[ur, phantom, "\ulcorner"] & Y_{<i}\otimes\Delta^1\cup_{Y_{<i}\otimes\{1\}} X_i\otimes\{1\}
\end{tikzcd}$$

Define $Y_i = Y_{<i} \otimes \Delta^1 \cup_{Y_{<i} \otimes \{1\}} X_i$. This defines a diagram $Y|_{I_{\leq i}} : I_{\leq i} \to \mathbf{T}$. The above factorization of $Y_{<i} \to X_i$ shows that we obtain a natural equivalence $N(Y|_{I_{\leq i}}) \simeq X|_{N(I_{\leq i})}$ extending the equivalence $N(Y|_{I_{<i}}) \simeq X|_{N(I_{<i})}$. $\square$

We are now ready to prove Theorem A.2.

Proof of Theorem A.2 The terms "colimit", "homotopy colimit" and "∞–colimit" are used in the same way as in the proof of Lemma A.5. We will denote mapping spaces in a simplicial category by "Map", while mapping spaces in a quasicategory are denoted by "map"; that is, with a lowercase m.

We will prove that the functor $F : \mathrm{Ind}(N(\mathbf{T})) \to \mathrm{Ind}(\mathbf{C})_\infty$ is fully faithful and essentially surjective. To see that F is fully faithful, we need to show that

$$\mathrm{map}_{\mathrm{Ind}(N(\mathbf{T}))}(X, Y) \to \mathrm{map}_{\mathrm{Ind}(\mathbf{C})_\infty}(F(X), F(Y))$$

is a weak equivalence for any $X, Y \in \mathrm{Ind}(N(\mathbf{T}))$. Since F preserves filtered ∞–colimits, it suffices to show this for $X \in N(\mathbf{T})$. Write $Y = \mathrm{colim}_i\, Y_i$ as a filtered ∞–colimit of a diagram $Y : I \to N(\mathbf{T})$ (which we also denote by Y). By Proposition 5.3.1.18 of [24] and Lemma E.1.6.4 of [25], we may assume without loss of generality that I is the nerve of a directed poset, which we also denote by I, with the property that $I_{<i}$ is finite for any $i \in I$. By Lemma A.5, we may replace Y by a strict diagram $Z : I \to \mathbf{T}$. Since a diagram as described in Lemma A.5 is cofibrant in the projective model structure on $\mathrm{Ind}(\mathbf{C})^I$, we see that the ind-object $Z = \{Z_i\}_{i \in I}$ is the homotopy colimit of the diagram $i \mapsto Z_i$. By Theorem 4.2.4.1 of [24], the object Z is an ∞–colimit of the diagram $Y : N(I) \to \mathrm{Ind}(\mathbf{C})_\infty$, hence Z is equivalent to $F(Y)$ (note that F preserves filtered colimits). In particular, we obtain a commutative diagram

$$\begin{tikzcd}
\mathrm{colim}_i\, \mathrm{map}_{\mathrm{Ind}(N(\mathbf{T}))}(X, Z_i) \arrow[r] \arrow[d] & \mathrm{colim}_i\, \mathrm{map}_{\mathrm{Ind}(\mathbf{C})_\infty}(FX, FZ_i) \arrow[d] \\
\mathrm{map}_{\mathrm{Ind}(N(\mathbf{T}))}(X, Y) \arrow[r] & \mathrm{map}_{\mathrm{Ind}(\mathbf{C})_\infty}(FX, FY)
\end{tikzcd}$$

Here the left-hand vertical map is an equivalence since objects of $\mathbf{T}$ are compact (in the ∞–categorical sense), while the right-hand vertical map is an equivalence since it is equivalent to $\mathrm{colim}_i\, \mathrm{Map}_{\mathrm{Ind}(\mathbf{C})}(X, Z_i) \to \mathrm{Map}_{\mathrm{Ind}(\mathbf{C})}(X, Z)$, which is an isomorphism since X is compact in $\mathrm{Ind}(\mathbf{C})$. The top horizontal map is an equivalence since F is by construction fully faithful when restricted to $N(\mathbf{T}) \subseteq \mathrm{Ind}(N(\mathbf{T}))$. We conclude that the bottom map is an equivalence and hence that F is fully faithful.

To see that F is essentially surjective, let X be a fibrant-cofibrant object in $\mathrm{Ind}(\mathbf{C})$. By Lemma 7.11, X is a directed colimit $\mathrm{colim}_i\, t_i$ of objects in $\mathbf{T}$. By Lemma 3.11, X is also a homotopy colimit of this diagram, hence X is an ∞–colimit of the diagram $\{t_i\}_i$ in the underlying ∞–category $\mathrm{Ind}(\mathbf{C})_\infty$. View $\{t_i\}_i$ as a diagram in $N(\mathbf{T})$ and let Y denote the ∞–colimit of this diagram in $\mathrm{Ind}(N(\mathbf{T}))$. Since F preserves filtered ∞–colimits, it follows that $F(Y) \simeq X$ and hence that F is essentially surjective. □

We automatically obtain the following dual result. Note that item (iv) of Theorem 1.1 stated in the introduction is a direct consequence of this theorem.

Theorem A.7 *Let* $(\mathbf{C}, \mathbf{T})$ *be a fibration test category and suppose that* $\mathbf{T}$ *is closed under pullbacks along fibrations* (see Definition 7.10)*. Then the canonical functor*

$$\mathrm{Pro}(N(\mathbf{T})) \to \mathrm{Pro}(\mathbf{C})_\infty$$

is an equivalence of quasicategories.

The main theorems of this appendix can be used to determine the underlying ∞–categories of many of the examples that were mentioned throughout this paper. Moreover, it shows that the homotopy theory of $\mathrm{Pro}(\mathbf{C})$ is often fully determined by the full simplicial subcategory $\mathbf{T}$ of $\mathbf{C}$. By way of illustration, we will single out one specific example. Namely, we will relate the "profinite" Joyal–Kan model structure (see Example 5.8) to the profinite stratified spaces defined in [8, Section 2.5]. Note that one can use similar arguments to determine the underlying ∞–categories of Quick's and Morel's model structures on $\widehat{\mathbf{sSet}}$ (cf [4, Section 7]) and of the profinite Joyal model structure (see Remark A.11).

Example A.8 Let P be a finite poset and let $\mathbf{L}_{/P}$ be the fibration test category defined in Example 5.8. The full subcategory of test objects $\mathbf{T}$ in this fibration test category consists of the fibrant objects of the Joyal–Kan model structure on $\mathbf{sSet}_{/P}$ whose total space is a lean simplicial set. We will call these lean P–stratified Kan complexes. They can be described explicitly as those inner fibrations $f\colon X \twoheadrightarrow P$ for which X is lean and the fiber above any point is a Kan complex. We prove in Lemma A.9 below that the homotopy-coherent nerve $N(\mathbf{T})$ of the category of lean P–stratified Kan complexes is equivalent to $\mathbf{Str}_{\pi,P}$, the ∞–category of π–finite P–stratified spaces defined in Definition 2.4.3 of [8]. By Theorem A.7, it now follows that the underlying ∞–category of the profinite Joyal–Kan model structure on $\widehat{\mathbf{sSet}}_{/P}$ is equivalent to $\mathrm{Pro}(\mathbf{Str}_{\pi,P})$, which is equivalent to the ∞–category of profinite P–stratified spaces defined in [8, Section 2.5].

We conclude this appendix by proving the lemma used in the above example.

Lemma A.9 *Let P be (the nerve of) a finite poset and let $\mathbf{T}$ be the full simplicial subcategory of $\mathbf{sSet}_{/P}$ spanned by the lean P–stratified Kan complexes. Then the homotopy-coherent nerve $N(\mathbf{T})$ is equivalent to the ∞–category of π–finite P–stratified spaces as defined in* [8, Definition 2.4.3].

Proof By slightly rephrasing the definition of "π–finite" given in [8], this comes down to proving that if $X \twoheadrightarrow P$ is a lean P–stratified Kan complex, then

(i) for any $p \in P$, the set $\pi_0(f^{-1}(p))$ is finite,

(ii) there exists an $n \in \mathbb{N}$ such that for all $x, y \in X$, the homotopy groups of $\mathrm{map}_X(x, y)$ vanish above degree n, and

(iii) for all $x, y \in X$, the Kan complex $\mathrm{map}_X(x, y)$ has finite homotopy groups,

and conversely that any P–stratified Kan complex $X \twoheadrightarrow P$ satisfying these properties is equivalent to a lean P–stratified Kan complex. If X is lean, then items (i) and (iii) follow since X is degreewise finite, while (ii) follows since X is coskeletal. For the converse, let a P–stratified Kan complex $X \twoheadrightarrow P$ satisfy these items. If we replace $X \twoheadrightarrow P$ by a minimal inner fibration $\widetilde{X} \to P$ (cf [24, Section 2.3.3]), then it is still a P–stratified Kan complex satisfying items (i)–(iii), so it suffices to show that $\widetilde{X}$ is lean. Since pullbacks of minimal fibrations are again minimal, it follows from (i) that $f^{-1}(p) \subseteq \widetilde{X}$ has finitely many 0–simplices for any $p \in P$, and hence that $\widetilde{X}$ has finitely many 0–simplices. Since P is (the nerve of) a poset, two maps $\Delta^n \to \widetilde{X}$ are homotopic relative to the boundary if and only if they are so over P. This implies that $\widetilde{X}$ is itself a minimal quasicategory, and hence degreewise finite by (iii) and Lemma A.10 below. It is proved in Proposition 2.3.4.18 of [24] that if $\widetilde{X}$ is a minimal quasicategory satisfying (ii), then it is coskeletal, so we conclude that $\widetilde{X}$ is lean. □

Lemma A.10 *Let X be a minimal quasicategory with finitely many 0–simplices and with the property that for any $x, y \in X_0$, the homotopy groups of $\mathrm{map}_X(x, y)$ are finite. Then X is degreewise finite.*

Proof Since X has finitely many 0–simplices, it suffices to show that for any $n \geq 1$ and any map $D\colon \partial\Delta^n \to X$, there exist finitely many n–simplices filling D. For $n = 1$ this is clear: by minimality, the number of 1–simplices from x to y in X agrees with $\pi_0\,\mathrm{map}(x, y)$, which is finite by assumption. Now assume $n > 1$ and let D be given.

Write E for the restriction of D to the face opposite to the n-th vertex, and write ∂E for the restriction of E to $\partial\Delta^{n-1}$. This restriction induces a left fibration $X_{E/} \to X_{\partial E/}$, where these slice categories are defined as in [21, Section 3]. Let z be the 0–simplex of X obtained by restricting D to the top vertex, and denote the fibers of $X_{E/}$ and $X_{\partial E/}$ above z by $\mathrm{map}(E,z)$ and $\mathrm{map}(\partial E,z)$, respectively. Since these are fibers of left fibrations over X, we see that these are Kan complexes. Note that the restriction of D to Λ^n_n defines a 0–simplex in $\mathrm{map}(\partial E,z)$. Now define $\mathrm{Fill}(D)$ as the pullback

$$\begin{tikzcd} \mathrm{Fill}(D) \arrow[r] \arrow[d] \arrow[dr, phantom, "\lrcorner"] & \mathrm{map}(E,z) \arrow[d, two heads] \\ \{D|_{\Lambda^n_n}\} \arrow[r] & \mathrm{map}(\partial E,z) \end{tikzcd}$$

It is clear that the 0–simplices of $\mathrm{Fill}(D)$ correspond to n–simplices in X that fill D. A 1–simplex in $\mathrm{Fill}(D)$ between two such n–simplices f,g in X is exactly an $(n+1)$–simplex $h\colon \Delta^{n+1} \to X$ such that $d_n h = f$, $d_{n+1} h = g$ and $d_i h = d_i s_m f$ for any $i<n$. Given such an $(n+1)$–simplex h, the sequence $(s_0 f, s_1 f, \dots, s_{n-1} f, h)$ defines a homotopy $\Delta^n \times \Delta^1 \to X$ between f and g relative to $\partial\Delta^n$.[3] In particular, by minimality of X, the existence of such an $(n+1)$–simplex h implies that $f=g$, and hence the number of elements in $\pi_0(\mathrm{Fill}(D))$ equals the number of fillers of $D\colon \partial\Delta^n \to X$.

Since $X_{E/} \to X_{\partial E/}$ is a left fibration and $\mathrm{map}(\partial E,z)$ is a Kan complex, the restriction $\mathrm{map}(E,z) \to \mathrm{map}(\partial E,z)$ is a Kan fibration and hence $\mathrm{Fill}(D)$ is the homotopy fiber of $\mathrm{map}(E,z) \twoheadrightarrow \mathrm{map}(\partial E,z)$. In particular, if $\mathrm{map}(E,z)$ and $\mathrm{map}(\partial E,z)$ have finite homotopy groups, then $\mathrm{Fill}(D)$ does as well, and hence D has finitely many fillers. If we let y denote the top vertex of the $(n-1)$–simplex E, then $\mathrm{map}(E,z) \simeq \mathrm{map}(y,z)$, which has finite homotopy groups by assumption. To see that $\mathrm{map}(\partial E,z)$ has finite homotopy groups, note that

$$\mathrm{map}(\partial E,z) = \lim_{x \in \mathrm{nd}(\partial\Delta^n)^{\mathrm{op}}} \mathrm{map}(E|_x, z),$$

where $\mathrm{nd}(\partial\Delta^n)$ denotes the poset of nondegenerate simplices of $\partial\Delta^{n-1}$. This follows from the fact that the join of simplicial sets $\star$ preserves connected colimits. We see that for any $x \in \mathrm{nd}(\partial\Delta^n)$, the Kan complex $\mathrm{map}(E|_x, z)$ is equivalent to $\mathrm{map}(y,z)$, where y denotes the top vertex of $E|_x$. In particular, it has finite homotopy groups. Note that

[3]The converse is also true: if there is a homotopy between f and f' relative to $\partial\Delta^n$, then there exists an $(n+1)$–simplex h in X with the given property. A proof of this statement can be obtained by slightly modifying the proof of Theorem I.8.2 in [23] in such a way that one only needs to fill inner horns.

the diagram $x \mapsto \mathrm{map}(E|_x, z)$ is injectively fibrant since the diagram $\{x\}_{x \in \mathrm{nd}(\partial\Delta^n)}$ is cofibrant in the projective model structure on $\mathbf{sSet}^{\mathrm{nd}(\partial\Delta^n)}$. In particular, $\mathrm{map}(\partial E, z)$ is a finite homotopy limit of spaces with finite homotopy groups, so it has finite homotopy groups as well. We conclude that there are finitely many n–simplices filling $D\colon \partial\Delta^n \to X$. □

Remark A.11 It follows as in the proofs of Lemmas A.9 and A.10 that a quasicategory is equivalent to a lean quasicategory if and only if it has finitely many objects up to equivalence and all its mapping spaces have finite homotopy groups that vanish above a certain dimension; let us call such quasicategories *π–finite*. Applying Theorem A.7 to the fibration test category $\mathbf{L}_{\mathrm{J}}$ of Example 5.7 shows that the underlying ∞–category of the profinite Joyal model structure $\widehat{\mathbf{sSet}}_{\mathrm{J}}$ (and hence also of $\widehat{\mathbf{bisSet}}_{\mathrm{CSS}}$) is equivalent to $\mathrm{Pro}(\mathbf{Cat}_{\infty,\pi})$, where $\mathbf{Cat}_{\infty,\pi}$ denotes the ∞–category of π–finite ∞–categories.

References

[1] **D W Anderson**, *Fibrations and geometric realizations*, Bull. Amer. Math. Soc. 84 (1978) 765–788 MR Zbl

[2] **M Artin**, **B Mazur**, *Etale homotopy*, Lecture Notes in Math. 100, Springer (1969) MR Zbl

[3] **I Barnea**, *The two out of three property in ind-categories and a convenient model category of spaces*, Theory Appl. Categ. 32 (2017) 620–651 MR Zbl

[4] **I Barnea**, **Y Harpaz**, **G Horel**, *Pro-categories in homotopy theory*, Algebr. Geom. Topol. 17 (2017) 567–643 MR Zbl

[5] **I Barnea**, **T M Schlank**, *A new model for pro-categories*, J. Pure Appl. Algebra 219 (2015) 1175–1210 MR Zbl

[6] **I Barnea**, **T M Schlank**, *From weak cofibration categories to model categories*, preprint (2016) arXiv

[7] **I Barnea**, **T M Schlank**, *A projective model structure on pro-simplicial sheaves, and the relative étale homotopy type*, Adv. Math. 291 (2016) 784–858 MR Zbl

[8] **C Barwick**, **S Glasman**, **P Haine**, *Exodromy*, preprint (2018) arXiv

[9] **T Blom**, **I Moerdijk**, *Profinite ∞–operads*, Adv. Math. 408 (2022) art. id. 108601 MR Zbl

[10] **P Boavida de Brito**, **G Horel**, **M Robertson**, *Operads of genus zero curves and the Grothendieck–Teichmüller group*, Geom. Topol. 23 (2019) 299–346 MR Zbl

[11] **D Dugger**, *Combinatorial model categories have presentations*, Adv. Math. 164 (2001) 177–201 MR Zbl

[12] **D A Edwards**, **H M Hastings**, *Čech and Steenrod homotopy theories with applications to geometric topology*, Lecture Notes in Math. 542, Springer (1976) MR Zbl

[13] **P Gabriel**, **M Zisman**, *Calculus of fractions and homotopy theory*, Ergebnisse der Math. 35, Springer (1967) MR Zbl

[14] **P G Goerss**, **J F Jardine**, *Simplicial homotopy theory*, Progr. Math. 174, Birkhäuser, Basel (1999) MR Zbl

[15] **A Grothendieck**, **J L Verdier**, *Préfaisceaux*, from "Théorie des topos et cohomologie étale des schémas, Tome 1: Théorie des topos (SGA 4_1)" (M Artin, A Grothendieck, J L Verdier, editors), Lecture Notes in Math. 269, Springer (1972) Exposé I MR Zbl

[16] **P J Haine**, *On the homotopy theory of stratified spaces*, preprint (2018) arXiv

[17] **P S Hirschhorn**, *Model categories and their localizations*, Mathematical Surveys and Monographs 99, Amer. Math. Soc., Providence, RI (2003) MR Zbl

[18] **G Horel**, *Profinite completion of operads and the Grothendieck–Teichmüller group*, Adv. Math. 321 (2017) 326–390 MR Zbl

[19] **M Hovey**, *Model categories*, Mathematical Surveys and Monographs 63, Amer. Math. Soc., Providence, RI (1999) MR Zbl

[20] **D C Isaksen**, *Calculating limits and colimits in pro-categories*, Fund. Math. 175 (2002) 175–194 MR Zbl

[21] **A Joyal**, *Quasi-categories and Kan complexes*, J. Pure Appl. Algebra 175 (2002) 207–222 MR Zbl

[22] **A Joyal**, **M Tierney**, *Quasi-categories vs Segal spaces*, from "Categories in algebra, geometry and mathematical physics" (A Davydov, M Batanin, M Johnson, S Lack, A Neeman, editors), Contemp. Math. 431, Amer. Math. Soc., Providence, RI (2007) 277–326 MR Zbl

[23] **K Lamotke**, *Semisimpliziale algebraische Topologie*, Grundl. Math. Wissen. 147, Springer (1968) MR Zbl

[24] **J Lurie**, *Higher topos theory*, Annals of Mathematics Studies 170, Princeton Univ. Press (2009) MR Zbl

[25] **J Lurie**, *Spectral algebraic geometry*, preprint (2018) Available at `https://url.msp.org/Lurie-SAG`

[26] **S MacLane**, *Categories for the working mathematician*, Graduate Texts in Math. 5, Springer (1971) MR Zbl

[27] **M Makkai**, **J Rosický**, **L Vokřínek**, *On a fat small object argument*, Adv. Math. 254 (2014) 49–68 MR Zbl

[28] **J P May**, *Simplicial objects in algebraic topology*, Van Nostrand Mathematical Studies 11, Van Nostrand, Princeton, NJ (1967) MR Zbl

[29] **C V Meyer**, *Approximation filtrante de diagrammes finis par Pro-C*, Ann. Sci. Math. Québec 4 (1980) 35–57 MR Zbl

[30] **F Morel**, *Ensembles profinis simpliciaux et interprétation géométrique du foncteur T*, Bull. Soc. Math. France 124 (1996) 347–373 MR Zbl

[31] **G Quick**, *Profinite homotopy theory*, Doc. Math. 13 (2008) 585–612 MR Zbl

[32] **G Quick**, *Continuous group actions on profinite spaces*, J. Pure Appl. Algebra 215 (2011) 1024–1039 MR Zbl Corrected and updated version available at https://folk.ntnu.no/gereonq/ContinuousGspaces_updated

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

[34] **C Rezk**, *A model category of categories*, preprint (2000) Available at `https://faculty.math.illinois.edu/~rezk/papers.html`

[35] **C Rezk**, *A model for the homotopy theory of homotopy theory*, Trans. Amer. Math. Soc. 353 (2001) 973–1007 MR Zbl

[36] **L Ribes**, **P Zalesskii**, *Profinite groups*, 2nd edition, Ergebnisse der Math. (3) 40, Springer (2010) MR Zbl

[37] **E Riehl**, **D Verity**, *Elements of ∞–category theory*, Cambridge Studies in Advanced Mathematics 194, Cambridge Univ. Press (2022) MR Zbl

Department of Mathematical Sciences, University of Copenhagen
Copenhagen, Denmark

Mathematisch Instituut, Universiteit Utrecht
Utrecht, Netherlands

`blom@math.ku.dk`, `i.moerdijk@uu.nl`

Received: 24 May 2022

Guidelines for Authors

Submitting a paper to Algebraic & Geometric Topology

Papers must be submitted using the upload page at the AGT website. You will need to choose a suitable editor from the list of editors' interests and to supply MSC codes.

The normal language used by the journal is English. Articles written in other languages are acceptable, provided your chosen editor is comfortable with the language and you supply an additional English version of the abstract.

Preparing your article for Algebraic & Geometric Topology

At the time of submission you need only supply a PDF file. Once accepted for publication, the paper must be supplied in LaTeX, preferably using the journal's class file. More information on preparing articles in LaTeX for publication in AGT is available on the AGT website.

`arXiv` papers

If your paper has previously been deposited on the `arXiv`, we will need its `arXiv` number at acceptance time. This allows us to deposit the DOI of the published version on the paper's `arXiv` page.

References

Bibliographical references should be listed alphabetically at the end of the paper. All references in the bibliography should be cited at least once in the text. Use of BibTeX is preferred but not required. Any bibliographical citation style may be used, but will be converted to the house style (see a current issue for examples).

Figures

Figures, whether prepared electronically or hand-drawn, must be of publication quality. Fuzzy or sloppily drawn figures will not be accepted. For labeling figure elements consider the pinlabel LaTeX package, but other methods are fine if the result is editable. If you're not sure whether your figures are acceptable, check with production by sending an email to graphics@msp.org.

Proofs

Page proofs will be made available to authors (or to the designated corresponding author) in PDF format. Failure to acknowledge the receipt of proofs or to return corrections within the requested deadline may cause publication to be postponed.

ALGEBRAIC & GEOMETRIC TOPOLOGY

Volume 23 Issue 8 (pages 3417–3908) 2023