Geometry & Topology

Volume 21 (2017)

Issue 6 (pages 3191–3810)

msp

GEOMETRY & TOPOLOGY

msp.org/gt

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

The subscription price for 2017 is US $495/year for the electronic version, and $730/year (+$60, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP. Geometry & Topology is indexed by Mathematical Reviews, Zentralblatt MATH, Current Mathematical Publications and the Science Citation Index.

Geometry & Topology (ISSN 1465-3060 printed, 1364-0380 electronic) at Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840, is published 6 times per year and continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices. POSTMASTER: send address changes to Mathematical Sciences Publishers, c/o Department of Mathematics, University of California, 798 Evans Hall #3840, Berkeley, CA 94720-3840.

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

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

Geometry & Topology 21 (2017) 3191–3211

Independence of satellites of torus knots in the smooth concordance group

JUANITA PINZÓN-CAICEDO

The main goal of this article is to obtain a condition under which an infinite collection $\mathscr{F}$ of satellite knots (with companion a positive torus knot and pattern similar to the Whitehead link) freely generates a subgroup of infinite rank in the smooth concordance group. This goal is attained by examining both the instanton moduli space over a 4–manifold with tubular ends and the corresponding Chern–Simons invariant of the adequate 3–dimensional portion of the 4–manifold. More specifically, the result is derived from Furuta's criterion for the independence of Seifert fibred homology spheres in the homology cobordism group of oriented homology 3–spheres. Indeed, we first associate to $\mathscr{F}$ the corresponding collection of 2–fold covers of the 3–sphere branched over the elements of $\mathscr{F}$ and then introduce definite cobordisms from the aforementioned covers of the satellites to a number of Seifert fibered homology spheres. This allows us to apply Furuta's criterion and thus obtain a condition that guarantees the independence of the family $\mathscr{F}$ in the smooth concordance group.

57M25; 57N70, 58J28

1 Introduction

A knot is a smooth embedding of S^1 into S^3. Two knots K_0 and K_1 are said to be smoothly concordant if there is a smooth embedding of $S^1 \times [0,1]$ into $S^3 \times [0,1]$ that restricts to the given knots at each end. Requiring such an embedding to be locally flat instead of smooth gives rise to the weaker notion of topological concordance. Both kinds of concordance are equivalence relations, and the sets of smooth and topological concordance classes of knots, denoted by $\mathcal{C}_\infty$ and $\mathcal{C}_{\mathrm{TOP}}$, respectively, are abelian groups with connected sum as their binary operation. In both cases the identity element is the concordance class of the unknot and the knots in that class are known as smoothly slice and topologically slice, respectively. The algebraic structure of $\mathcal{C}_\infty$ and $\mathcal{C}_{\mathrm{TOP}}$ is a much studied object in low-dimensional topology, as is the concordance class of the unknot. Identifying the set of knots that are topologically slice but not smoothly slice is a challenging topic, among other reasons because these knots reveal subtle properties of differentiable structures in dimension four; see Gompf and Stipsicz [10, page 522].

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3191

One way to approach this problem is by using satellite operations to construct families of knots and studying their concordance properties. To define a satellite operation we start with a given knot B, embedded in an unknotted solid torus $V \subseteq S^3$, and a second knot $K \subseteq S^3$. The satellite knot with pattern $B \subseteq V$ and companion K is denoted by $B(K)$ and is obtained as the image of B under the embedding of V in S^3 that knots V as a tubular neighborhood of K. Freedman's theorem [6; 7] (see also Freedman and Quinn [8]) implies that if the pattern B is an unknot in S^3 and is trivial in $H_1(V;\mathbb{Z})$, then the satellite $B(K)$ is topologically slice.

Whitehead doubles are an important example of such satellites and are obtained by using the Whitehead link (Figure 1, left) as the pattern of the operation. Similar examples arise by considering Whitehead-like patterns D_n (Figure 2). Because the knot D_n is trivial in S^3, every satellite knot with pattern D_n is topologically slice, and classical invariants do not detect information about their smooth concordance type. Thus, smooth techniques like gauge theory are necessary to obtain that information. In this article we use the theory of SO(3) instantons to establish an obstruction for a family of Whitehead-like satellites of positive torus knots to be dependent in the smooth concordance group. The main result is the following:

Theorem 6.2 *Let $\{(p_i, q_i)\}_i$ be a sequence of relatively prime positive integers and n_i a positive and even integer for $i = 1, 2, \dots$. Then, if*

$$p_i q_i (2n_i p_i q_i - 1) < p_{i+1} q_{i+1} (n_{i+1} p_{i+1} q_{i+1} - 1),$$

the collection $\{D_{n_i}(T_{p_i q_i})\}_{i=1}^{\infty}$ is an independent family in $\mathcal{C}_\infty$.

It is important to mention that the case $n_i = 2$ is a result of Hedden and Kirk [12] and the previous theorem is a generalization of their work.

The proof of Theorem 6.2 is based on a technique pioneered by Akbulut (and made public at a 1983 NSF-CBMS Regional Conference in Santa Barbara) and later expanded by Cochran and Gompf [2] among others. The starting point of Akbulut's technique is to assign to each satellite knot $D_n(T_{p,q})$ the 2–fold cover of S^3 branched over the knot $D_n(T_{p,q})$, since an obstruction to the cover from bounding results in an obstruction to $D_n(T_{p,q})$ from being slice. The next step is to construct a negative definite cobordism W from the 2–fold cover $\Sigma = \Sigma_2(D_n(T_{p,q}))$ to the Seifert fibered homology sphere $-\Sigma(2,3,5)$ and then glue W to the negative definite 4–manifold E_8 along their common boundary. The last step is to notice that if Σ bounded a $\mathbb{Z}/2$–homology 4–ball Q, then the manifold $X = Q \cup W \cup E_8$ would be a closed 4–manifold with negative definite intersection form given by $m\langle -1\rangle \oplus E_8$ for some integer $m > 0$. However, Donaldson's diagonalization theorem prevents the existence of such a manifold, thus showing that Q cannot exist and that $D_n(T_{p,q})$ is not smoothly slice. This same technique can be used to prove that $D_n(T_{p,q})$ has infinite order in the smooth concordance group $\mathcal{C}_\infty$ and so we see that (1) nontriviality and order in $\mathcal{C}_\infty$ can be obtained by studying 2–fold branched covers and the 4–manifolds they bound, and (2) the existence of certain 4–manifolds with a fixed boundary can be obstructed using gauge-theoretical techniques. This can be further extended to get to independence by focusing on the 3–manifolds. Indeed, a 3–manifold counterpart to the group $\mathcal{C}_\infty$ can be obtained by considering $\mathbb{Z}/2$–homology 3–spheres under an equivalence relation stemming from the notion of cobordisms; the details of this correspondence will be explained in Lemma 4.1. Since cobordisms will play a fundamental role in this paper, and to clarify some terminology, we include a precise definition of an oriented cobordism.

Definition 1.1 Two closed, oriented 3–manifolds Y_0 and Y_1 are said to be oriented cobordant if there exists a compact, oriented 4–manifold W with oriented boundary $\partial W = -Y_0 \sqcup Y_1$. The manifold W is called a cobordism from Y_0 to Y_1, with Y_0 referred to as the incoming boundary component and Y_1 the outcoming boundary component. Moreover, if W is positive (negative) definite, then W is called a positive (negative) definite cobordism.

The 3–manifold equivalence relation corresponding to concordance is the following: call two oriented $\mathbb{Z}/2$–homology spheres Σ_0 and Σ_1 homology cobordant if there is a cobordism W from Σ_0 to Σ_1 such that $H_*(W;\mathbb{Z}/2) = H_*(I \times S^3;\mathbb{Z}/2)$. The set of homology cobordism classes of $\mathbb{Z}/2$–homology spheres forms an abelian group $\Theta^3_{\mathbb{Z}/2}$ with connected sum as the group operation. The same notion with $\mathbb{Z}/2$ replaced with $\mathbb{Z}$ gives rise to the $\mathbb{Z}$–homology cobordism group $\Theta^3_{\mathbb{Z}}$. Independence of infinite families of knots in $\mathcal{C}_\infty$ can then be proven by establishing independence of the corresponding families of 2–fold branched covers in $\Theta^3_{\mathbb{Z}/2}$. To prove the latter we will use a generalization of the following gauge-theoretical result:

Theorem (Furuta [9]; see also Fintushel and Stern [5]) *Let $R(p,q,r)$ be the Fintushel–Stern invariant for $\Sigma(p,q,r)$ and suppose that a sequence $\Sigma_i = \Sigma(p_i,q_i,r_i)$ for $i = 1,2,\dots$ satisfies that $R(p_i,q_i,r_i) > 0$ for $i = 1,2,\dots$. Then, if*

$$p_i q_i r_i < p_{i+1} q_{i+1} r_{i+1},$$

the homology classes $[\Sigma_i]$ for $i = 1,2,\dots$ are linearly independent over $\mathbb{Z}$ in $\Theta^3_{\mathbb{Z}}$.

In a manner similar to Akbulut's technique, the gauge-theoretical result cannot be applied directly; in both cases it is necessary to first construct definite cobordisms from the 2–fold cover $\Sigma_2(D_n(T_{p,q}))$ to Seifert fibered homology spheres and then apply Furuta's criterion for independence. This approach was used by Hedden and Kirk [12] to establish conditions under which an infinite family of Whitehead doubles of positive torus knots is independent in $\mathcal{C}_\infty$. Nonetheless, their proof involves a complicated computation of bounds for the minimal Chern–Simons invariant of $\Sigma_2(D_2(T_{p,q}))$ and this can be sidestepped by introducing definite cobordisms from $\Sigma_2(D_2(T_{p,q}))$ to Seifert fibered homology 3–spheres. In this article we recover their result and generalize it to include more examples of satellite operations.

Outline In Section 2 we offer a brief description of satellite operations and present the important patterns. In Section 3 we review the theory of SO(3) instantons and the homology cobordism obstruction that derives from it. Then, in Section 4 we explore the topology of the 2–fold covers to later introduce the construction of the relevant cobordisms in Section 5. Finally, in Section 6 we prove the main result.

Acknowledgements The results in this paper originally formed the core of my PhD dissertation. I would like to thank my advisor Paul Kirk for his patient guidance and continual encouragement throughout my studies. I would also like to thank Charles Livingston for his careful reading of my dissertation and his helpful comments.

2 Patterns and satellite knots

The main goal of this article is to show independence of families of satellite knots in the smooth concordance group. This is done by considering satellite operations with pattern similar to the Whitehead link and companion a positive torus knot. In this section we describe the patterns of the relevant satellite operations.

Definition 2.1 Let $B \sqcup A$ be a 2–component link in S^3, where A is an unknot, and so $V = S^3 \setminus N(A)$ is an unknotted solid torus in S^3. For K any knot, consider $h\colon V \to S^3$ an orientation-preserving embedding taking V to a tubular neighborhood of K in such a way that a longitude of V (which is a meridian of A) is sent to a longitude of K; then $h(B)$ is the untwisted satellite knot with pattern $B \sqcup A$ and companion K and is usually denoted by $B(K)$.

A notable example of a satellite operation is provided by using the Whitehead link (Figure 1, left) as the pattern of the operation. The knots obtained in this way are called Whitehead doubles. The following figures show the pattern, companion, and satellite whenever we take the pattern $B \sqcup A$ to be the Whitehead link and the companion knot to be the right-handed trefoil, $T_{2,3}$.

In greater generality, we can add more twists to the clasp of Figure 1, left, to obtain the patterns included in Figure 2. These are the Whitehead-like patterns under consideration in the present article.

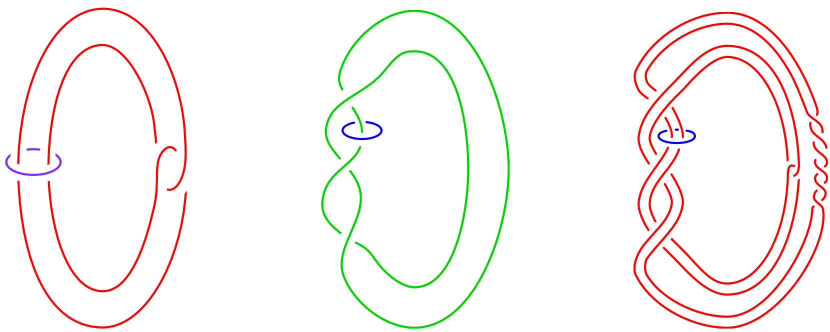


Figure 1: An example of a satellite. The Whitehead link (left), trefoil (center), and untwisted Whitehead double of the trefoil (right).

Since the pattern D_n, as a knot in S^3, is unknotted and $\mathrm{lk}(A, D_n) = 0$ whenever n is an even integer, the Alexander polynomial of the satellite knot $D_n(K)$ is $\Delta_{D_n(K)}(t) = 1$ [13, Theorem 6.15]. A theorem of Freedman [6; 7; 8] states that every knot with Alexander polynomial 1 is topologically slice. This implies that for any companion knot K, the satellite $D_n(K)$ is a topologically slice knot. We will later show that whenever $K = T_{p,q}$ with (p,q) a pair of positive and relatively prime integers, the satellite knots $D_n(T_{p,q})$ are not smoothly slice.

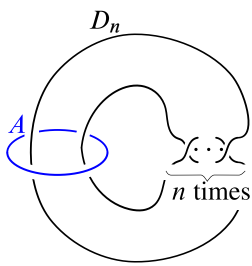


Figure 2: The Whitehead-like patterns D_n. In this figure, $n > 0$ denotes the number of positive half twists. Also, since we require $\mathrm{lk}(A, D_n) = 0$, we will further assume that n is an even integer.

3 Instanton cobordism obstruction

In this section we survey the theory of instantons on $\mathrm{SO}(3)$–bundles developed by Furuta [9] and Fintushel and Stern [4; 5] in the setting of orbifolds (ie manifolds with a special kind of singularities), and recast by Hedden and Kirk [11] in the setting of manifolds with tubular or cylindrical ends. Additionally, in this section we introduce the instanton cobordism obstruction, that is, the way in which the topology of the instanton moduli space obstructs the existence of certain 4–manifolds.

Following [9; 4; 5; 11], let p, q, r be positive and relatively prime integers, and consider the Seifert fibered sphere $\Sigma = \Sigma(p,q,r)$ and the mapping cylinder $\mathcal{W}$ of the Seifert projection $\Sigma \to S^2$. The latter space is a negative definite orbifold with boundary Σ and with three singularities, each of which has a neighborhood homeomorphic to a cone on a lens space. To avoid singularities, form a manifold $W = W(p,q,r)$ by removing the aforementioned neighborhoods from the mapping cylinder $\mathcal{W}$, and notice that

$$H^2(W;\mathbb{Z}) \cong H^2(W,\Sigma;\mathbb{Z}) \cong \mathbb{Z}.$$

One of the key components of the theory is that these groups have a preferred generator. Let e be the generator of $H^2(W;\mathbb{Z})$ and notice that this cohomology class determines an SO(2)–vector bundle $\mathcal{L}$ over W, which is trivial over Σ. In addition, if ε is the trivial real vector bundle of rank 1 over W, the bundle $\mathcal{L} \oplus \varepsilon$ is an SO(3)–vector bundle over W. Then, if X is a 4–manifold with Σ as one of its boundary components, one can form $M = X \cup_\Sigma W$ and, since $\mathcal{L} \oplus \varepsilon$ is trivial over Σ, it can be extended trivially to an SO(3)–vector bundle E over M.

For technical reasons originating from analytical considerations, it is necessary to attach to M cylindrical ends isometric to $[0,\infty) \times \partial M$ to form a noncompact manifold M_∞. One then considers the corresponding extension of the bundle E to M_∞ and studies connections A on E with finite energy, that is, connections for which the energy integral satisfies

$$\mathcal{E}(A) = \int_{M_\infty} \operatorname{Tr}(F_A \wedge *F_A) < \infty.$$

Here F_A is the curvature of A and $*$ is the Hodge star operator. However, one of the subtle variations present in the cylindrical end formulation of the theory of instantons is the presence of limiting connections on E that are determined by the cohomology class e. Modulo gauge equivalence, the class e uniquely determines a flat connection β_i on the restriction of $\mathcal{L}$ to each of the lens spaces in the boundary of W. Furthermore, if ϑ_i is the trivial connection on the restriction of ε to the i^{th} lens space, we can form $\alpha_i = (\beta_i, \vartheta_i)$ to obtain an SO(3)–connection on the restriction of E to the i^{th} lens space. Then, if we choose the trivial SO(3)–connection over every other boundary component of $X \cup W$, the tuple $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \theta, \dots, \theta)$ is the limiting flat connection and $(E, \boldsymbol{\alpha})$ is the adapted bundle (in the sense of [3]) to be considered.

For a positive number δ and an appropriate weighted Sobolev norm $\|\cdot\|_\delta$, the moduli space $\mathscr{M} = \mathscr{M}_\delta(E,\boldsymbol{\alpha})$ is the set of gauge equivalence classes of finite weighted norm SO(3)–connections A on E that limit to $\boldsymbol{\alpha}$ and that satisfy the anti-self-dual (ASD) equation

$$-F_A = *F_A.$$

In other words, $\mathscr{M}$ is the moduli space of instantons over M_∞. Then, perhaps after perturbing either the metric of M_∞ or the anti-self-dual equation, $\mathscr{M}$ can be shown to have the structure of a smooth manifold with some singular points. An in-depth account of the theory of instantons over manifolds with cylindrical ends can be found in [3].

In summary, the cohomology class e determines the adapted bundle $(E,\boldsymbol{\alpha})$. The next theorem shows that if X is a negative definite 4–manifold, the choice of e also gives information about the topology of the instanton moduli space $\mathscr{M}$ and thus, all the gauge theory over M_∞.

Theorem 3.1 *Let X be a negative definite 4–manifold whose boundary consists of the union of some Seifert fibered homology spheres $\Sigma_i = \Sigma(p_i,q_i,k_i p_i q_i - 1)$ for $i = 1,\dots,N$. Consider $W = W(p_N,q_N,k_N p_N q_N - 1)$ and form $M = X \cup_{\Sigma_N} W$. Let E be the* SO(3)*–bundle over M_∞ determined by the generator e of $H^2(W;\mathbb{Z})$.*

The moduli space $\mathscr{M}$ of finite energy instantons on E is a (possibly noncompact) smooth 1–manifold with boundary and with the following properties:

(a) *The number of boundary points of $\mathscr{M}$ is given by $C(e) = T/2^\beta$, where T is the order of the torsion subgroup of $H_1(X;\mathbb{Z})$ and*

$$\beta = \operatorname{rank}(H_1(X;\mathbb{Z}/2)) - \operatorname{rank}(H_1(X;\mathbb{Z})).$$

(b) *If $p_i q_i (k_i p_i q_i - 1) < p_{i+1} q_{i+1} (k_{i+1} p_{i+1} q_{i+1} - 1)$, then $\mathscr{M}$ is compact.*

In what follows we offer a broad idea of the proof. For a precise account we refer the reader to [11].

Using the theory of singular bundles over orbifolds, Fintushel and Stern compute the index for the bundle $L \oplus \varepsilon$ over $W(a_1,a_2,a_3)$ and give an explicit formula as

$$R(a_1,a_2,a_3) = \frac{2}{a_1a_2a_3} + \sum_{i=1}^{3} \frac{2}{a_i} \sum_{k=1}^{a_i-1} \cot\Big(\frac{\pi a k}{a_i^2}\Big)\cot\Big(\frac{\pi k}{a_i}\Big)\sin^2\Big(\frac{\pi k}{a_i}\Big).$$

Furthermore, Hedden and Kirk [11] show that whenever $R(a_1,a_2,a_3)$ is positive, it equals the dimension of the moduli space of instantons over the noncompact manifold M_∞ obtained from the augmented manifold $M = X \cup W$ for any 4–manifold X. A calculation using the Neumann–Zagier formula [16] shows that when p and q are relatively prime positive integers and $k \geq 1$, the Fintushel–Stern invariant for $\Sigma(p,q,kpq-1)$ is such that

$$R(p,q,kpq-1) = 1,$$

thus proving that $\mathscr{M}$ is a 1–dimensional space.

It can be shown that the boundary points of $\mathscr{M}$ correspond to reducible connections. Using results found in [18; 11] and some basic algebraic topology one can show that the number of reducible connections is given by $C(e) = T/2^{\beta}$, where T is the order of the torsion subgroup of $H_1(X;\mathbb{Z})$ and $\beta = \operatorname{rank}(H_1(X;\mathbb{Z}/2)) - \operatorname{rank}(H_1(X;\mathbb{Z}))$ as claimed in Theorem 3.1(a).

Finally, the question about compactness is in fact a question about convergence. To address this question, for any connection A on E that limits to $\boldsymbol{\alpha}$, consider the integral

$$-\frac{1}{8\pi^2}\int_{M_\infty} \operatorname{Tr}(F_A \wedge F_A).$$

The value of this integral can be shown to be independent of the choice of A and thus an invariant of the bundle $(E,\boldsymbol{\alpha})$. This invariant is usually denoted by $p_1(E,\boldsymbol{\alpha})$ and known as the Pontryagin number of $(E,\boldsymbol{\alpha})$. In addition, $p_1(E,\boldsymbol{\alpha})$ captures convergence of sequences of connections in $\mathscr{M}$ modulo gauge equivalence. Indeed, Uhlenbeck compactness for noncompact manifolds characterizes lack of convergence in $\mathscr{M}$ as taking one of two different forms: "bubbling" and "breaking". Bubbling happens when the curvature accumulates near a point inside a compact set in M_∞ and results in a change of the Pontryagin number of the bundle. In fact, by Uhlenbeck's removable singularities theorem [19], this change comes in multiples of 4 and so, if $p_1(E,\boldsymbol{\alpha})$ is less than 4, bubbling cannot occur. Breaking happens when a region appears in one of the cylindrical ends of M_∞ where the connection looks like an instanton on a tube that limits to a flat connection at either end of the tube. Furuta [9] shows that the curvature of the connection at such a region is nonzero and the energy of the connection is greater than or equal to the Chern–Simons invariant of the limiting connections. For ease of notation, for Y a 3–manifold denote by $\tau(Y)$ the minimum of the differences $\operatorname{cs}(Y,b) - \operatorname{cs}(Y,\boldsymbol{\alpha}|_Y) \in (0,4]$, where b ranges over all flat connections on $E|_Y$. So, if $p_1(E,\boldsymbol{\alpha})$ is less than $\tau(Y,e)$ for every connected component Y of ∂M, breaking cannot occur. In conclusion, if $p_1(E,\boldsymbol{\alpha}) < 4$ and $p_1(E,\boldsymbol{\alpha}) < \min\{\tau(Y) \mid Y \subseteq \partial M\}$, neither bubbling nor breaking can occur, and thus the previous inequalities constitute a compactness criterion for the moduli space $\mathscr{M}$. Computations of these quantities for the case at hand and proofs of the inequalities will show compactness of $\mathscr{M}$. First, an argument involving the intersection form of W shows that

$$p_1(E,\boldsymbol{\alpha}) = \frac{1}{p_N q_N (k_N p_N q_N - 1)} < 4$$

and can be found in [11]. Further, if L is any of the lens spaces in the boundary of W, then its minimum Chern–Simons invariant satisfies

$$\tau(L) \geq \min\left\{\frac{1}{p_N}, \frac{1}{q_N}, \frac{1}{k_N p_N q_N - 1}\right\} > p_1(E,\boldsymbol{\alpha}).$$

In addition, it is also known [9; 5] that $\tau(\Sigma(p,q,kpq-1)) = 1/(pq(kpq-1))$. Then, the condition

$$p_i q_i (k_i p_i q_i - 1) < p_{i+1} q_{i+1} (k_{i+1} p_{i+1} q_{i+1} - 1)$$

implies $p_1(E,\boldsymbol{\alpha}) < \min\{\tau(\Sigma_i) \mid i = 1,\dots,N-1\}$ and so, by the compactness criterion previously described, $\mathscr{M}$ is in fact a compact space as asserted in Theorem 3.1(b). This completes the sketch of the proof of Theorem 3.1.

To obtain the instanton cobordism obstruction further assume that $H_1(X;\mathbb{Z}/2) = 0$. In that case $H_1(X;\mathbb{Z})$ would be a torsion group with no even torsion and so β would be 0 and $T = |H_1(X;\mathbb{Z})|$ would be an odd integer. Therefore, $C(e) = |H_1(X;\mathbb{Z})|$ and the moduli space $\mathscr{M}$ would contain an odd number of reducible connections. However, by Theorem 3.1, $\mathscr{M}$ is a compact 1–dimensional manifold. Since a compact 1–dimensional manifold cannot have an odd number of boundary components, Theorem 3.1 obstructs the existence of a negative definite 4–manifold satisfying $H_1(X;\mathbb{Z}/2) = 0$. The following theorem is a reformulation of Theorem 3.1, with the additional hypothesis $H_1(X;\mathbb{Z}/2) = 0$, expressed in purely topological terms.

Theorem 3.2 *Let p_i and q_i be relatively prime integers and k_i a positive integer for $i = 1,\dots,N$. If $\{\Sigma_i\}_{i=1}^N$ is a family of Seifert fibred homology 3–spheres such that $\Sigma_i = \Sigma(p_i, q_i, k_i p_i q_i - 1)$ and satisfying*

(1) $$p_i q_i (k_i p_i q_i - 1) < p_{i+1} q_{i+1} (k_{i+1} p_{i+1} q_{i+1} - 1),$$

then no combination of elements in $\{\Sigma_i\}_{i=1}^N$ cobounds a smooth 4–manifold X with negative definite intersection form and such that $H_1(X;\mathbb{Z}/2) = 0$.

In summary, the crucial idea is that the topology of the instanton moduli space obstructs the existence of some definite 4–manifolds. Also key is the fact that the cohomology class e and the minimum Chern–Simons invariant of the boundary 3–manifolds provide important information about the topology of the moduli space. Note that the compactness criterion presented in Theorem 3.1(b) is precisely the criterion for the independence of a family of satellites of the form $D_n(T_{p,q})$.

4 Topological description of 2–fold covers

A useful method to study the algebraic structure of a group G is to consider homomorphisms $G \to H$ and use information about the algebraic structure of H. In the case of the smooth concordance group $\mathcal{C}_\infty$ it is common to associate to the concordance class of a knot K the equivalence class of the 2–fold cover of S^3 branched over K, $\Sigma_2(K)$, in the homology cobordism group of oriented $\mathbb{Z}/2$–homology spheres, $\Theta^3_{\mathbb{Z}/2}$. The following lemma and the comment after it establish the precise relationship between the smooth concordance group and the $\mathbb{Z}/2$ homology cobordism group.

Lemma 4.1 [1, Lemma 2] *Let $K \subseteq S^3$ be a knot. Then:*

(a) *$\Sigma_2(K)$ is a $\mathbb{Z}/2$ homology 3–sphere, that is, $H_*(\Sigma_2(K);\mathbb{Z}/2) \cong H_*(S^3;\mathbb{Z}/2)$.*

(b) *If K is slice, then $\Sigma_2(K) = \partial Q$, where Q is a $\mathbb{Z}/2$–homology 4–ball, that is, $H_*(Q;\mathbb{Z}/2) \cong H_*(B^4;\mathbb{Z}/2)$.*

Moreover, $\Sigma_2(K_1 \# K_2) = \Sigma_2(K_1) \# \Sigma_2(K_2)$, where the separating sphere is obtained as the lift of the embedded 2–sphere in S^3 that appears in the definition of $K_1 \# K_2$ as the connected sum of pairs $(S^3, K_1) \# (S^3, K_2)$. All these observations show that the assignment $K \to \Sigma_2(K)$ is a group homomorphism

$$\Sigma_2\colon \mathcal{C}_\infty \to \Theta^3_{\mathbb{Z}/2}.$$

Therefore, the end result of the present article is in fact a result about independence in $\Theta^3_{\mathbb{Z}/2}$. With all the previous in mind, in this section we include a topological description of $\Sigma_2(D_n(K))$.

In [17; 14] the authors offer a description of the infinite cyclic cover of a satellite knot $B(K)$ in terms of some covers of the companion and pattern knots. Since finite cyclic covers may be regarded as quotients of the infinite cyclic covers, their description can be adapted to the case of 2–fold cyclic covers of satellite knots. The branched covers are obtained by compactifying the cyclic cover and attaching to it a solid torus in such a way that a meridian of the solid torus matches with the preimage of the meridian of the knot in the cyclic cover. In what follows we reproduce without proof the modified version of the description found in [17; 14].

Theorem 4.2 *Let $B \sqcup A$ be a pattern link satisfying* $\mathrm{lk}(A, B) = 0$ *and K a knot in S^3. There are splittings*

$$\Sigma_2(B) = V_2 \cup N_2 \quad \textit{and} \quad \Sigma_2(B(K)) = W_2 \cup M_2$$

such that:

(a) *The space N_2 consists of two copies of $N(A)$ and M_2 of two copies of $S^3 \setminus N(K)$.*

(b) *If N^i is the i^{th} copy of $N(A)$ in N_2 and X^i the i^{th} copy of $S^3 \setminus N(K)$ in M_2, then*

$$V_2 \cap N^i = T^i \quad \textit{and} \quad W_2 \cap X^i = U^i,$$

where T^i and U^i are 2–tori for $i = 1, 2$.

(c) *The embedding h from Definition 2.1 induces a homeomorphism $h_2\colon V_2 \to W_2$.*

(d) *If q_i and α_i are, respectively, the lift of the meridian and longitude of N to T^i, then the gluing map of $\Sigma_2(B)$ identifies $(\mu_A)_i$ with q_i, and $(\lambda_A)_i$ with α_i. Analogously, the gluing map of $\Sigma_2(B(K))$ identifies $(\lambda_K)_i$ with the image of q_i under h_2, and $(\mu_K)_i$ with the image of α_i under h_2.*

In conclusion, there is an isomorphism

$$\Sigma_2(B(K)) \cong V_2 \cup_\phi 2(S^3 \setminus N(K)),$$

where the gluing map ϕ identifies each copy of λ_K with q_i, and each corresponding copy of μ_K with the corresponding lift α_i. Thus, the 2–fold branched cover of S^3 over a satellite knot is determined by the 2–fold cover of a solid torus branched over the pattern B, and the curves α_i $(i = 1, 2)$. The following proposition makes these choices explicit for the patterns presented in Figure 2.

Proposition 4.3 *Given a knot $K \subseteq S^3$, the 2–fold branched cover $\Sigma_2(D_n(K))$ of S^3 branched over $D_n(K)$ has a decomposition*

$$\Sigma_2(D_n(K)) = S^3 \setminus N(T_{2,-2n}) \cup_\varphi 2(S^3 \setminus N(K)),$$

where $T_{2,-2n}$ is the $(2,-2n)$ torus link with unknotted components $A_1 \sqcup A_2$. Additionally, the gluing map φ is determined by

$$\varphi_*(\mu_K)_i = -n \cdot \mu_{A_i} + \lambda_{A_i} \quad \text{and} \quad \varphi_*(\lambda_K)_i = \mu_{A_i},$$

where μ_{A_i} and λ_{A_i} for $i = 1,\ 2$ denote the standard meridian–longitude pairs for the components of the link $T_{2,-2n} = A_1 \sqcup A_2$, and $(\mu_K)_i$ and $(\lambda_K)_i$ for $i = 1, 2$ denote the standard meridian–longitude pairs for K.

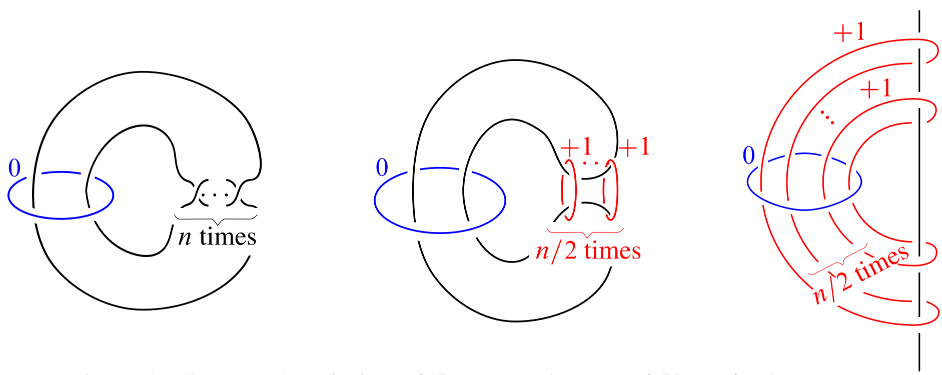


Figure 3: Surgery description of D_n as a subspace of V. Left: the pattern D_n and the pair $(A, 0)$. Center: surgery description of the pattern D_n as a subset of V. Right: an isotopy of the center diagram.

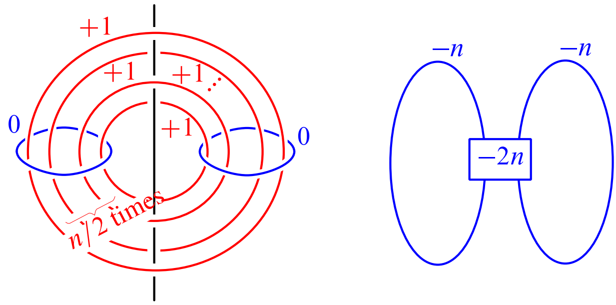


Figure 4: Left: surgery diagram of the 2–fold cover of V branched over D_n. Right: performing the surgeries one obtains the 2–fold cover of V branched over D_n and the lifts of $(A,0)$; the box represents half twists.

Proof By Theorem 4.2, to obtain a description of $\Sigma_2(D_n(K))$ it is enough to understand V_2, the 2–fold cover of $V = S^3 \setminus N(A)$ branched over D_n, and α_i for $i = 1, 2$, the lifts of λ_A to V_2. Since the longitude λ_A of A is the 0–framing of A, it suffices to consider the framed knot $(A,0)$ and its framed lifts (A_i, f_i) for $i = 1, 2$. Indeed, if $A_1 \sqcup A_2$ is the lift of A to $\Sigma_2(D_n)$, then

$$V_2 = \Sigma_2(D_n) \setminus N(A_1 \sqcup A_2) \cong S^3 \setminus N(A_1 \sqcup A_2),$$

and α_i is the $(f_i, 1)$ curve in $\partial N(A_i)$ for $i = 1, 2$. Therefore, the proof amounts to getting a description of the cover $\Sigma_2(D_n)$, which, since D_n is trivial in S^3, is simply S^3. This uses the surgery description of the pattern $D_n \sqcup A$, shown in Figure 3; Figure 4 shows the description of the 2–fold branched cover. □

5 Definite cobordisms

The main result will be obtained in terms of the instanton cobordism obstruction presented in Theorem 3.2 for a collection of Seifert fibered homology 3–spheres to cobound a negative definite 4–manifold. The issue here is that the 3–dimensional manifold $\Sigma_2(D_n(T_{p,q}))$ is not Seifert fibered. However, this obstacle can be overcome by introducing definite cobordisms with (unoriented) boundary $\Sigma_2(D_n(T_{p,q}))$ and some Seifert fibered spaces. The following theorem introduces the sought-after cobordisms.

Theorem 5.1 *Let (p,q) be relatively prime positive integers and $n > 0$ an even integer. If $\Sigma_2(D_n(T_{p,q}))$ is the 2–fold cover of S^3 branched over the satellite knot $D_n(T_{p,q})$, then there exist:*

(a) *A negative definite cobordism $Z(n,p,q)$ from the manifold $\Sigma_2(D_n(T_{p,q}))$ to the manifold $-\Sigma(p,q,npq-1)$.*

(b) *A negative definite cobordism $R(n,p,q)$ from the manifold $\Sigma_2(D_n(T_{p,q}))$ to the empty manifold.*

(c) *A positive definite cobordism $P(n,p,q)$ from the manifold $\Sigma_2(D_n(T_{p,q}))$ to the manifold $-\Sigma(p,q,2npq-1) \sqcup -\Sigma(p,q,2npq-1)$.*

Moreover, these cobordisms have trivial first homology group $H_1(-;\mathbb{Z})$.

By the handle decomposition theorem, every cobordism with incoming boundary component Y is obtained by attaching handles to $I \times Y$. Requiring the cobordism to be oriented is equivalent to requiring the attaching maps of the 1–handles to preserve orientations. In the case being considered, all the cobordisms will be obtained by attaching 2–handles to the 4–manifold $I \times \Sigma_2(D_n(T_{p,q}))$ along framed knots in $\{1\} \times \Sigma_2(D_n(T_{p,q}))$. To that end, we first recall the precise definition of framings to later compute the relevant ones.

Definition 5.2 Let J be a knot in a $\mathbb{Z}$–homology sphere Y and $N(J)$ a tubular neighborhood of J in Y. A framing of J is a choice of a simple closed curve J' in the boundary $N(J)$ that wraps once around J in the longitudinal direction. Similarly, the framing coefficient of J is the oriented intersection number of J' and any Seifert surface for J in Y.

With the definition of framing at hand, we start with the construction of the cobordism $Z(n,p,q)$.

Proof of Theorem 5.1(a) Any torus knot $T_{p,q}$ with (p,q) relatively prime positive integers admits a planar diagram with only positive crossings. This implies that $T_{p,q}$ can be unknotted by a sequence of positive-to-negative crossing changes in such a way that the i^{th} crossing change is obtained by performing -1 surgery on S^3 along a trivial knot γ_i that lies in the complement of $T_{p,q}$ and encloses the crossing. Then, if c is the number of crossings changed and $L = \gamma_i \sqcup \cdots \sqcup \gamma_c$, there exists an isomorphism

$$\psi\colon S^3_{-1}(L) \to S^3 \tag{2}$$

that identifies the restriction of $T_{p,q}$ to the complement of L with the unknot. Next, notice that since L is contained in $S^3 \setminus N(T_{p,q})$, it can be regarded as a subset of $\Sigma_n = \Sigma_2(D_n(T_{p,q}))$. Thus, one can form a 4–manifold Z by attaching 2–handles to $I \times \Sigma_n$ along the framed link $(L,-1)$. Specifically, if $\boldsymbol{h}_i$ is a 4–dimensional 2–handle,

$$Z = (I \times \Sigma_n) \cup_L (\boldsymbol{h}_1 \sqcup \cdots \sqcup \boldsymbol{h}_c).$$

It is then a matter of routine to check that the incoming boundary component of Z is the manifold Σ_n and its outcoming boundary component, Y, is the result of surgery on Σ_n along the framed link $(L,-1)$. In what follows, we will first obtain a description of Y as surgery and then we will show that Z is a negative definite manifold.

First, using the description of Σ_n included in Proposition 4.3, Y can be seen to split as the union of $(S^3 \setminus N(T_{p,q})) \cup_{\varphi_1} (S^3 \setminus N(T_{2,-2n}))$ and the result of surgery on $S^3 \setminus N(T_{p,q})$ along the framed link $(L,-1)$. The restriction of the isomorphism ψ from (2) to the latter space shows that surgery on $S^3 \setminus N(T_{p,q})$ along the framed link $(L,-1)$ is isomorphic to the unknot complement and therefore isomorphic to a standard solid torus $D^2 \times S^1$. Furthermore, choosing γ_i to have linking number 0 with the knot $T_{p,q}$ guarantees that the Seifert longitude of $T_{p,q}$ gets sent to the Seifert longitude of the unknot, and thus to a meridional curve $\partial D^2 \times \{\mathrm{pt}\}$ of $D^2 \times S^1$. The aforementioned choice also guarantees that the meridian of $T_{p,q}$ gets sent to the longitudinal curve $\{\mathrm{pt}\} \times S^1$ of the solid torus $D^2 \times S^1$. In other words, if h is the isomorphism between surgery on $S^3 \setminus N(T_{p,q})$ and the standard solid torus $D^2 \times S^1$, there is an isomorphism

$$Y \cong (S^3 \setminus N(T_{p,q})) \cup_{\varphi_1} (S^3 \setminus N(T_{2,-2n})) \cup_{\varphi_2 \circ h} D^2 \times S^1.$$

To simplify notation call A_1 and A_2 the components of the link $T_{2,-2n}$ and let

$$X = (S^3 \setminus N(T_{2,-2n})) \cup_{\varphi_2 \circ h} D^2 \times S^1 = (S^3 \setminus N(A_1 \sqcup A_2)) \cup_{\varphi_2 \circ h} D^2 \times S^1.$$

Notice that since the gluing map $\varphi_2 \circ h \colon \partial D^2 \times S^1 \to \partial N(A_1 \sqcup A_2)$ satisfies

$$(\varphi_2 \circ h)_*([S^1]) = (\varphi_2)_*(\mu_K) = -n\mu_{A_2} + \lambda_{A_2},$$
$$(\varphi_2 \circ h)_*([\partial D^2]) = (\varphi_2)_*(\lambda_K) = \mu_{A_2},$$

it extends to the interior of $D^2 \times S^1$. This implies that X is the result of filling the space left by $N(A_2)$ in S^3 with a solid torus in a way that makes X isomorphic to $S^3 \setminus N(A_1)$. Then, since A_1 is unknotted, X is isomorphic to a standard solid torus and thus Y is isomorphic to the union of $S^3 \setminus N(T_{p,q})$ and a solid torus. In other words, Y is the result of performing surgery on S^3 along $T_{p,q}$. To make explicit the coefficient of the surgery, recall that

$$(\varphi_1)_*(\mu_K) = -n\mu_{A_1} + \lambda_{A_1} \quad \text{and} \quad (\varphi_1)_*(\lambda_K) = \mu_{A_1}.$$

Then, since λ_{A_1} is identified with the meridian ∂D^2 and μ_{A_1} with the longitude S^1, simple arithmetic shows that

$$(\varphi_1)_*(\mu_K + n\lambda_K) = [\partial D^2],$$

thus showing that the surgery coefficient is $1/n$. Finally, since, for $p, q, n > 0$, the result of $1/n$ surgery on S^3 along the torus knot $T_{p,q}$ is diffeomorphic to the Seifert fibred homology sphere $-\Sigma(p,q,npq-1)$ [15, Proposition 3.1], the outcoming boundary component of Z is $-\Sigma(p,q,npq-1)$, as sought.

As for definiteness, since Σ_n is a homology sphere, the second homology group $H_2(Z;\mathbb{Z})$ admits a basis determined by the 2–handles. In addition, the matrix representation of the intersection form of Z in terms of such basis is given by the linking matrix of the framed link $(L,-1)$. This, in turn, can be seen to be the matrix $-I_c$, where I_c is the $c\times c$ identity matrix. We thus see that Z is negative definite, as sought. $\square$

The remaining statements Theorem 5.1(b)–(c) will be obtained as a corollary to the following theorem.

Theorem 5.3 *Let K be any knot and Σ_n the 2–fold cover of S^3 branched over $D_n(K)$. Then there exist 4–manifolds $P_n(K)$ and $R_n(K)$ such that:*

(a) *$P_n(K)$ is a positive definite cobordism from Σ_n to $S^3_{1/2n}(K)\,\#\,S^3_{1/2n}(K)$.*

(b) *$R_n(K)$ is a negative definite cobordism from Σ_n to S^3.*

The cobordisms will be constructed explicitly from $I\times\Sigma_n$ by attaching some 2–handles to it along framed knots in Σ_n. Specifically, the attachment will take place along the links $\boldsymbol{\gamma}^\pm=\gamma_1^\pm\sqcup\cdots\sqcup\gamma_n^\pm$ shown in Figure 5 and will be completely determined after establishing the appropriate framing and framing coefficient of the link components. Notice that since $\boldsymbol{\gamma}^\pm$ is completely contained in $S^3\setminus N(T_{2,-2n})$, any tubular neighborhood $N(\gamma_i^\pm)$ in S^3 small enough to be completely contained in $S^3\setminus N(T_{2,-2n})$ is also a tubular neighborhood of $\gamma_i^\pm$ in Σ_n. Definition 5.2 and the previous statement show that there is no difference between framings of $\boldsymbol{\gamma}^\pm$ in S^3 and Σ_n. To see that the same holds for framing coefficients we need to analyze the Seifert surfaces for $\gamma_i^\pm$ in both S^3 and Σ_n. First, since $\gamma_i^\pm$ is an unknot in S^3, any embedded 2–disk in S^3 bounding $\gamma_i^\pm$ is a Seifert surface for $\gamma_i^\pm$ in S^3. Call this disk D_i and choose it to be disjoint from every other component of $\boldsymbol{\gamma}^\pm$. Notice also that each curve $\gamma_i^\pm$ encloses a crossing of $T_{2,-2n}$ in such a way that D_i intersects the boundary of $N(T_{2,-2n})$ in two disjoint curves, one homologous to $-\mu_{A_1}$ and the other to $-\mu_{A_2}$ (see Figure 6). Next, to obtain a Seifert surface S_i for $\gamma_i^\pm$ in Σ_n, let F_j be a Seifert surface for K in S^3 contained in the j^{th} copy of $S^3\setminus N(K)$ in Σ_n

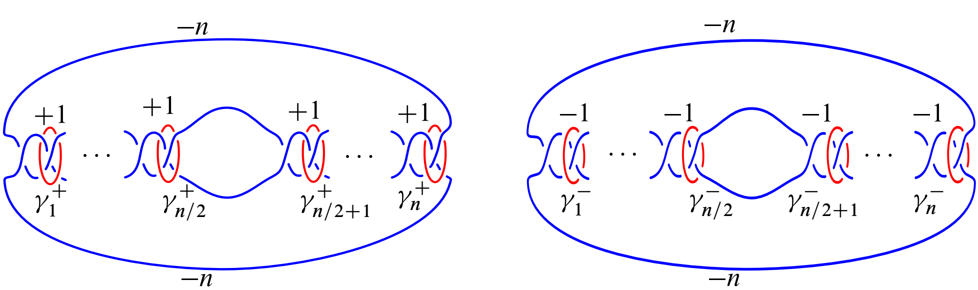


Figure 5: Descriptions of $P_n(K)$ (left) and $R_n(K)$ (right), showing Σ_n and the links $\boldsymbol{\gamma}^\pm$.

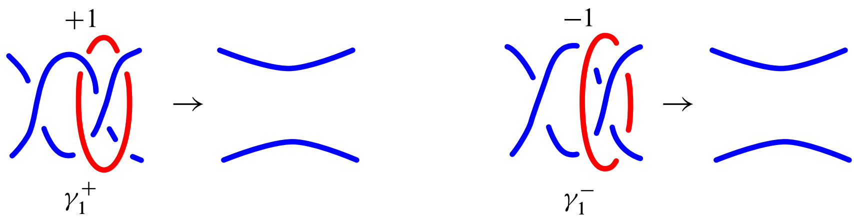


Figure 6: Local depiction of one component of γ^+ (left) and γ^- (right) and the corresponding crossing changes.

and recall that the gluing map φ from Proposition 4.3 identifies μ_{A_j} with a longitude of K in the j^{th} copy of $S^3 \setminus N(K) \subseteq \Sigma$. The surface F_j can then be glued to D_i along φ_j and so we can form

$$S_i = D_i \cap S^3 \setminus N(T_{2,-2n}) \cup_\varphi (F_1 \sqcup F_2). \tag{3}$$

Hence, if β_i is any framing of γ_i, its framing coefficient in S^3 is given by the number of points in $\beta_i \cap D_i$ counted with sign and its framing coefficient in Σ_n is given by the number of points in $\beta_i \cap S_i$ counted with sign. Since any choice of β_i is contained in the interior of $S^3 \setminus N(T_{2,-2n})$, it is disjoint from each copy of $S^3 \setminus N(K)$ that appears in the description of Σ_n. Thus, β_i is disjoint from both F_1 and F_2 and so

$$\beta_i \cap S_i = \beta_i \cap (D_i \cap S^3 \setminus N(T_{2,-2n})) = \beta_i \cap D_i.$$

This shows that the framing coefficient of $\gamma_i^\pm$ in both S^3 and Σ_n agree.

So, let $\boldsymbol{\gamma}^\pm = \gamma_1^\pm \sqcup \cdots \sqcup \gamma_n^\pm$ with the framings as shown in Figure 5, and form

$$P_n(K) = (I \times \Sigma_n) \cup_{\boldsymbol{\gamma}^+} (\boldsymbol{h}_1 \sqcup \cdots \sqcup \boldsymbol{h}_n) \quad \text{and} \quad R_n(K) = (I \times \Sigma_n) \cup_{\boldsymbol{\gamma}^-} (\boldsymbol{h}_1 \sqcup \cdots \sqcup \boldsymbol{h}_n).$$

These two 4–manifolds are the sought-after cobordisms, as will be established next.

Proof of Theorem 5.3 The boundary of $P_n(K)$ is the disjoint union of $-\Sigma_n$ and M^+, the result of surgery on Σ_n along the framed link $\boldsymbol{\gamma}^+$. Analogously, the boundary of $R_n(K)$ is the disjoint union of $-\Sigma_n$ and M^-, the result of surgery on Σ_n along the framed link $\boldsymbol{\gamma}^-$. Then, since $\boldsymbol{\gamma}^\pm$ is a link in $S^3 \setminus N(T_{2,-2n})$, the space $M^\pm$ can be expressed as the union of two disjoint copies of $S^3 \setminus N(K)$ and surgery on $S^3 \setminus N(T_{2,-2n})$ along the framed link $\boldsymbol{\gamma}^\pm$. The latter manifold can be better understood by first performing the surgery on S^3 and then examining the effect such surgery has on $S^3 \setminus N(T_{2,-2n})$.

Since the surgery is done along unknots with framing ± 1, the result is a space isomorphic to S^3. Also, notice that every component of $\boldsymbol{\gamma}^\pm$ encloses a crossing of the link $T_{2,-2n}$. Then, it is well-known that surgery on S^3 along $\boldsymbol{\gamma}^\pm$ can be interpreted as a sequence of n crossing changes on the link $T_{2,-2n}$ that unlink its components. In other words, there is an isomorphism

$$\psi^{\pm}\colon S^3_{\pm 1}(\boldsymbol{\gamma}^{\pm}) \to S^3$$

that sends the restriction of $T_{2,-2n}$ to the complement of $\boldsymbol{\gamma}^{\pm}$ to the 2–component unlink $\boldsymbol{U} = U_1 \sqcup U_2$. Thus, after restricting, $\psi^{\pm}$ gives us an isomorphism between surgery on $S^3 \setminus N(T_{2,-2n})$ along $\boldsymbol{\gamma}^{\pm}$ and $S^3 \setminus N(\boldsymbol{U})$. The previous shows that

$$M^{\pm} \cong (S^3 \setminus N(\boldsymbol{U})) \cup_{\psi \circ \varphi} 2(S^3 \setminus N(K)).$$

Furthermore, since $\boldsymbol{U}$ is a 2–component unlink, there exists a 2–sphere S^2 that separates $S^3 \setminus N(\boldsymbol{U})$ into $S^3 \setminus N(U_1) \# S^3 \setminus N(U_1) \cong D^2 \times S^1 \# D^2 \times S^1$. Then, the same sphere decomposes $M^{\pm}$ as

$$M^{\pm} \cong \big((D^2 \times S^1) \cup_{h_1^{\pm}} (S^3 \setminus N(K))\big) \# \big((D^2 \times S^1) \cup_{h_2^{\pm}} (S^3 \setminus N(K))\big). \tag{4}$$

For simplicity in notation, set $X^{\pm} = (D^2 \times S^1) \cup_{h^{\pm}} (S^3 \setminus N(K))$ and notice that, being the union of the complement of K and a solid torus, $X^{\pm}$ is surgery on S^3 along K. The coefficient of the surgery is given by the homology class of the curve that maps to the meridian $\partial D^2 \times \{\mathrm{pt}\}$ of $D^2 \times S^1$ under the gluing map $h^{\pm}$, and so it is important to understand $h^{\pm}$. This can be done by analyzing the identifications that took place to get (4), and the effect they have on μ_k and λ_K. With that in mind, let $\{\mu_{A_i}, \lambda_{A_i}\}$ be the meridian–longitude pair of the component A_i of $T_{2,-2n}$, and let $\{\mu_{U_i}, \lambda_{U_i}\}$ be the meridian–longitude pair of the component U_i of $\boldsymbol{U}$. Also, recall that φ is such that

$$(\varphi)_*(\mu_K) = -n \cdot \mu_{A_i} + \lambda_{A_i} \quad \text{and} \quad (\varphi)_*(\lambda_K) = \mu_{A_i},$$

and that, since $\mathrm{lk}(\gamma_j^{\pm}, A_i) = 1$ and $\psi^{\pm}$ can be interpreted as a sequence of n crossing changes,

$$\psi_*^{\pm}(\mu_{A_i}) = \mu_{U_i} \quad \text{and} \quad \psi_*^{\pm}(\lambda_{A_i}) = (\mp n) \cdot \mu_{U_i} + \lambda_{U_i}.$$

Similarly, the isomorphism θ between $S^3 \setminus N(U_i)$ and the standard solid torus $D^2 \times S^1$ identifies μ_{U_i} with $l = [S^1]$ and λ_{U_i} with $m = [\partial D^2]$, so that

$$(h_i^{\pm})_*(\mu_K) = m + (-n \mp n) \cdot l \quad \text{and} \quad (h_i^{\pm})_*(\lambda_K) = l.$$

Therefore $(h_i^{\pm})_*(\mu_K + (n \pm n) \cdot \lambda_K) = m$, showing that the slope of the surgery is $1/(n \pm n)$. This shows that

$$M^{+} \cong S^3_{1/2n}(K) \# S^3_{1/2n}(K) \quad \text{and} \quad M^{-} \cong S^3_{1/0}(K) \# S^3_{1/0}(K) \cong S^3,$$

thus proving that $P_n(K)$ is a cobordism from Σ_n to $S^3_{1/2n}(K) \# S^3_{1/2n}(K)$, and $R_n(K)$ is one from Σ_n to S^3.

To show definiteness, it is enough to understand the intersection form of the 4–manifolds being considered. Let $\{b_1, b_2, \dots, b_n\}$ be the basis for $H_2(-;\mathbb{Z})$ determined by the handles. To find a surface that represents b_j consider S_j the Seifert surface for γ_j^+ in Σ_n described in (3) and push $\operatorname{int}(S_j)$ into the interior of $I \times \Sigma \subset P_n(K)$. Then add the core of the i^{th} handle along γ_j^+ to obtain a closed surface $\widehat{S}_j$. Next, denote by Q the intersection form of $P_n(K)$. It is well-known that the value of $Q(b_j, b_k)$ is given by the number of points in $\widehat{S}_j \cap \widehat{S}_k$, counted with sign. Then, using (3) we get

$$\widehat{S}_j \cap \widehat{S}_k = S_j \cap \gamma_k^+ = D_j \cap \gamma_k^+.$$

Here D_j is a 2–disk in S^3 bounding γ_j and disjoint from every other component of $\boldsymbol{\gamma}^+$. Since the disk D_j is disjoint from every other component of $\boldsymbol{\gamma}^+$, and γ_j^+ has framing $+1$, the signed number of points in $D_j \cap \gamma_k^+$ is given by the Kronecker delta number δ_{ik}. This shows that the $n \times n$ identity matrix I_n represents the intersection form Q in terms of the basis $\{b_1, b_2, \dots, b_n\}$, and thus that $P_n(K)$ is a positive definite manifold.

The analogous argument applied to $\boldsymbol{\gamma}^-$ shows that $-I_n$ represents the intersection form of $R_n(K)$ and so $R_n(K)$ is negative definite. □

The following corollary establishes Theorem 5.1(b)–(c):

Corollary 5.4 *Let $p, q > 0$ and consider the satellite knot $D_n(T_{p,q})$. If $\Sigma = \Sigma_2(D_n(T_{p,q}))$ is the 2–fold cover of S^3 branched over $D_n(T_{p,q})$ then:*

(a) *There exists a positive definite 4–manifold, $P(n, p, q)$, with boundary components $-\Sigma$ and two copies of $-\Sigma(p, q, 2npq - 1)$.*

(b) *There exists a negative definite 4–manifold, $R(n, p, q)$, with boundary $-\Sigma$.*

Proof First, to construct $P(n, p, q)$ attach a 3–handle to the manifold $P_n(T_{p,q})$ along its outcoming boundary component to transform the connected sum of manifolds into disjoint union. Next, recall that for $p, q, n > 0$, the result of $1/2n$ surgery on S^3 along the torus knot $T_{p,q}$ is diffeomorphic to the Seifert fibred homology sphere $-\Sigma(p, q, 2npq - 1)$ [15, Proposition 3.1].

Similarly, the manifold $R(n, p, q)$ is obtained from $R_n(T_{p,q})$ by capping off its outcoming boundary component S^3 with a 4–ball. □

6 Main result

Theorem 6.1 *Let $\{(p_i, q_i)\}_i$ be a sequence of relatively prime positive integers and $\{n_i\}_i$ a sequence of positive and even integers. If*

$$p_i q_i (2 n_i p_i q_i - 1) < p_{i+1} q_{i+1} (n_{i+1} p_{i+1} q_{i+1} - 1),$$

the family $\mathscr{F} = \{\Sigma_2(D_{n_i}(T_{p_i q_i}))\}_{i=1}^{\infty}$ is independent in $\Theta^3_{\mathbb{Z}/2}$.

Proof Denote by $[Y]$ the homology cobordism class of the $\mathbb{Z}/2$–homology sphere Y and suppose by contradiction that there exist integral coefficients $c_1,\dots,c_N \in \mathbb{Z}$ such that

$$\sum_{i=1}^{N} c_i[\Sigma_2(D_{n_i}(T_{p_i q_i}))] = 0$$

in $\Theta^3_{\mathbb{Z}/2}$. The supposition implies the existence of an oriented 4–manifold Q with the $\mathbb{Z}/2$ homology of a punctured 4–ball and with boundary

$$\partial Q = \mathop{\#}_{i=1}^{N}\Bigl(\mathop{\#}_{j=1}^{c_i} \Sigma_2(D_{n_i}(T_{p_i q_i}))\Bigr).$$

Attaching 3–handles to Q we can further assume that

$$\partial Q = \bigsqcup_{i=1}^{N} c_i \Sigma_2(D_{n_i}(T_{p_i q_i})).$$

Here we use cY to denote the disjoint union of c copies of Y if $c > 0$, and $-c$ copies of $-Y$ if $c < 0$. In addition, and without loss of generality, further assume that $c_N \geq 1$. Augment Q using the cobordisms constructed in Theorem 5.1, namely (see Figure 7), let

$$X = Q \cup Z(n_N, p_N, q_N) \cup \Bigl(\bigsqcup_{c_i>0} R(n_i, p_i, q_i)\Bigr) \cup \Bigl(\bigsqcup_{c_i<0} -P(n_i, p_i, q_i)\Bigr).$$

Recall by Theorem 5.1 that $Z(n,p,q)$, $-P(n,p,q)$ and $R(n,p,q)$ are negative definite cobordisms from Σ to $-\Sigma(p,q,npq-1)$, $2\Sigma(p,q,2npq-1)$ and the empty set, respectively. Thus, X is a negative definite 4–manifold with oriented boundary

$$\partial X = -\Sigma(p_N, q_N, n_N p_N q_N - 1) \sqcup \Bigl(\bigsqcup_{c_i<0} 2\Sigma(p_i, q_i, 2n_i p_i q_i - 1)\Bigr).$$

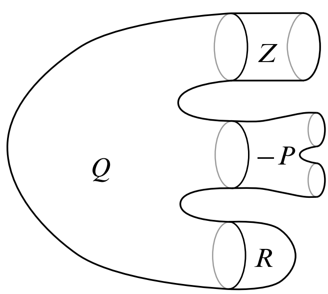


Figure 7: The manifold X

Also, since the first $\mathbb{Z}/2$–homology groups of $Z(n,p,q)$, $-P(n,p,q)$, $R(n,p,q)$ and Q are trivial, the Mayer–Vietoris theorem shows that $H_1(X,\mathbb{Z}/2)=0$. This would imply that the Seifert fibered spaces

$$\{-\Sigma(p_N,q_N,n_Np_Nq_N-1)\}\cup\{\Sigma(p_i,q_i,2n_ip_iq_i-1)\}_{c_i<0}$$

cobound a smooth 4–manifold that has negative definite intersection form and that satisfies $H_1(X,\mathbb{Z}/2)=0$, contradicting Theorem 3.2. Therefore, Q cannot exist and so the 3–manifolds $\Sigma_2(D_{n_i}(T_{p_iq_i}))$ are independent in the $\mathbb{Z}/2$ homology cobordism group. □

Theorem 6.2 *Let $\{(p_i,q_i)\}_i$ be a sequence of relatively prime positive integers and $\{n_i\}_i$ a sequence of positive and even integers. Then, if*

$$p_iq_i(2n_ip_iq_i-1)<p_{i+1}q_{i+1}(n_{i+1}p_{i+1}q_{i+1}-1),$$

the collection $\{D_{n_i}(T_{p_iq_i})\}_{i=1}^{\infty}$ is an independent family in $\mathcal{C}_\infty$.

Proof If $c_1D_{n_1}(T_{p_1q_1})\#c_2D_{n_2}(T_{p_2q_2})\#\cdots\#c_ND_{n_N}(T_{p_Nq_N})$ is slice for some integral coefficients $c_1,\dots,c_N\in\mathbb{Z}$, then Lemma 4.1 shows that

$$\Sigma_2(c_1D_{n_1}(T_{p_1q_1})\#\cdots\#c_ND_{n_N}(T_{p_Nq_N}))
= c_1\Sigma_2(D_{n_1}(T_{p_1q_1}))\#\cdots\#c_N\Sigma_2(D_{n_N}(T_{p_Nq_N}))$$

is the boundary of a $\mathbb{Z}/2$–homology ball Q. However, Theorem 6.1 shows that Q does not exist and the result thus follows. □

References

[1] **A J Casson**, **C M Gordon**, *Cobordism of classical knots*, from "À la recherche de la topologie perdue" (L Guillou, A Marin, editors), Progr. Math. 62, Birkhäuser, Boston (1986) 181–199 MR

[2] **T D Cochran**, **R E Gompf**, *Applications of Donaldson's theorems to classical knot concordance, homology 3–spheres and property P*, Topology 27 (1988) 495–512 MR

[3] **S K Donaldson**, *Floer homology groups in Yang–Mills theory*, Cambridge Tracts in Math. 147, Cambridge Univ. Press (2002) MR

[4] **R Fintushel**, **R J Stern**, *Pseudofree orbifolds*, Ann. of Math. 122 (1985) 335–364 MR

[5] **R Fintushel**, **R J Stern**, *Instanton homology of Seifert fibred homology three spheres*, Proc. London Math. Soc. 61 (1990) 109–137 MR

[6] **M H Freedman**, *The topology of four-dimensional manifolds*, J. Differential Geom. 17 (1982) 357–453 MR

[7] **M H Freedman**, *The disk theorem for four-dimensional manifolds*, from "Proceedings of the International Congress of Mathematicians, I" (Z Ciesielski, C Olech, editors), PWN, Warsaw (1984) 647–663 MR

[8] **M H Freedman**, **F Quinn**, *Topology of 4–manifolds*, Princeton Math. Series 39, Princeton Univ. Press (1990) MR

[9] **M Furuta**, *Homology cobordism group of homology 3–spheres*, Invent. Math. 100 (1990) 339–355 MR

[10] **R E Gompf**, **A I Stipsicz**, *4–manifolds and Kirby calculus*, Graduate Studies in Math. 20, Amer. Math. Soc., Providence, RI (1999) MR

[11] **M Hedden**, **P Kirk**, *Chern–Simons invariants,* SO(3) *instantons, and* $\mathbb{Z}/2$ *homology cobordism*, from "Chern–Simons gauge theory: 20 years after" (J r E Andersen, H U Boden, A Hahn, B Himpel, editors), AMS/IP Stud. Adv. Math. 50, Amer. Math. Soc., Providence, RI (2011) 83–114 MR

[12] **M Hedden**, **P Kirk**, *Instantons, concordance, and Whitehead doubling*, J. Differential Geom. 91 (2012) 281–319 MR

[13] **W B R Lickorish**, *An introduction to knot theory*, Graduate Texts in Math. 175, Springer (1997) MR

[14] **C Livingston**, **P Melvin**, *Abelian invariants of satellite knots*, from "Geometry and topology" (J Alexander, J Harer, editors), Lecture Notes in Math. 1167, Springer (1985) 217–227 MR

[15] **L Moser**, *Elementary surgery along a torus knot*, Pacific J. Math. 38 (1971) 737–745 MR

[16] **W D Neumann**, **D Zagier**, *A note on an invariant of Fintushel and Stern*, from "Geometry and topology" (J Alexander, J Harer, editors), Lecture Notes in Math. 1167, Springer (1985) 241–244 MR

[17] **H Seifert**, *On the homology invariants of knots*, Quart. J. Math., Oxford Ser. 1 (1950) 23–32 MR

[18] **C H Taubes**, *Gauge theory on asymptotically periodic 4–manifolds*, J. Differential Geom. 25 (1987) 363–430 MR

[19] **K K Uhlenbeck**, *Removable singularities in Yang–Mills fields*, Comm. Math. Phys. 83 (1982) 11–29 MR

Department of Mathematics, University of Georgia
Athens, GA 30605, United States

jpinzon@uga.edu

http://alpha.math.uga.edu/~juanita/

Proposed: Peter Teichner — Received: 16 January 2015
Seconded: András I. Stipsicz, Rob Kirby — Revised: 13 October 2016

Geometry & Topology 21 (2017) 3213–3230

The chromatic splitting conjecture at $n = p = 2$

AGNÈS BEAUDRY

We show that the strongest form of Hopkins' chromatic splitting conjecture, as stated by Hovey, cannot hold at chromatic level $n=2$ at the prime $p=2$. More precisely, for $V(0)$, the mod 2 Moore spectrum, we prove that $\pi_k L_1 L_{K(2)} V(0)$ is not zero when k is congruent to -3 modulo 8. We explain how this contradicts the decomposition of $L_1 L_{K(2)} S$ predicted by the chromatic splitting conjecture.

55P60, 55Q45

1 Introduction

Fix a prime p. Let S be the p–local sphere spectrum, and $L_n S$ be the Bousfield localization of S at the Johnson–Wilson spectrum $E(n)$. Let $K(n)$ be Morava K-theory. There is a homotopy pullback square called the chromatic fracture square:

$$\begin{tikzcd} L_n S \arrow[r] \arrow[d] & L_{K(n)} S \arrow[d] \\ L_{n-1} S \arrow[r, "\iota"] & L_{n-1} L_{K(n)} S \end{tikzcd}$$

Let F_n be the fiber of the map $L_n S \to L_{K(n)} S$. Note that F_n is weakly equivalent to the fiber of ι. It was shown by Hovey [12, Lemma 4.1] that F_n is weakly equivalent to the function spectrum $F(L_{n-1}S, L_n S)$. Hopkins' chromatic splitting conjecture, as stated by Hovey [12, Conjecture 4.2], stipulates that ι is the inclusion of a wedge summand, so that

$$L_{n-1} L_{K(n)} S \simeq L_{n-1} S \vee \Sigma F_n. \tag{1-1}$$

We will call this the *weak* form of the chromatic splitting conjecture. However, [12, Conjecture 4.2] also gives an explicit decomposition of ΣF_n as a wedge of suspensions of spectra of the form $L_i S_p$ for $0 \le i < n$. We will call this the *strong* form of the chromatic splitting conjecture.

The conjectured decomposition comes from the connection between the $K(n)$–local category and the cohomology of a certain group called the Morava stabilizer group $\mathbb{G}_n$.

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3213

Let $\mathbb{S}_n$ be the group of automorphisms of the formal group law of $K(n)$ over $\mathbb{F}_{p^n}$. Then $\mathbb{G}_n$ is the extension of $\mathbb{S}_n$ by the Galois group $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$. Let $\mathbb{W}$ be the Witt vectors on $\mathbb{F}_{p^n}$. There is a spectral sequence

$$H^s(\mathbb{G}_n, (E_n)_t) \implies \pi_{t-s} L_{K(n)} S. \tag{1-2}$$

Note that $\mathbb{W}$ sits naturally in $(E_n)_0 \cong \mathbb{W}[\![u_1, \dots, u_{n-1}]\!]$. The inclusion induces a map

$$H^*(\mathbb{G}_n, \mathbb{W}) \to H^*(\mathbb{G}_n, (E_n)_0). \tag{1-3}$$

Morava proves in [16, Remark 2.2.5], using the work of Lazard, that

$$H^*(\mathbb{G}_n, \mathbb{W}) \otimes \mathbb{Q}_p \cong E(e_1, \dots, e_n)$$

for classes e_i of degree $2i-1$. Therefore, $H^*(\mathbb{G}_n, \mathbb{W})$ contains an exterior algebra $E(x_1, \dots, x_n)$ for appropriate integral multiples x_i of the generators e_i. The chromatic splitting conjecture stipulates that, for some choice of $x_1, \dots, x_n$, the exterior algebra $E(x_1, \dots, x_n)$ injects into $H^*(\mathbb{G}_n, (E_n)_0)$ under the map (1-3), and that the nonzero products $x_{i_1} \cdots x_{i_j}$ survive in (1-2) to nontrivial elements in $\pi_{-2(\sum i_k)+j} L_{K(n)} S$. Further, it states that there is a factorization

$$\begin{tikzcd}
S_p^{-2(\sum i_k)+j} \arrow[r] \arrow[d] & L_{n-\max(i_k)} S_p^{-2(\sum i_k)+j} \arrow[d, dotted] \\
L_{K(n)} S \arrow[r] & \Sigma F_n
\end{tikzcd}$$

where S_p^m is the p–completion of S^m, and that these maps decompose ΣF_n as

$$\Sigma F_n \simeq \bigvee_{\substack{1 \le j \le n \\ 1 \le i_1 < \cdots < i_j \le n}} L_{n-\max(i_k)} S_p^{-2(\sum i_k)+j}. \tag{1-4}$$

The chromatic splitting conjecture has been shown for $n \le 2$ and for all primes p, except in the case $n = p = 2$. For $n = 1$, it follows immediately from a computation of $\pi_* L_1 S_p$; see Ravenel [19, Theorems 8.10 and 8.15]. At $n = 2$ and $p \ge 5$, it is due to Hopkins, and follows from Shimomura and Yabe's computations [23]. The proof can be found in Behrens' account of their work [4, Remark 7.8]. At $n = 2$ and $p = 3$, the conjecture was proved recently by Goerss, Henn and Mahowald [9].

In this paper, we show that the chromatic splitting conjecture as stated above cannot hold for $n = p = 2$. More precisely, we show that [12, Conjecture 4.2(iv)] fails in this case. At $n = 2$, (1-1) and (1-4) imply that

$$L_1 L_{K(2)} S \simeq L_1 S_p \vee L_1 S_p^{-1} \vee L_0 S_p^{-3} \vee L_0 S_p^{-4}. \tag{1-5}$$

We show that the right-hand side of (1-5) has too few homotopy groups for the equivalence to hold. However, our results do not contradict the possibility that ι is the inclusion of a wedge summand. Giving an alternative description for the fiber in this case is work in progress.

That our methods could disprove (1-5) was first suggested to the author by Paul Goerss. He and Mark Mahowald had been studying the computations of Shimomura and Wang [22] and Shimomura [21] and noticed that these suggest that the right-hand side of (1-5) is too small.

Statement of the results Let $V(0)$ be the cofiber of multiplication by p on S. Note that for any p–local spectrum X, there is a cofiber sequence

$$X \xrightarrow{p} X \to X \wedge V(0).$$

Since Bousfield localization of spectra preserves exact triangles, it follows that

$$L_E V(0) \simeq L_E S \wedge V(0)$$

for any spectrum E. This has the following consequence.

Proposition 1.1 *The strong form of the chromatic splitting conjecture at $n = 2$ implies that $L_1 L_{K(2)} V(0) \simeq L_1 V(0) \vee L_1 \Sigma^{-1} V(0)$.*

We now fix our attention to the case when $p = 2$. Since $L_0 V(0)$ is contractible, it follows from the chromatic fracture square that $L_1 V(0) \simeq L_{K(1)} V(0)$. Computing $\pi_* L_{K(1)} V(0)$ is a routine exercise using the spectral sequence

$$E_2^{s,t} = H^s(\mathbb{G}_1, (E_1)_* V(0)) \implies \pi_{t-s} L_{K(1)} V(0). \tag{1-6}$$

The E_∞–term is given in Figure 1. At $p = 2$, we have that $V(0)$ is not a ring spectrum. This manifests itself by the fact that $\pi_* L_{K(1)} V(0)$ is not a ring. In fact,

$$\pi_* L_{K(1)} V(0) = \big(\mathbb{Z}_2[\eta, \beta^{\pm 1}, \zeta_1]/(2\eta, \eta^3, \zeta_1^2)\big)\{e_0, v_1 e_0\}/(2e_0, 2v_1 e_0 - \eta^2 e_0),$$

where $\eta \in \pi_1$ is the Hopf map, $\beta \in \pi_8$ is the v_1–self-map detected by v_1^4, and $\zeta_1 \in \pi_{-1}$ is detected by a generator of $H^1(\mathbb{G}_1, \mathbb{Z}_2) \cong H^1(\mathbb{Z}_2^\times, \mathbb{Z}_2)$. The element $e_0 \in \pi_0$ represents the inclusion of the bottom cell $S^0 \hookrightarrow V(0)$, and $v_1 e_0 \in \pi_2$ is a lift of $\Sigma\eta$ to the top cell:

$$\begin{tikzcd}
 & & S^2 \arrow[dl, dotted, "v_1 e_0"'] \arrow[d, "\Sigma\eta"] & \\
S^0 \arrow[r, "e_0"] & V(0) \arrow[r] & S^1 \arrow[r, "2"] & S^1
\end{tikzcd}$$

The following result is a consequence of Proposition 1.1.

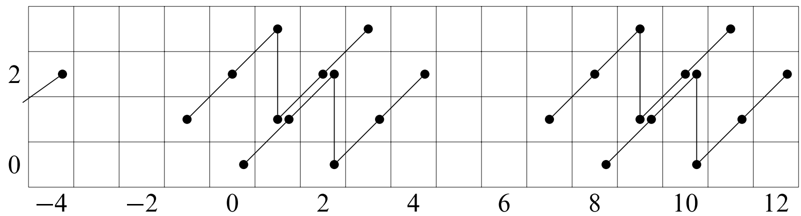


Figure 1: The E_∞–term of (1-6) computing $\pi_* L_{K(1)} V(0)$. Vertical lines denote extensions by multiplication by 2, and lines of slope one denote multiplication by η.

Corollary 1.2 *The chromatic splitting conjecture implies that $\pi_k L_1 L_{K(2)} V(0)$ is zero when $k \equiv -3$ modulo 8.*

However, in this paper, we prove the following result.

Theorem 1.3 *There are nontrivial homotopy classes $\beta^t x$ in $\pi_{8t-3} L_1 L_{K(2)} V(0)$ and $\zeta_2 \beta^t x$ in $\pi_{8t-4} L_1 L_{K(2)} V(0)$.*

This has the following immediate consequence.

Theorem 1.4 *The homotopy group $\pi_k L_1 L_{K(2)} V(0)$ is nonzero when $k \equiv -3$ modulo 8. Therefore, the decomposition (1-5) of the chromatic splitting conjecture does not hold when $n = 2$ and $p = 2$.*

The broad strokes of the proof of Theorem 1.3 when $t = 0$ are as follows. Let $G_{24} \cong Q_8 \rtimes C_3$ be a representative of the unique conjugacy class of maximal finite subgroups of $\mathbb{S}_2$. Let C_6 be a subgroup of G_{24} of order 6. Let $\mathbb{S}_2^1$ be the norm one subgroup so that $\mathbb{S}_2 \cong \mathbb{S}_2^1 \rtimes \mathbb{Z}_2$ (see Section 2). It follows from the duality resolution techniques of Goerss, Henn, Mahowald and Rezk and the work of Bobkova [6] that, for any X, there is a spectral sequence

$$E_1^{p,t} = \pi_t(\mathscr{E}_p \wedge X) \implies \pi_{t-p}(E_2^{h\mathbb{S}_2^1} \wedge X),$$

where $\mathscr{E}_p$ are spectra such that $\mathscr{E}_0 \simeq E_2^{hG_{24}}$, $\mathscr{E}_p \simeq E_2^{hC_6}$ if $p = 1, 2$ and $(E_2)_* \mathscr{E}_3 \cong (E_2)_* E_2^{hG_{24}}$ as Morava modules. Localizing at $E(1)$, we obtain a spectral sequence

$$E_1^{p,t} = \pi_t L_1(\mathscr{E}_p \wedge X) \implies \pi_{t-p} L_1(E_2^{h\mathbb{S}_2^1} \wedge X). \tag{1-7}$$

We use this spectral sequence to show that $\pi_{-3} L_1(E_2^{h\mathbb{S}_2^1} \wedge V(0)) \cong \mathbb{F}_4$, in Lemma 4.1 and Proposition 4.2. After taking Galois invariants, we obtain a nonzero element x in $\pi_{-3} L_1(E_2^{h\mathbb{G}_2^1} \wedge V(0))$. In the cofiber sequence

$$L_1 L_{K(2)} V(0) \to L_1(E_2^{h\mathbb{G}_2^1} \wedge V(0)) \to L_1(E_2^{h\mathbb{G}_2^1} \wedge V(0)),$$

which is obtained from the cofiber sequence $L_{K(2)}S \to E_2^{h\mathbb{G}_2^1} \to E_2^{h\mathbb{G}_2^1}$ by smashing with $V(0)$ and localizing at $E(1)$; this class gives rise to nonzero elements $x \in \pi_{-3}L_1L_{K(2)}V(0)$ and $\zeta_2 x \in \pi_{-4}L_1L_{K(2)}V(0)$.

Warning 1.5 We use the notation ζ_2 to denote the homotopy class defined by

$$\begin{tikzcd} & & S^0 \arrow[d, "1"'] \arrow[dr, "\zeta_2"] & \\ L_{K(2)}S \arrow[r] & E_2^{h\mathbb{G}_2^1} \arrow[r] & E_2^{h\mathbb{G}_2^1} \arrow[r] & \Sigma L_{K(2)}S \end{tikzcd}$$

Experts will notice that this clashes with Ravenel [17, Lemma 2.1], but this is the natural generalization of what is now commonly denoted by ζ_n at odd primes.

Organization of the paper In Section 2, we specialize to the case $n=2$ and $p=2$ and describe the duality resolution spectral sequence and its $E(1)$–localization. In Section 3, we compute the E_1–page of this spectral sequence for $V(0)$. In Section 4, we prove Theorem 1.3.

Acknowledgements I thank Paul Goerss, Hans-Werner Henn and Peter May for their constant help and support. I thank Tobias Barthel, Daniel G Davis and Douglas Ravenel for helping me sort out some of the details for the proofs of Section 2. I also thank Mark Behrens, Irina Bobkova, Michael Hopkins, Jack Morava, Niko Naumann and Zhouli Xu for helpful conversations. Finally, I thank Mark Mahowald for the insight he shared with all of us throughout his life.

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1612020.

2 The $E(1)$–local duality resolution spectral sequence

We take the point of view that, at height 2, the Honda formal group law may be replaced by the formal group law of a supersingular elliptic curve. This was carefully explained in [3, Section 1]. (The reader who wants to ignore this subtlety may take $\mathbb{S}_{\mathcal{C}}$, $\mathbb{G}_{\mathcal{C}}$ and $E_{\mathcal{C}}$ to mean $\mathbb{S}_2$, $\mathbb{G}_2$ and E_2, respectively.)

Let $\mathbb{S}_{\mathcal{C}}$ be the group of automorphisms of the formal group law of the supersingular elliptic curve

$$\mathcal{C}: \quad y^2+y=x^3$$

of height two over $\mathbb{F}_4$; see [3, Section 3] for the comparison. It admits an action of the Galois group $\mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2)$. Define

$$\mathbb{G}_{\mathcal{C}} = \mathbb{S}_{\mathcal{C}} \rtimes \mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2).$$

Let $E_{\mathcal{C}}$ be the spectrum which classifies the deformations of the formal group law of $\mathcal{C}$ over $\mathbb{F}_4$ as described, for example, in Rezk [20]. It can be chosen to be a complex oriented ring spectrum with

$$(E_{\mathcal{C}})_* = \mathbb{W}[\![u_1]\!][u^{\pm 1}]$$

for $|u_1| = 0$, $|u| = -2$, whose formal group law is the formal group law of the curve

$$\mathcal{C}_U: \quad y^2 + 3u_1xy + (u_1^3 - 1)y = x^3. \tag{2-1}$$

It admits an action of $\mathbb{G}_{\mathcal{C}}$, and for any finite spectrum X,

$$L_{K(2)}X \simeq E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}} \wedge X \simeq (E_{\mathcal{C}} \wedge X)^{h\mathbb{G}_{\mathcal{C}}};$$

see Behrens and Davis [5, page 5]. The group of automorphisms $\mathrm{Aut}(\mathcal{C})$ of $\mathcal{C}$ is of order 24 and injects into $\mathbb{S}_{\mathcal{C}}$. We let G_{24} denote the image of $\mathrm{Aut}(\mathcal{C})$. We note that

$$G_{24} \cong Q_8 \rtimes C_3,$$

where Q_8 is a quaternion subgroup and C_3 a cyclic group of order 3. The group $\mathbb{S}_{\mathcal{C}}$ contains a central subgroup of order 2, which we denote by C_2. We define

$$C_6 = C_2 \times C_3.$$

There is a surjective homomorphism $N\colon \mathbb{S}_{\mathcal{C}} \to \mathbb{Z}_2^\times/(\pm 1) \cong \mathbb{Z}_2$, which we call the *norm*. It is constructed using the determinant of a representation $\rho\colon \mathbb{S}_{\mathcal{C}} \to GL_2(\mathbb{W})$; see [3, Section 3]. Further, it can be extended to $\mathbb{G}_{\mathcal{C}}$. We let $\mathbb{S}_{\mathcal{C}}^1$ and $\mathbb{G}_{\mathcal{C}}^1$ be the kernels of the norms, and note that the elements of finite order in $\mathbb{S}_{\mathcal{C}}$ and $\mathbb{G}_{\mathcal{C}}$ are contained in $\mathbb{S}_{\mathcal{C}}^1$ and $\mathbb{G}_{\mathcal{C}}^1$ respectively. Further,

$$\mathbb{S}_{\mathcal{C}} \cong \mathbb{S}_{\mathcal{C}}^1 \rtimes \mathbb{Z}_2 \quad \text{and} \quad \mathbb{G}_{\mathcal{C}} \cong \mathbb{G}_{\mathcal{C}}^1 \rtimes \mathbb{Z}_2. \tag{2-2}$$

The formal group law $F_{\mathcal{C}_U}$ of $\mathcal{C}_U$, is not 2–typical. Nonetheless, it is strictly isomorphic to a 2–typical formal group law classified by a map $BP_* \to (E_{\mathcal{C}})_*$. Further, $[2]_{F_{\mathcal{C}_U}}(x) \equiv u_1u^{-1}x^2$ modulo $(2, x^4)$; see [3, Section 6.1] for details on $F_{\mathcal{C}_U}$. The strict isomorphism between $F_{\mathcal{C}_U}$ and its 2–typification preserves this identity. Hence, v_1 is mapped to u_1u^{-1} modulo (2). Since we are working primarily modulo (2), we abuse notation and let $v_1 = u_1u^{-1} \in (E_{\mathcal{C}})_2$.

We will need the following result, which can be found in Henn [11, Theorem 13] and is also discussed in greater detail in Bobkova [6]. We restate it here using our notation for convenience.

Theorem 2.1 (Goerss, Henn, Mahowald, Rezk and Bobkova) *There is a resolution of spectra in the $K(2)$–local category given by*

$$\begin{tikzcd} E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \arrow[r] & E_{\mathcal{C}}^{hG_{24}} \arrow[r] \arrow[d, equal] & E_{\mathcal{C}}^{hC_6} \arrow[r] \arrow[d, equal] & E_{\mathcal{C}}^{hC_6} \arrow[r] \arrow[d, equal] & \mathscr{E}_3 \arrow[d, equal] \\ & \mathscr{E}_0 \arrow[r] & \mathscr{E}_1 \arrow[r] & \mathscr{E}_2 \arrow[r] & \mathscr{E}_3 \end{tikzcd}$$

where $(E_{\mathcal{C}})_*\mathscr{E}_3 \cong (E_{\mathcal{C}})_* E_{\mathcal{C}}^{hG_{24}}$ as Morava modules. Further, for any spectrum X, the resolution gives rise to a tower of fibrations spectral sequence

$$E_1^{p,t} = \pi_t(\mathscr{E}_p \wedge X) \overset{SS_1}{\Longrightarrow} \pi_{t-p}(E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \wedge X) \tag{2-3}$$

with differentials $d_r\colon E_r^{p,t} \to E_r^{p+r,t+r-1}$.

We call the resolution of Theorem 2.1 the *duality resolution*. Let π generate $\mathbb{Z}_2$ in the decompositions (2-2), and let $G'_{24} = \pi G_{24}\pi^{-1}$. Recall from [3] or [2] that there is also an *algebraic* duality resolution:

$$\begin{tikzcd} 0 \arrow[r] & \mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1/G'_{24}]\!] \arrow[r] \arrow[d, equal] & \mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1/C_6]\!] \arrow[r] \arrow[d, equal] & \mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1/C_6]\!] \arrow[r] \arrow[d, equal] & \mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1/G_{24}]\!] \arrow[r] \arrow[d, equal] & \mathbb{Z}_2 \arrow[r] & 0 \\ & \mathscr{C}_3 \arrow[r] & \mathscr{C}_2 \arrow[r] & \mathscr{C}_1 \arrow[r] & \mathscr{C}_0 \end{tikzcd} \tag{2-4}$$

Now, let X be a finite spectrum. Resolving (2-4) into a double complex of projective $\mathbb{S}_{\mathcal{C}}^1$–modules and applying the functor $\mathrm{Hom}_{\mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1]\!]}(-, (E_{\mathcal{C}})_t X)$ gives rise to a spectral sequence

$$E_1^{p,q,t} = \mathrm{Ext}^q_{\mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1]\!]}(\mathscr{C}_p, (E_{\mathcal{C}})_t X) \overset{SS_2}{\Longrightarrow} H^{p+q}(\mathbb{S}_{\mathcal{C}}^1, (E_{\mathcal{C}})_t X) \tag{2-5}$$

with differentials $d_r\colon E_r^{p,q,t} \to E_r^{p+r,q-r+1,t}$. Further, in each fixed degree p, there are spectral sequences

$$E_1^{p,q,t} = \mathrm{Ext}^q_{\mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1]\!]}(\mathscr{C}_p, (E_{\mathcal{C}})_t X) \overset{SS_3}{\Longrightarrow} \pi_{t-q}(\mathscr{E}_p \wedge X) \tag{2-6}$$

with differentials $d_r\colon E_r^{p,q,t} \to E_r^{p,q+r,t+r-1}$. Finally, there is also a spectral sequence

$$E_2^{s,t} = H^s(\mathbb{S}_{\mathcal{C}}^1, (E_{\mathcal{C}})_t X) \overset{SS_4}{\Longrightarrow} \pi_{t-s}(E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \wedge X) \tag{2-7}$$

with differentials $d_r\colon E_r^{s,t} \to E_r^{s+r,t+r-1}$. Thus, for X finite, we obtain a diagram of spectral sequences:

$$\begin{tikzcd} \mathrm{Ext}^q_{\mathbb{Z}_2[\![\mathbb{S}^1_{\mathcal{C}}]\!]}(\mathscr{C}_p, (E_{\mathcal{C}})_t X) \arrow[r, Rightarrow, "SS_2"] \arrow[d, Rightarrow, "SS_3"'] & H^{p+q}(\mathbb{S}^1_{\mathcal{C}}, (E_{\mathcal{C}})_t X) \arrow[d, Rightarrow, "SS_4"] \\ \pi_{t-q}(\mathscr{E}_p \wedge X) \arrow[r, Rightarrow, "SS_1"'] & \pi_{t-(p+q)}(E_{\mathcal{C}}^{h\mathbb{S}^1_{\mathcal{C}}} \wedge X) \end{tikzcd} \tag{2-8}$$

Remark 2.2 For elements of Adams–Novikov filtration $s=0$ in $E_1^{p,t}(SS_1)$, the differentials d_1 are related to the d_1–differentials in the algebraic duality resolution spectral sequence SS_2 in the following way. If X is finite, as in [10, Proposition 2.4 and (2.7)], for G a closed subgroup of $\mathbb{G}_{\mathcal{C}}$, there are isomorphisms of Morava modules

$$(E_{\mathcal{C}})_t(E_{\mathcal{C}}^{hG} \wedge X) \cong \mathrm{Hom}^c(\mathbb{G}_{\mathcal{C}}/G, (E_{\mathcal{C}})_t X) \cong \mathrm{Hom}_{\mathbb{Z}_2}(\mathbb{Z}_2[\![\mathbb{G}_{\mathcal{C}}/G]\!], (E_{\mathcal{C}})_t X). \tag{2-9}$$

Let

$$E_1(SS_1)^{p,t} \cong \pi_t(\mathscr{E}_p \wedge X) \xrightarrow{h} H^0(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_t(\mathscr{E}_p \wedge X)) \cong E_1^{p,0,t}(SS_2)$$

be the edge homomorphism for the spectral sequence

$$H^s(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_t(\mathscr{E}_p \wedge X)) \implies \pi_{t-s}(\mathscr{E}_p \wedge X).$$

The spectral sequence SS_1 is constructed so that the following diagram commutes:

$$\begin{tikzcd} E_1^{p,t}(SS_1) \arrow[r, "h"] \arrow[d, "d_1"'] & E_1^{p,0,t}(SS_2) \arrow[d, "d_1"] \\ E_1^{p+1,t}(SS_1) \arrow[r, "h"] & E_1^{p+1,0,t}(SS_2) \end{tikzcd}$$

When both horizontal maps h are injective, one can deduce information in SS_1 from information in SS_2.

For the statement of the next result, recall that for any closed subgroup F of $\mathbb{G}_{\mathcal{C}}$ and finite spectrum X, there is a spectral sequence

$$E_2^{s,t}(F, X) = H^s(F, (E_{\mathcal{C}})_t X) \implies \pi_{t-s}(E_{\mathcal{C}}^{hF} \wedge X). \tag{2-10}$$

The author learned the proof of the following result from Paul Goerss.

Lemma 2.3 *Let S a closed subgroup of $\mathbb{S}_{\mathcal{C}}$ which is invariant under the action of $\mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2)$. Let $G \cong S \rtimes \mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2)$ be the corresponding closed subgroup of $\mathbb{G}_{\mathcal{C}}$. Then for any finite X and any $2 \le r \le \infty$,*

$$E_r^{s,t}(S, X) \cong \mathbb{W} \otimes_{\mathbb{Z}_2} E_r^{s,t}(G, X),$$

and the differentials of the spectral sequence $E_r^{s,t}(S, X)$ are $\mathbb{W}$–linear.

Proof The action of $\mathrm{Gal}(\mathbb{F}_4/\mathbb{F}_2)$ on $(E_{\mathcal{C}})_*X$ is semilinear over $\mathbb{W}$, so there is an isomorphism $E_2^{*,*}(S,X) \cong \mathbb{W} \otimes_{\mathbb{Z}_2} E_2^{*,*}(G,X)$. Now consider, $E_r^{s,t}(\mathbb{S}_{\mathcal{C}}, S^0)$. We have $E_2^{0,0}(\mathbb{S}_{\mathcal{C}}, S^0) \cong \mathbb{W}$ and the subring $\mathbb{Z}_2$ of $\mathbb{W}$ consists of permanent cycles. The spectral sequence $E_r^{*,*}(\mathbb{S}_{\mathcal{C}}, S^0)$ is multiplicative, so the differentials $d_r\colon E_r^{0,0} \to E_r^{r,r-1}$ are $\mathbb{Z}_2$–derivations. Since $\mathbb{W}$ is an étale extension of $\mathbb{Z}_2$, for any r, the $\mathbb{Z}_2$–derivations from $\mathbb{W}$ to the $\mathbb{W}$–module $E_r^{r,r-1}$ are zero. Hence, $E_2^{0,0}(\mathbb{S}_{\mathcal{C}}, S^0) \cong \mathbb{W}$ consists of permanent cycles and the differentials are $\mathbb{W}$–linear. Since the spectral sequence $E_r^{*,*}(S,X)$ is one of modules over $E_r^{*,*}(\mathbb{S}_{\mathcal{C}}, S^0)$, the differentials of $E_r^{*,*}(S,X)$ are also $\mathbb{W}$–linear, and the result follows. □

In what follows, we will use the following remark.

Remark 2.4 Let X be a finite spectrum and F be a closed subgroup of $\mathbb{G}_{\mathcal{C}}$. As noted by Devinatz in the proof of [7, Lemma 3.5], it follows from the fact that $E_{\mathcal{C}}^{hF}$ is $(K_{\mathcal{C}})_*$–local $E_{\mathcal{C}}$–nilpotent, (see Devinatz and Hopkins [8, Proposition A.3]) that the descent spectral sequence (2-10) has a horizontal vanishing line.

Now, recall that the telescope conjecture holds at height $n = 1$. This was proved at odd primes by Miller [15] and at $p = 2$ by Mahowald [14]. In particular, we have the following result.

Theorem 2.5 (Mahowald and Miller) *Let Y admit a v_1–self-map $v_1^k\colon \Sigma^{2k}Y \to Y$. Then*

$$L_1 Y \simeq L_{K(1)} Y \simeq v_1^{-1} Y,$$

where

$$v_1^{-1}Y := \operatorname{colim}\Big(\cdots \xrightarrow{v_1^k} \Sigma^{2k} Y \xrightarrow{v_1^k} Y \xrightarrow{v_1^k} \cdots\Big).$$

Proposition 2.6 *For any finite type-1 spectrum X, with self map $v_1^k\colon \Sigma^{2k}X \to X$, there is a diagram of strongly convergent spectral sequences:*

$$\begin{tikzcd}
v_1^{-1}\operatorname{Ext}^q_{\mathbb{Z}_2[\![\mathbb{S}_{\mathcal{C}}^1]\!]}(\mathscr{C}_p, (E_{\mathcal{C}})_t X) \arrow[r, Rightarrow, "L_1 SS_2"] \arrow[d, Rightarrow, "L_1 SS_3"'] & v_1^{-1} H^{p+q}(\mathbb{S}_{\mathcal{C}}^1, (E_{\mathcal{C}})_t X) \arrow[d, Rightarrow, "L_1 SS_4"] \\
\pi_{t-q} L_1(\mathscr{E}_p \wedge X) \arrow[r, Rightarrow, "L_1 SS_1"'] & \pi_{t-(p+q)} L_1(E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \wedge X)
\end{tikzcd}$$

Proof The spectral sequence $L_1 SS_2$ is obtained from SS_2 by inverting the element $v_1^k \in (E_{\mathcal{C}})_{2k} X$, and $L_1 SS_1$ is obtained by the applying L_1 to the tower of fibrations which gives rise to SS_1. The spectral sequences $L_1 SS_3$ and $L_1 SS_4$ are obtained by inverting the algebraic element v_1^k in the spectral sequences SS_3 or SS_4, and using the fact that

$$v_1^{-1}\pi_*(\mathscr{E}_p \wedge X) \cong \pi_* L_1(\mathscr{E}_p \wedge X).$$

With regards to the strong convergence of the four spectral sequences, note that localization with respect to v_1 is exact. Therefore, the localized spectral sequences will converge strongly if they have horizontal vanishing lines at the E_∞–term. The spectral sequences SS_1 and SS_2 have a vanishing line at $p = 4$ for all $r \geq 1$. As noted in Remark 2.4, the descent spectral sequences SS_3 and SS_4 have horizontal vanishing lines. Therefore, the spectral sequences $L_1 SS_i$ exist and converge. □

Remark 2.7 As in Remark 2.2, the differentials d_1 in $L_1 SS_1$ and $L_1 SS_2$ commute with the edge homomorphisms

$$E_1(L_1 SS_1)^{p,t} \cong \pi_t L_1(\mathscr{E}_p \wedge X) \xrightarrow{h} v_1^{-1} H^0(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_t(\mathscr{E}_p \wedge X)) \cong E_1^{p,0,t}(L_1 SS_2).$$

Remark 2.8 For X as in Proposition 2.6, the element $v_1^{2k} \in (E_{\mathcal{C}})_{2k} X$ can be chosen to be Galois invariant. Therefore, the results of Lemma 2.3 also hold for the localized spectral sequences. That is, let

$$v_1^{-1} E_2^{s,t}(F, X) \cong v_1^{-1} H^s(F, (E_{\mathcal{C}})_t X) \implies \pi_{t-s} L_1(E_{\mathcal{C}}^{hF} \wedge X).$$

Then for S and G as in Lemma 2.3, we have

$$v_1^{-1} E_r^{s,t}(S, X) \cong \mathbb{W} \otimes_{\mathbb{Z}_2} v_1^{-1} E_r^{s,t}(G, X)$$

for $2 \leq r \leq \infty$, and the differentials are $\mathbb{W}$–linear.

3 The homotopy of $L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0))$ and $L_1(E_{\mathcal{C}}^{hC_6} \wedge V(0))$

The spectrum $V(0)$ has a self map

$$\beta\colon \Sigma^8 V(0) \xrightarrow{v_1^4} V(0),$$

and in this section, we give the E_1–term for

$$E_1^{p,q}(L_1 SS_1) = \pi_q L_1(\mathscr{E}_p \wedge V(0)) \xrightarrow[L_1 SS_1]{} \pi_{q-p} L_1(E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \wedge V(0)).$$

In order to do so, we must compute $\pi_* L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0))$ and $\pi_* L_1(E_{\mathcal{C}}^{hC_6} \wedge V(0))$. We do this using the descent spectral sequences

$$v_1^{-1} H^s(G, (E_{\mathcal{C}})_t V(0)) \implies \pi_{t-s} L_1(E_{\mathcal{C}}^{hG} \wedge V(0)).$$

Notation 3.1 We use the following conventions. First,

$$v_1 = u_1 u^{-1}, \quad v_2 = u^{-3} \quad \text{and} \quad j_0 = u_1^3.$$

The element Δ is the discriminant of $\mathcal{C}_U$, and hence is given by

$$\Delta = 27v_2(v_1^3 - v_2)^3 \equiv v_2(v_1^3 + v_2)^3 \mod (2),$$

and

$$c_4 = 9v_1^4 + 72v_1v_2 \equiv v_1^4 \mod (2).$$

The j–invariant is

$$j = c_4^3\Delta^{-1} \equiv v_1^{12}\Delta^{-1} \mod (2).$$

These identities can be computed using Silverman [24, Section III.1]; see also [3, Section 4.2]. We abuse notation and let

$$\eta = \delta(v_1),$$

where δ is the Bockstein associated to

$$0 \to (E_{\mathcal{C}})_*/2 \xrightarrow{2} (E_{\mathcal{C}})_*/4 \to (E_{\mathcal{C}})_*/2 \to 0.$$

This is justified by the fact that $\delta(v_1)$ detects the homotopy class η (see [3, Section 4.1]).

The v_1–torsion-free elements of $H^*(G_{24}, (E_{\mathcal{C}})_*V(0))$ generate a submodule isomorphic to

$$\mathbb{F}_4[\![j]\!][v_1, \eta, \Delta^{\pm 1}, k]/(\eta^4 - v_1^4k, j\Delta - v_1^{12})$$

for elements of degrees (s,t), where s is the cohomological grading, t is the internal grading, and

$$|v_1| = (0,2), \quad |\eta| = (1,2), \quad |\Delta| = (0,24), \quad |k| = (4,0), \quad |j| = (0,0);$$

see Section 4.2 or the appendix of [3]. On the other hand, $H^*(C_6, (E_{\mathcal{C}})_*V(0))$ is v_1–torsion-free and is isomorphic to

$$\mathbb{F}_4[\![j_0]\!][v_1, \eta, v_2^{\pm 1}, h]/(\eta - v_1h, j_0v_2 - v_1^3),$$

where $|v_2| = (0,6)$, $|h| = (1,0)$ and $|j_0| = (0,0)$; see Section 4.2 of [3].

The next proposition is an immediate consequence of these results. In its statement, we let $\mathbb{F}_4(\!(x)\!)$ denote the Laurent series on x.

Proposition 3.2 *There are isomorphisms*

$$v_1^{-1}H^*(G_{24}, (E_{\mathcal{C}})_*V(0)) \cong \mathbb{F}_4(\!(j)\!)[v_1^{\pm 1}, \eta]$$

and

$$v_1^{-1}H^*(C_6; (E_{\mathcal{C}})_*V(0)) \cong \mathbb{F}_4(\!(j_0)\!)[v_1^{\pm 1}, \eta].$$

The degrees (s,t) are given by $|v_1| = (0,2)$, $|\eta| = (1,2)$, $|j| = (0,0)$ and $|j_0| = (0,0)$. The restriction associated to the inclusion of C_6 in G_{24} maps j to $j_0^4(1 + j_0)^{-3}$.

Proof This follows from [3, Section 4.2] after inverting v_1. □

To compute the differentials, we will use the following observation.

Remark 3.3 There is a class α_3 in $\mathrm{Ext}^{1,6}_{BP_*BP}(BP_*, BP_*)$ (see Ravenel [18, page 430]) such that $d_3(\alpha_3) = \eta^4$. Further, α_3 reduces to ηv_1^2 in $\mathrm{Ext}^{1,6}_{BP_*BP}(BP_*, BP_*V(0))$, so $\eta d_3(v_1^2) = \eta^4$.

In general, for a 2–local BP–algebra spectrum E, the E–Adams spectral sequence for any spectrum X is a module over $\mathrm{Ext}_{BP_*BP}(BP_*, BP_*)$. There is a universal d_3–differential $d_3(\alpha_3 z) = \eta^4 z + \alpha_3 d_3(z)$. Further, if 2 annihilates $E_*(X)$, this reduces to $d_3(\eta v_1^2 z) = \eta^4 z + \eta v_1^2 d_3(z)$. If there is no η–torsion on the E_3–term as in our examples below, this gives a universal differential $d_3(v_1^2 z) = \eta^3 z + v_1^2 d_3(z)$.

Lemma 3.4 *Let G be a closed subgroup of $\mathbb{G}_{\mathcal{C}}$. Let X be a $K(2)$–local spectrum such that $(E_{\mathcal{C}})_*X \cong (E_{\mathcal{C}})_*E_{\mathcal{C}}^{hG}$. Then the $K(2)$–local, $E_{\mathcal{C}}$–Adams spectral sequence computing π_*X has E_2–term isomorphic to $H^*(G, (E_{\mathcal{C}})_*)$.*

Proof We first prove that the E_2–term is isomorphic to $H^*(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_*X)$. This can be deduced directly from Barthel and Heard [1, Theorem 4.3]. Nonetheless, we sketch the proof here. The assumption on $(E_{\mathcal{C}})_*X$ implies that it is profree as an $(E_{\mathcal{C}})_*$–module. An inductive argument using [13, Proposition 8.4] and [10, Proposition 2.4] shows that

$$\pi_* L_{K(2)}(E_{\mathcal{C}}^{\wedge k} \wedge X) \cong \mathrm{Hom}^c(\mathbb{G}_{\mathcal{C}}^{k-1}, (E_{\mathcal{C}})_*X),$$

which allows us to identify the E_2–term as $H^*(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_*X)$. Now, using the fact that $(E_{\mathcal{C}})_*X \cong (E_{\mathcal{C}})_*E_{\mathcal{C}}^{hG}$ as Morava modules, (2-9) and Shapiro's lemma imply that $H^*(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_*X) \cong H^*(G, (E_{\mathcal{C}})_*)$. □

Lemma 3.5 *Let X be a $K(2)$–local spectrum such that $(E_{\mathcal{C}})_*X \cong (E_{\mathcal{C}})_*E_{\mathcal{C}}^{hG_{24}}$ as Morava modules. Then the $K(2)$–local, $E_{\mathcal{C}}$–Adams spectral sequence computing $\pi_*(X \wedge V(0))$ has E_2–term isomorphic to $H^*(G_{24}, (E_{\mathcal{C}})_*V(0))$. Further, in this spectral sequence, the elements Δ^k and $v_1\Delta^k$ are d_3–cycles for all k.*

Proof The identification of the E_2–term follows from Lemma 3.4 and the five lemma. There are no d_2–differentials, so all elements survive to the E_3–term. Let $\epsilon = 0, 1$. It follows from [2, Theorem 4.2.2], that $d_3(v_1^{\epsilon}\Delta^k) = v_1^{10+\epsilon}\eta^3 p(j)\Delta^{k-1}$ for $p(j) \in \mathbb{F}_4[\![j]\!]$. Suppose that $p(j)$ is not zero. Then $p(j) = j^r p_0(j)$ for $r \geq 0$ and $p_0(j) \in \mathbb{F}_4[\![j]\!]$ such that $p_0(j) \equiv \ell$ modulo (j) for some $\ell \in \mathbb{F}_4^{\times}$. Using the fact that the differentials are η– and v_1^4–linear (since $X \wedge V(0)$ has a v_1^4–self map), Remark 3.3 and the identity $j = v_1^{12}\Delta^{-1}$, we have

$$\begin{aligned} 0 &= d_3(v_1^{10+\epsilon}\eta^3 p(j)\Delta^{k-1}) \\ &= v_1^{12r+8}\eta^3 d_3(v_1^{2+\epsilon} p_0(j)\Delta^{k-r-1}) \\ &= v_1^{12r+8+\epsilon}\eta^6 p_0(j)\Delta^{k-r-1} + v_1^{12r+10}\eta^3 d_3(v_1^{\epsilon} p_0(j)\Delta^{k-r-1}). \end{aligned}$$

Again, by [2, Theorem 4.2.2], $H^3(G_{24}, (E_{\mathcal{C}})_t V(0))$ is $\mathbb{F}_4[v_1, \eta]$–torsion-free in degrees $t \equiv 6 + 2\epsilon$ modulo (24), so we can conclude that

$$\eta^3 p_0(j)\Delta^{k-r-1} = v_1^{2-\epsilon} d_3(v_1^{\epsilon} p_0(j)\Delta^{k-r-1}).$$

Since $\epsilon = 0$ or 1, the right-hand side is divisible by v_1, while the left-hand side is not, a contradiction. Therefore, we must have $p(j) = 0$. □

In the next two propositions, we let

$$R(-) = \mathbb{W}((-))[\beta^{\pm 1}, \eta]/(2\eta, \eta^3).$$

Proposition 3.6 *Let X be as in Lemma 3.5. The $E(1)$–localization of the $K(2)$–local, $E_{\mathcal{C}}$–Adams spectral sequence*

$$E_2^{s,t} = v_1^{-1} H^s\big(\mathbb{G}_{\mathcal{C}}, (E_{\mathcal{C}})_t(X \wedge V(0))\big) \Longrightarrow \pi_{t-s} L_1(X \wedge V(0))$$

satisfies

$$E_\infty^{s,t} \cong R(j)\{x, v_1 x\}/(2 \cdot x, 2v_1 x)$$

for x in $(0,0)$ and $v_1 x \in (0,2)$. Further, $\pi_{8t} L_1(X \wedge V(0)) \cong \mathbb{F}_4((j))\{\beta^t\}$ and the edge homomorphisms

$$h\colon \pi_{8t} L_1(X \wedge V(0)) \to v_1^{-1} H^0(G_{24}, (E_{\mathcal{C}})_{8t} V(0))$$

are isomorphisms.

Proof By Lemma 3.5 and naturality, $E_2^{s,t}$ is isomorphic to $v_1^{-1}H^s(G_{24}, (E_{\mathcal{C}})_t V(0))$ and $j^k = v_1^{12k}\Delta^{-k}$ and $v_1 j^k$ are d_3–cycles. By Remark 3.3, there are differentials $d_3(v_1^2 j^k) = \eta^3 j^k$ and $d_3(v_1^3 j^k) = v_1\eta^3 j^k$. This, together with the fact that the differentials are v_1^4–linear, determines all d_3–differentials. The E_4–term has a horizontal vanishing line at $s = 3$. Therefore, there cannot be any higher differentials. Letting x be the element detected by $1 \in H^0(G_{24}, (E_{\mathcal{C}})_0 V(0))$, $v_1 x$ the element detected by $v_1 \in H^0(G_{24}, (E_{\mathcal{C}})_2 V(0))$ and β^t the element detected by v_1^{4t}, we obtain the desired description of the E_∞–term. For degree reasons, $\pi_{8t} L_1(X \wedge V(0)) \cong \mathbb{F}_4((j))\{\beta^t\}$. That the edge homomorphisms are isomorphisms in degrees $8t$ follows since $v_1^{-1}H^0(G_{24}, (E_{\mathcal{C}})_{8t} V(0)) \cong \mathbb{F}_4((j))\{v_1^{4t}\}$ and $h(j^k\beta^t) = j^k v_1^{4t}$. □

Remark 3.7 When $X = V(0)$, the class x can be described as the composite $S^0 \to L_1 E_{\mathcal{C}}^{hG_{24}} \xrightarrow{1\wedge e_0} L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0))$, where the first map is the unit and e_0 is the

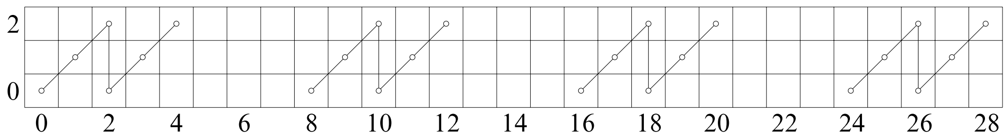


Figure 2: This picture is both an illustration of the homotopy groups $\pi_* L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0))$ and of the homotopy groups $\pi_* L_1(E_{\mathcal{C}}^{hC_6} \wedge V(0))$. For the former, a $\circ$ denotes a copy of $\mathbb{F}_4((j))$, and for the latter, it denotes a copy of $\mathbb{F}_4((j_0))$.

inclusion of the bottom cell. In $\pi_* V(0)_{(2)}$, there is a relation $2v_1 e_0 = \eta^2 e_0$ for $v_1 e_0$ detected by $v_1 \in BP_2 V(0)$ in the Adams–Novikov spectral sequence. This then implies that $2v_1 x = \eta^2 x$ in $\pi_* L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0))$, so

$$\pi_* L_1(E_{\mathcal{C}}^{hG_{24}} \wedge V(0)) \cong R(j)\{x, v_1 x\}/(2 \cdot x, 2v_1 x - \eta^2 x).$$

With some work, one can show that the relation $2v_1 x = \eta^2 x$ holds for arbitrary X satisfying the condition of Lemma 3.5. However, this fact is not needed here.

Proposition 3.8 *There is an isomorphism*

$$\pi_* L_1(E_{\mathcal{C}}^{hC_6} \wedge V(0)) \cong R(j_0)\{y, v_1 y\}/(2 \cdot y, 2v_1 y - \eta^2 y)$$

for y in $(0,0)$ and $v_1 y \in (0,2)$; see Figure 2. Hence, $\pi_* L_1(E_2^{hC_6} \wedge V(0))$ is 8–periodic with periodicity generator β. Further, the edge homomorphisms

$$h\colon \pi_{8t} L_1(E_{\mathcal{C}}^{hC_6} \wedge V(0)) \to v_1^{-1} H^0(C_6, (E_{\mathcal{C}})_{8t} V(0))$$

are isomorphisms.

Proof We prove that j_0^k is a d_3–cycle for all integers k. Then an argument similar to that of Proposition 3.6 finishes the computation of the E_∞–term, where we let y be the element detected by $1 \in H^0(C_6, (E_{\mathcal{C}})_0 V(0))$ and $v_1 y$ be the element detected by $v_1 y \in H^0(C_6, (E_{\mathcal{C}})_2 V(0))$. The extension is obtained as in Remark 3.7.

The spectral sequence $H^*(C_6, (E_{\mathcal{C}})_*) \Rightarrow \pi_* E_{\mathcal{C}}^{hC_6}$ is multiplicative; hence, in this spectral sequence, all elements of the form a^2 are d_3–cycles. Note that j_0 lifts to an invariant in $H^0(C_6, (E_{\mathcal{C}})_0)$. This implies that $d_3(j_0^{2r}) = 0$ and $d_3(j_0^{2r+1}) = j_0^{2r} d_3(j_0)$. Hence, it suffices to prove that j_0 is a d_3–cycle. The restriction induced by the inclusion of C_6 in G_{24}, maps j to $j_0^4(1+j_0)^{-3}$. By naturality, the element $d_3(j_0^4(1+j_0)^{-3}) = 0$. However,

$$d_3(j_0^4(1+j_0)^{-3}) = j_0^4(1+j_0)^{-4} d_3(1+j_0) = j_0^4(1+j_0)^{-4} d_3(j_0),$$

which implies that $d_3(j_0) = 0$. $\square$

4 Some elements in $\pi_* L_1 L_{K(2)} V(0)$

We now turn to examining the spectral sequence

$$E_1^{p,q}(L_1 SS_1) = \pi_q L_1(\mathscr{E}_p \wedge V(0)) \xRightarrow{L_1 SS_1} \pi_{q-p} L_1(E_{\mathcal{C}}^{h\mathbb{S}^1_{\mathcal{C}}} \wedge V(0)).$$

The idea is to use knowledge of the differentials in the spectral sequence

$$E_1^{p,q,t}(L_1 SS_2) = v_1^{-1} \operatorname{Ext}^q_{\mathbb{Z}_2[\![\mathbb{S}^1_{\mathcal{C}}]\!]}(\mathscr{C}_p, (E_{\mathcal{C}})_t V(0)) \xRightarrow{L_1 SS_2} v_1^{-1} H^{p+q}(\mathbb{S}^1_{\mathcal{C}}.(E_{\mathcal{C}})_t V(0))$$

to deduce information about the differentials of $L_1 SS_1$.

Lemma 4.1 *In the spectral sequence $L_1 SS_1$, we have $E_2^{3,8t} \cong \mathbb{F}_4\{\beta^t\}$.*

Proof From Section 3, we have that

$$E_1^{p,8t} \cong \begin{cases} \mathbb{F}_4((j))\{\beta^t\}, & p = 0, 3, \\ \mathbb{F}_4((j_0))\{\beta^t\}, & p = 1, 2. \end{cases}$$

From Remark 2.7 and the fact that the edge homomorphisms are isomorphisms in these degrees, we obtain a commutative diagram

$$\begin{array}{ccccccc} E_1^{0,8t}(L_1 SS_1) & \xrightarrow{d_1} & E_1^{1,8t}(L_1 SS_1) & \xrightarrow{d_1} & E_1^{2,8t}(L_1 SS_1) & \xrightarrow{d_1} & E_1^{3,8t}(L_1 SS_1) \\ \downarrow \cong & & \downarrow \cong & & \downarrow \cong & & \downarrow \cong \\ E_1^{0,0,8t}(L_1 SS_2) & \xrightarrow{d_1} & E_1^{1,0,8t}(L_1 SS_2) & \xrightarrow{d_1} & E_1^{2,0,8t}(L_1 SS_2) & \xrightarrow{d_1} & E_1^{3,0,8t}(L_1 SS_2) \end{array}$$

where β^{4t} maps to v_1^{4t}. Theorem 1.2.1 and Corollary 1.2.3 of [3] give a computation of the spectral sequence $L_1 SS_2$. In particular, it follows immediately from these results that

$$E_2^{3,0,8t}(L_1 SS_2) \cong \mathbb{F}_4((j))\{v_1^{4t}\}/(j) \cong \mathbb{F}_4\{v_1^{4t}\}.$$

The claim follows. □

Proposition 4.2 *If $k \equiv -3$ modulo 8, then $\pi_k L_1(E_{\mathcal{C}}^{h\mathbb{S}^1_{\mathcal{C}}} \wedge V(0)) \cong \mathbb{F}_4$.*

Proof We use the spectral sequence $E_r^{p,q} = E_r^{p,q}(L_1 SS_1)$. From Proposition 3.6 applied to $X = \mathscr{E}_0$ and $X = \mathscr{E}_3$, and from Proposition 3.8, it follows that for $r = 1$, 2 or 3 and for any p,

$$E_1^{p,8t-r} = \pi_{8t-r} L_1(\mathscr{E}_p \wedge V(0)) = 0.$$

By Lemma 4.1, $E_2^{3,8t} \cong \mathbb{F}_4\{\beta^{8t}\}$, which proves the claim. □

Proposition 4.3 *If $k \equiv -3$ modulo 8, then $\pi_k L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0)) \cong \mathbb{F}_2$.*

Proof It follows from Remark 2.8 that

$$v_1^{-1} E_\infty^{*,*}(\mathbb{S}_{\mathcal{C}}^1, V(0)) \cong \mathbb{W} \otimes_{\mathbb{Z}_2} v_1^{-1} E_\infty^{*,*}(\mathbb{G}_{\mathcal{C}}^1, V(0)).$$

Since $\pi_k L_1(E_{\mathcal{C}}^{h\mathbb{S}_{\mathcal{C}}^1} \wedge V(0)) \cong \mathbb{F}_4$, there is a unique $s_0 \geq 0$ such that $E_\infty^{s_0,k+s_0}(\mathbb{S}_{\mathcal{C}}^1, V(0))$ is nonzero, and $E_\infty^{s_0,k+s_0}(\mathbb{S}_{\mathcal{C}}^1, V(0)) \cong \mathbb{F}_4$. Therefore, $E_\infty^{s,k+s}(\mathbb{G}_{\mathcal{C}}^1, V(0)) = 0$ if $s \neq s_0$ and $E_\infty^{s_0,k+s_0}(\mathbb{G}_{\mathcal{C}}^1, V(0)) \cong \mathbb{F}_2$. □

Definition 4.4 Define the class $x \in \pi_{-3} L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0))$ to be the nonzero element.

Recall that

$$\mathbb{G}_{\mathcal{C}} \cong \mathbb{G}_{\mathcal{C}}^1 \rtimes \mathbb{Z}_2.$$

Let π be a topological generator of the subgroup $\mathbb{Z}_2$ in $\mathbb{G}_{\mathcal{C}}$. There is a cofiber sequence

$$L_{K(2)} S \to E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \xrightarrow{\pi - 1} E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1}. \tag{4-1}$$

We can now prove our main result.

Proof of Theorem 1.3 Since $L_{K(2)} S \wedge V(0) \simeq L_{K(2)} V(0)$ and localization preserves exact triangles, the fiber sequence (4-1) gives rise to a fiber sequence

$$L_1 L_{K(2)} V(0) \to L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0)) \xrightarrow{\pi - 1} L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0)). \tag{4-2}$$

Since π acts by automorphisms and the only automorphism of $\mathbb{F}_2$ is the identity, the map $\pi - 1$ acts trivially on $\pi_{8t-3} L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0))$. Therefore, in the long exact sequence on homotopy groups, the class $\beta^t x$ is in the kernel of $\pi - 1$, and the image of $\beta^t x$ under the map $L_1(E_{\mathcal{C}}^{h\mathbb{G}_{\mathcal{C}}^1} \wedge V(0)) \to \Sigma L_1 L_{K(2)} V(0)$ is nonzero. We denote it by $\zeta_2 \beta^t x$. □

References

[1] **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

[2] **A Beaudry**, *The algebraic duality resolution at $p = 2$*, Algebr. Geom. Topol. 15 (2015) 3653–3705 MR

[3] **A Beaudry**, *Towards the homotopy of the $K(2)$–local Moore spectrum at $p = 2$*, Adv. Math. 306 (2017) 722–788 MR

[4] **M Behrens**, *The homotopy groups of $S_{E(2)}$ at $p \geq 5$ revisited*, Adv. Math. 230 (2012) 458–492 MR

[5] **M Behrens**, **D G Davis**, *The homotopy fixed point spectra of profinite Galois extensions*, Trans. Amer. Math. Soc. 362 (2010) 4983–5042 MR

[6] **I Bobkova**, *Resolutions in the $K(2)$–local category at the prime 2*, PhD thesis, Northwestern University (2014) MR Available at `https://search.proquest.com/docview/1558126694`

[7] **E S Devinatz**, *A Lyndon–Hochschild–Serre spectral sequence for certain homotopy fixed point spectra*, Trans. Amer. Math. Soc. 357 (2005) 129–150 MR

[8] **E S Devinatz**, **M J Hopkins**, *Homotopy fixed point spectra for closed subgroups of the Morava stabilizer groups*, Topology 43 (2004) 1–47 MR

[9] **P G Goerss**, **H-W Henn**, **M Mahowald**, *The rational homotopy of the $K(2)$–local sphere and the chromatic splitting conjecture for the prime 3 and level 2*, Doc. Math. 19 (2014) 1271–1290 MR

[10] **P Goerss**, **H-W Henn**, **M Mahowald**, **C Rezk**, *A resolution of the $K(2)$–local sphere at the prime 3*, Ann. of Math. 162 (2005) 777–822 MR

[11] **H-W Henn**, *On finite resolutions of $K(n)$–local spheres*, from "Elliptic cohomology" (H R Miller, D C Ravenel, editors), London Math. Soc. Lecture Note Ser. 342, Cambridge Univ. Press (2007) 122–169 MR

[12] **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

[13] **M Hovey**, **N P Strickland**, *Morava K–theories and localisation*, Mem. Amer. Math. Soc. 666, Amer. Math. Soc., Providence, RI (1999) MR

[14] **M Mahowald**, *The image of J in the EHP sequence*, Ann. of Math. 116 (1982) 65–112 MR

[15] **H R Miller**, *On relations between Adams spectral sequences, with an application to the stable homotopy of a Moore space*, J. Pure Appl. Algebra 20 (1981) 287–312 MR

[16] **J Morava**, *Noetherian localisations of categories of cobordism comodules*, Ann. of Math. 121 (1985) 1–39 MR

[17] **D C Ravenel**, *The cohomology of the Morava stabilizer algebras*, Math. Z. 152 (1977) 287–297 MR

[18] **D C Ravenel**, *A novice's guide to the Adams–Novikov spectral sequence*, from "Geometric applications of homotopy theory, II" (M G Barratt, M E Mahowald, editors), Lecture Notes in Math. 658, Springer (1978) 404–475 MR

[19] **D C Ravenel**, *Localization and periodicity in homotopy theory*, from "Homotopy theory" (E Rees, J D S Jones, editors), London Math. Soc. Lecture Note Ser. 117, Cambridge Univ. Press (1987) 175–194 MR

[20] **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

[21] **K Shimomura**, *The Adams–Novikov E_2–term for computing $\pi_*(L_2V(0))$ at the prime 2*, Topology Appl. 96 (1999) 133–152 MR

[22] **K Shimomura**, **X Wang**, *The Adams–Novikov E_2–term for $\pi_*(L_2S^0)$ at the prime 2*, Math. Z. 241 (2002) 271–311 MR

[23] **K Shimomura**, **A Yabe**, *The homotopy groups $\pi_*(L_2S^0)$*, Topology 34 (1995) 261–289 MR

[24] **J H Silverman**, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics 106, Springer (1986) MR

Department of Mathematics, University of Colorado
Boulder, CO, United States

agnes.beaudry@colorado.edu

Proposed: Mark Behrens | Received: 3 April 2015
Seconded: Stefan Schwede, Ralph Cohen | Revised: 7 December 2016

Geometry & Topology 21 (2017) 3231–3311

Virtual fundamental classes for moduli spaces of sheaves on Calabi–Yau four-folds

DENNIS BORISOV
DOMINIC JOYCE

Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ be a separated, -2–shifted symplectic derived $\mathbb{C}$–scheme, in the sense of Pantev, Toën, Vezzosi and Vaquié (2013), of complex virtual dimension $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X} = n \in \mathbb{Z}$, and X_{an} the underlying complex analytic topological space. We prove that X_{an} can be given the structure of a derived smooth manifold $\boldsymbol{X}_{\mathrm{dm}}$, of real virtual dimension $\mathrm{vdim}_{\mathbb{R}} \boldsymbol{X}_{\mathrm{dm}} = n$. This $\boldsymbol{X}_{\mathrm{dm}}$ is not canonical, but is independent of choices up to bordisms fixing the underlying topological space X_{an}. There is a one-to-one correspondence between orientations on $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ and orientations on $\boldsymbol{X}_{\mathrm{dm}}$.

Because compact, oriented derived manifolds have virtual classes, this means that proper, oriented -2–shifted symplectic derived $\mathbb{C}$–schemes have virtual classes, in either homology or bordism. This is surprising, as conventional algebrogeometric virtual cycle methods fail in this case. Our virtual classes have half the expected dimension.

Now derived moduli schemes of coherent sheaves on a Calabi–Yau 4–fold are expected to be -2–shifted symplectic (this holds for stacks). We propose to use our virtual classes to define new Donaldson–Thomas style invariants "counting" (semi)stable coherent sheaves on Calabi–Yau 4–folds Y over $\mathbb{C}$, which should be unchanged under deformations of Y.

14A20; 14N35, 14J35, 14F05, 55N22, 53D30

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3231

1 Introduction

This paper will relate two apparently rather different classes of "derived" geometric spaces. The first class is *derived* $\mathbb{C}$*–schemes* $\boldsymbol{X}$, in the derived algebraic geometry of Toën and Vezzosi [34; 36], equipped with a -2*–shifted symplectic structure* $\omega_{\boldsymbol{X}}^*$ in the sense of Pantev, Toën, Vaquié and Vezzosi [31]. Such $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ are the expected structures on 4–Calabi–Yau derived moduli $\mathbb{C}$–schemes.

The second class is *derived smooth manifolds* $\boldsymbol{X}_{\mathrm{dm}}$, in derived differential geometry. There are several different models available: the *derived manifolds* of Spivak [32] and Borisov and Noël [3; 4] (which form ∞–categories $\mathbf{DerMan}_{\mathrm{Spi}}$, $\mathbf{DerMan}_{\mathrm{BoNo}}$), and Joyce's *d-manifolds* [18; 19; 20] (a strict 2–category $\mathbf{dMan}$) and *m-Kuranishi spaces* [21, Section 4.7] (a weak 2–category $\mathbf{mKur}$).

As it is known that equivalence classes of objects in all these higher categories are in natural bijection, these four models are interchangeable for our purposes. But we use theorems proved for d-manifolds or (m-)Kuranishi spaces.

Here is a summary of our main results, taken from Theorems 3.15, 3.16 and 3.24 and Propositions 3.17 and 3.18 below.

Theorem 1.1 *Let $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme, in the sense of Pantev et al [31], with complex virtual dimension $\mathrm{vdim}_{\mathbb{C}}\,\boldsymbol{X}=n$ in $\mathbb{Z}$, and write X_{an} for the set of $\mathbb{C}$–points of $X=t_0(\boldsymbol{X})$, with the complex analytic topology. Suppose that X is separated, and X_{an} is second countable. Then we can make the topological space X_{an} into a derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ of real virtual dimension $\mathrm{vdim}_{\mathbb{R}}\,\boldsymbol{X}_{\mathrm{dm}}=n$, in the sense of any of Borisov and Noel [3; 4], Joyce [18; 19; 20; 21] and Spivak [32].*

There is a natural one-to-one correspondence between orientations on $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$, in the sense of Section 2.4, and orientations on $\boldsymbol{X}_{\mathrm{dm}}$, in the sense of Section 2.6.

The (oriented) derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ above depends on arbitrary choices made in its construction. However, $\boldsymbol{X}_{\mathrm{dm}}$ is independent of choices up to (oriented) bordisms of derived manifolds which fix the underlying topological space.

All the above extends to (oriented) -2–shifted symplectic derived schemes

$$(\boldsymbol{\pi}\colon \boldsymbol{X}\to Z,\ \omega_{\boldsymbol{X}/Z}^*)$$

over a base Z which is a smooth affine $\mathbb{C}$–scheme of pure dimension, yielding an (oriented) derived manifold $\boldsymbol{\pi}_{\mathrm{dm}}\colon \boldsymbol{X}_{\mathrm{dm}}\to Z_{\mathrm{an}}$ over the complex manifold Z_{an} associated to Z, regarded as an (oriented) real manifold.

In Section 2.5 we give a short definition of *Kuranishi atlases* $\mathcal{K}$ on a topological space X. These are families of "Kuranishi neighbourhoods" (V, E, s, ψ) on X and "coordinate changes" between them, based on work of Fukaya, Oh, Ohta and Ono [14; 15] in symplectic geometry. The hard work in proving Theorem 1.1 is using $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ to construct a Kuranishi atlas $\mathcal{K}$ on X_{an}. Then we use results from Borisov and Noel [3; 4] and Joyce [18; 19; 20; 21] to convert $(X_{\mathrm{an}}, \mathcal{K})$ into a derived manifold $\boldsymbol{X}_{\mathrm{dm}}$.

Readers of this papers do not need to understand derived manifolds, if they do not want to. They can just think in terms of Kuranishi atlases, as is common in symplectic geometry, without passing to derived manifolds.

We prove Theorem 1.1 using a "Darboux theorem" for k–shifted symplectic derived schemes by Brav, Bussi and Joyce [6]. This paper is related to the series Ben-Bassat, Brav, Bussi and Joyce [2], Brav, Bussi and Joyce [6], Brav, Bussi, Dupont, Joyce and Szendrői [5], Bussi, Joyce and Meinhardt [7] and Joyce [22], mostly concerning the -1–shifted (3–Calabi–Yau) case.

An important motivation for proving Theorem 1.1 is that *compact, oriented derived manifolds have virtual classes*, in both bordism and homology. As in Sections 3.6–3.7, from Theorem 1.1 we may deduce:

Corollary 1.2 *Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ be a proper, oriented -2–shifted symplectic derived $\mathbb{C}$–scheme, with $\mathrm{vdim}_{\mathbb{C}}\, \boldsymbol{X} = n$. Theorem 1.1 gives a compact, oriented derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ with $\mathrm{vdim}_{\mathbb{R}}\, \boldsymbol{X}_{\mathrm{dm}} = n$. We may define a **d-bordism class** $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{dbo}}$ in the bordism group $B_n(*)$, and a **virtual class** $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{virt}}$ in the homology group $H_n(X_{\mathrm{an}}; \mathbb{Z})$, depending only on $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ and its orientation.*

Let $\boldsymbol{X}$ be a derived $\mathbb{C}$–scheme, Z a connected $\mathbb{C}$–scheme, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ be proper, and $[\omega_{\boldsymbol{X}/Z}]$ a family of oriented -2–shifted symplectic structures on $\boldsymbol{X}/Z$, with $\mathrm{vdim}_{\mathbb{C}}\, \boldsymbol{X}/Z = n$. For each $z \in Z_{\mathrm{an}}$ we have a proper, oriented -2–shifted symplectic $\mathbb{C}$–scheme $(\boldsymbol{X}^z, \omega_{\boldsymbol{X}^z}^*)$ with $\mathrm{vdim}\, \boldsymbol{X}^z = n$. Then $[\boldsymbol{X}_{\mathrm{dm}}^{z_1}]_{\mathrm{dbo}} = [\boldsymbol{X}_{\mathrm{dm}}^{z_2}]_{\mathrm{dbo}}$ and $\iota_*^{z_1}([\boldsymbol{X}_{\mathrm{dm}}^{z_1}]_{\mathrm{virt}}) = \iota_*^{z_2}([\boldsymbol{X}_{\mathrm{dm}}^{z_2}]_{\mathrm{virt}})$ for all $z_1, z_2 \in Z_{\mathrm{an}}$, with $\iota_*^{z}([\boldsymbol{X}_{\mathrm{dm}}^{z}]_{\mathrm{virt}}) \in H_n(X_{\mathrm{an}}; \mathbb{Z})$ the pushforward under the inclusion $\iota^z\colon X_{\mathrm{an}}^z \hookrightarrow X_{\mathrm{an}}$.

So, *proper, oriented -2–shifted symplectic derived $\mathbb{C}$–schemes $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ have virtual classes*. This is not obvious; in fact it is rather surprising. Firstly, if $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ is -2–shifted symplectic then $X = t_0(\boldsymbol{X})$ has a natural obstruction theory $\mathbb{L}_{\boldsymbol{X}}|_X \to \mathbb{L}_X$ in the sense of Behrend and Fantechi [1], which is perfect in the interval $[-2, 0]$. But the Behrend–Fantechi construction of virtual cycles [1] works only for obstruction theories perfect in $[-1, 0]$, and does not apply here.

Secondly, our virtual cycle has real dimension $\mathrm{vdim}_{\mathbb{C}}\, \boldsymbol{X} = \frac{1}{2}\, \mathrm{vdim}_{\mathbb{R}}\, \boldsymbol{X}$, which is half what we might have expected. A heuristic explanation is that one should be able to make X into a "derived C^∞–scheme" X^{C^∞} (not a derived manifold), in some sense similar to Lurie [27, Section 4.5] or Spivak [32], and $(X^{C^\infty}, \operatorname{Im}\omega_X^*)$ should be a "real -2–shifted symplectic derived C^∞–scheme", with $\operatorname{Im}\omega_X^*$ the imaginary part of ω_X^*. There should be a morphism $X^{C^\infty} \to X_{\mathrm{dm}}$ which is a "Lagrangian fibration" of $(X^{C^\infty}, \operatorname{Im}\omega_X^*)$. So $\operatorname{vdim}_{\mathbb{R}} X_{\mathrm{dm}} = \frac{1}{2}\operatorname{vdim}_{\mathbb{R}} X^{C^\infty} = \frac{1}{2}\operatorname{vdim}_{\mathbb{R}} X$, as for Lagrangian fibrations $\pi\colon (S,\omega) \to B$ we have $\dim B = \frac{1}{2}\dim S$.

The main application that we intend for these results, motivated by Donaldson and Thomas [13] and explained in Sections 3.8–3.9, is to define new invariants "counting" (semi)stable coherent sheaves on Calabi–Yau 4–folds Y over $\mathbb{C}$, which should be unchanged under deformations of Y. These are similar to Donaldson–Thomas invariants found in Joyce and Song [25], Kontsevich and Soibelman [26] and Thomas [33] and could be called "holomorphic Donaldson invariants", as they are complex analogues of Donaldson invariants of 4–manifolds; see Donaldson and Kronheimer [12].

Pantev, Toën, Vaquié and Vezzosi [31, Section 2.1] show that any derived moduli stack $\boldsymbol{\mathcal{M}}$ of coherent sheaves (or complexes of coherent sheaves) on a Calabi–Yau m–fold has a $(2-m)$–shifted symplectic structure $\omega_{\boldsymbol{\mathcal{M}}}^*$, so in particular 4–Calabi–Yau moduli stacks are -2–shifted symplectic. Given an analogue of this for derived moduli schemes, and a way to define orientations upon them, Corollary 1.2 would give virtual classes for moduli schemes of (semi)stable coherent sheaves on Calabi–Yau 4–folds, and so enable us to define invariants.

It is well known that there is a great deal of interesting and special geometry, related to string theory, concerning Calabi–Yau 3–folds and 3–Calabi–Yau categories: mirror symmetry, Donaldson–Thomas theory, and so on. One message of this paper is that there should also be special geometry concerning Calabi–Yau 4–folds and 4–Calabi–Yau categories, which is not yet understood.

During the writing of this paper, Cao and Leung [8; 9; 10] also proposed a theory of invariants counting coherent sheaves on Calabi–Yau 4–folds, based on gauge theory rather than derived geometry. We discuss their work in Section 3.9.

Section 2 provides background material on derived schemes, shifted symplectic structures upon them, Kuranishi atlases, and derived manifolds. The heart of the paper is Section 3, with the definitions, main results, shorter proofs, and discussion. Longer proofs of results in Section 3 are deferred to Sections 4–6.

Acknowledgements We would like to thank Yalong Cao, Conan Leung, Bertrand Toën, Gabriele Vezzosi, and a referee for helpful conversations. This research was supported by EPSRC Programme Grant EP/I033343/1.

2 Background material

We begin with some background material and notation needed later. Some references are Toën and Vezzosi [34; 36] for Sections 2.1–2.2, Pantev, Toën, Vezzosi and Vaquié [31] and Brav, Bussi and Joyce [6] for Section 2.3, and Spivak [32], Borisov and Noël [3; 4] and Joyce [18; 19; 20; 21; 23; 24] for Section 2.6.

2.1 Commutative differential graded algebras

Definition 2.1 Write $\mathbf{cdga}_{\mathbb{C}}$ for the category of commutative differential graded $\mathbb{C}$–algebras in nonpositive degrees, and $\mathbf{cdga}_{\mathbb{C}}^{\mathrm{op}}$ for its opposite category. In fact $\mathbf{cdga}_{\mathbb{C}}$ has the additional structure of a model category (a kind of ∞–category), but we only use this in the proof of Theorem 3.1 in Section 4. In the rest of the paper we treat $\mathbf{cdga}_{\mathbb{C}}$, $\mathbf{cdga}_{\mathbb{C}}^{\mathrm{op}}$ just as ordinary categories.

Objects of $\mathbf{cdga}_{\mathbb{C}}$ are of the form $\cdots \to A^{-2} \xrightarrow{\mathrm{d}} A^{-1} \xrightarrow{\mathrm{d}} A^0$. Here A^k for $k = 0, -1, -2, \dots$ is the $\mathbb{C}$–vector space of degree-k elements of A, and we have a $\mathbb{C}$–bilinear, associative, supercommutative multiplication $A^k \times A^l \dot{\to} A^{k+l}$ for $k, l \leqslant 0$, an identity $1 \in A^0$, and differentials $\mathrm{d}\colon A^k \to A^{k+1}$ for $k < 0$ satisfying

$$\mathrm{d}(a \cdot b) = (\mathrm{d}a) \cdot b + (-1)^k a \cdot (\mathrm{d}b)$$

for all $a \in A^k$, $b \in A^l$. We write such objects as $A^\bullet$ or (A^*, d).

Here and throughout we will use the superscript “ $*$ ” to denote *graded* objects (eg graded algebras or vector spaces), where $*$ stands for an index in $\mathbb{Z}$, so that A^* means $(A^k : k \in \mathbb{Z})$. We will use the superscript “ $\bullet$ ” to denote *differential graded* objects (eg differential graded algebras or complexes), so that $A^\bullet$ means (A^*, d), the graded object A^* together with the differential d.

Morphisms $\alpha\colon A^\bullet \to B^\bullet$ in $\mathbf{cdga}_{\mathbb{C}}$ are $\mathbb{C}$–linear maps $\alpha^k\colon A^k \to B^k$ for all $k \leqslant 0$ commuting with all the structures on $A^\bullet$, $B^\bullet$.

A morphism $\alpha\colon A^\bullet \to B^\bullet$ is a *quasi-isomorphism* if $H^k(\alpha)\colon H^k(A^\bullet) \to H^k(B^\bullet)$ is an isomorphism on cohomology groups for all $k \leqslant 0$. A fundamental principle of derived algebraic geometry is that $\mathbf{cdga}_{\mathbb{C}}$ is not really the right category to work in, but instead one wants to define a new category (or better, ∞–category) by inverting (localizing) quasi-isomorphisms in $\mathbf{cdga}_{\mathbb{C}}$.

We will call $A^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ of *standard form* if A^0 is a smooth finitely generated $\mathbb{C}$–algebra of pure dimension, and the graded $\mathbb{C}$–algebra A^* is freely generated over A^0 by finitely many generators in each degree $i = -1, -2, \dots$. Here we require A^0 to be smooth *of pure dimension* so that $(\operatorname{Spec} A^0)_{\mathrm{an}}$ is a complex manifold, rather than a disjoint union of complex manifolds of different dimensions. This is not crucial, but will be convenient in Section 3.

Remark 2.2 Brav, Bussi and Joyce [6, Definition 2.9] work with a stronger notion of standard form cdgas than us, as they require A^* to be freely generated over A^0 by finitely many generators, all in negative degrees. In contrast, we allow infinitely many generators, but only finitely many in each degree $i = -1, -2, \dots$.

The important thing for us is that since standard form cdgas in the sense of [6] are also standard form in the (slightly weaker) sense of this paper, we can apply some of their results [6, Theorems 4.1, 4.2, 5.18] on the existence and properties of nice standard form cdga local models for derived schemes.

Definition 2.3 Let $A^\bullet \in \mathbf{cdga}_{\mathbb{C}}$, and write $D(\operatorname{mod} A)$ for the derived category of dg-modules over $A^\bullet$. Define a *derivation of degree* k from $A^\bullet$ to an $A^\bullet$–module $M^\bullet$ to be a $\mathbb{C}$–linear map $\delta\colon A^\bullet \to M^\bullet$ that is homogeneous of degree k with

$$\delta(fg) = \delta(f)g + (-1)^{k|f|} f\delta(g).$$

Just as for ordinary commutative algebras, there is a universal derivation into an $A^\bullet$–module of *Kähler differentials* $\Omega^1_{A^\bullet}$, which can be constructed as I/I^2 for $I = \operatorname{Ker}(m\colon A^\bullet \otimes A^\bullet \to A^\bullet)$. The universal derivation $\delta\colon A^\bullet \to \Omega^1_{A^\bullet}$ is given by $\delta(a) = a \otimes 1 - 1 \otimes a \in I/I^2$. One checks that δ is a universal degree-0 derivation, so that $\circ\,\delta\colon \operatorname{Hom}^\bullet_{A^\bullet}(\Omega^1_{A^\bullet}, M^\bullet) \to \operatorname{Der}^\bullet(A, M^\bullet)$ is an isomorphism of dg-modules.

Note that $\Omega^1_{A^\bullet} = ((\Omega^1_{A^\bullet})^*, \mathrm{d})$ is canonical up to strict isomorphism, not just up to quasi-isomorphism of complexes, or up to equivalence in $D(\operatorname{mod} A)$. Also, the underlying graded vector space $(\Omega^1_{A^\bullet})^*$, as a module over the graded algebra A^*, depends only on A^* and not on the differential d in $A^\bullet = (A^*, \mathrm{d})$.

Similarly, given a morphism of cdgas $\Phi\colon A^\bullet \to B^\bullet$, we can define the *relative Kähler differentials* $\Omega^1_{B^\bullet/A^\bullet}$.

The *cotangent complex* $\mathbb{L}_{A^\bullet}$ of $A^\bullet$ is related to the Kähler differentials $\Omega^1_{A^\bullet}$, but is not quite the same. If $\Phi\colon A^\bullet \to B^\bullet$ is a quasi-isomorphism of cdgas over $\mathbb{C}$, then $\Phi_*\colon \Omega^1_{A^\bullet} \otimes_{A^\bullet} B^\bullet \to \Omega^1_{B^\bullet}$ may not be a quasi-isomorphism of $B^\bullet$–modules. So Kähler differentials are not well behaved under localizing quasi-isomorphisms of cdgas, which is bad for doing derived algebraic geometry.

The cotangent complex $\mathbb{L}_{A^\bullet}$ is a substitute for $\Omega^1_{A^\bullet}$ which is well behaved under localizing quasi-isomorphisms. It is an object in $D(\operatorname{mod} A)$, canonical up to equivalence. We can define it by replacing $A^\bullet$ by a quasi-isomorphic, cofibrant (in the sense of model categories) cdga $B^\bullet$, and then setting $\mathbb{L}_{A^\bullet} = (\Omega^1_{B^\bullet}) \otimes_{B^\bullet} A^\bullet$. We will be interested in the p^{th} exterior power $\Lambda^p\mathbb{L}_{A^\bullet}$, and the dual $(\mathbb{L}_{A^\bullet})^\vee$, which is called the *tangent complex*, and written $\mathbb{T}_{A^\bullet}=(\mathbb{L}_{A^\bullet})^\vee$.

There is a *de Rham differential* $\mathrm{d}_{\mathrm{dR}}\colon \Lambda^p\mathbb{L}_{A^\bullet}\to\Lambda^{p+1}\mathbb{L}_{A^\bullet}$, a morphism of complexes, with ${\mathrm{d}_{\mathrm{dR}}}^2=0\colon \Lambda^p\mathbb{L}_{A^\bullet}\to\Lambda^{p+2}\mathbb{L}_{A^\bullet}$. Note that each $\Lambda^p\mathbb{L}_{A^\bullet}$ is also a complex with its own internal differential $\mathrm{d}\colon (\Lambda^p\mathbb{L}_{A^\bullet})^k\to(\Lambda^p\mathbb{L}_{A^\bullet})^{k+1}$, and d_{dR} being a morphism of complexes means that $\mathrm{d}\circ\mathrm{d}_{\mathrm{dR}}=\mathrm{d}_{\mathrm{dR}}\circ\mathrm{d}$.

Similarly, given a morphism of cdgas $\Phi\colon A^\bullet\to B^\bullet$, we can define the *relative cotangent complex* $\mathbb{L}_{B^\bullet/A^\bullet}$.

As in [6, Section 2.3], an important property of our standard form cdgas $A^\bullet$ in Definition 2.1 is that they are sufficiently cofibrant that the Kähler differentials $\Omega^1_{A^\bullet}$ provide a model for the cotangent complex $\mathbb{L}_{A^\bullet}$, so we can take $\Omega^1_{A^\bullet}=\mathbb{L}_{A^\bullet}$, without having to replace $A^\bullet$ by an unknown cdga $B^\bullet$. Thus standard form cdgas are convenient for doing explicit computations with cotangent complexes.

A morphism $\Phi\colon A^\bullet\to B^\bullet$ of cdgas will be called *quasifree* if $\Phi^0\colon A^0\to B^0$ is a smooth morphism of $\mathbb{C}$–algebras of pure relative dimension, and as a graded $(A^*\otimes_{A^0}B^0)$–algebra B^* is free and finitely generated in each degree. Here if $A^\bullet$ is of standard form and Φ is quasifree then $B^\bullet$ is of standard form, and a cdga $A^\bullet$ is of standard form if and only if the unique morphism $\mathbb{C}\to A^\bullet$ is quasifree. We will only consider quasifree morphisms when $A^\bullet$, $B^\bullet$ are of standard form.

If $\Phi\colon A^\bullet\to B^\bullet$ is a quasifree morphism then the relative Kähler differentials $\Omega^1_{B^\bullet/A^\bullet}$ are a model for the relative cotangent complex $\mathbb{L}_{B^\bullet/A^\bullet}$, and therefore we can take $\Omega^1_{B^\bullet/A^\bullet}=\mathbb{L}_{B^\bullet/A^\bullet}$. Thus quasifree morphisms are a convenient class of morphisms for doing explicit computations with cotangent complexes.

2.2 Derived algebraic geometry and derived schemes

Definition 2.4 Write $\mathbf{dSt}_{\mathbb{C}}$ for the ∞–category of *derived* $\mathbb{C}$*–stacks* (or D^-*–stacks*) defined by Toën and Vezzosi [36, Definition 2.2.2.14; 34, Definition 4.2]. Objects $\boldsymbol{X}$ in $\mathbf{dSt}_{\mathbb{C}}$ are ∞–functors

$$\boldsymbol{X}\colon \{\text{simplicial commutative } \mathbb{C}\text{–algebras}\}\to\{\text{simplicial sets}\}$$

satisfying sheaf-type conditions. There is a *spectrum functor*

$$\mathbf{Spec}\colon \mathbf{cdga}_{\mathbb{C}}^{\mathrm{op}}\to\mathbf{dSt}_{\mathbb{C}}.$$

A derived $\mathbb{C}$–stack $\boldsymbol{X}$ is called an *affine derived* $\mathbb{C}$*–scheme* if $\boldsymbol{X}$ is equivalent in $\mathbf{dSt}_{\mathbb{C}}$ to $\mathbf{Spec}\,A^\bullet$ for some cdga $A^\bullet$ over $\mathbb{C}$. As in [34, Section 4.2], a derived $\mathbb{C}$–stack $\boldsymbol{X}$ is called a *derived* $\mathbb{C}$*–scheme* if it may be covered by Zariski open $\boldsymbol{Y}\subseteq\boldsymbol{X}$ with $\boldsymbol{Y}$

an affine derived $\mathbb{C}$–scheme. Write $\mathbf{dSch}_{\mathbb{C}}$ for the full ∞–subcategory of derived $\mathbb{C}$–schemes in $\mathbf{dSt}_{\mathbb{C}}$, and $\mathbf{dSch}_{\mathbb{C}}^{\mathrm{aff}} \subset \mathbf{dSch}_{\mathbb{C}}$ for the full ∞–subcategory of affine derived $\mathbb{C}$–schemes. See also Toën [35] for a different but equivalent way to define derived $\mathbb{C}$–schemes, as an ∞–category of derived ringed spaces.

We shall assume throughout this paper that all derived $\mathbb{C}$–schemes $\boldsymbol{X}$ are *locally finitely presented* in the sense of Toën and Vezzosi [36, Definition 1.3.6.4]. Note that this is a strong condition, for instance it implies that the cotangent complex $\mathbb{L}_{\boldsymbol{X}}$ is perfect [36, Proposition 2.2.2.4]. A locally finitely presented classical $\mathbb{C}$–scheme X need not be locally finitely presented as a derived $\mathbb{C}$–scheme. A local normal form for locally finitely presented derived $\mathbb{C}$–schemes is given in [6, Theorem 4.1].

There is a *classical truncation functor* $t_0\colon \mathbf{dSch}_{\mathbb{C}} \to \mathbf{Sch}_{\mathbb{C}}$ taking a derived $\mathbb{C}$–scheme $\boldsymbol{X}$ to the underlying classical $\mathbb{C}$–scheme $X = t_0(\boldsymbol{X})$. On affine derived schemes $\mathbf{dSch}_{\mathbb{C}}^{\mathrm{aff}}$ the functor t_0 maps $\mathbf{Spec}\, A^\bullet \to \operatorname{Spec} H^0(A^\bullet) = \operatorname{Spec}(A^0/\mathrm{d}(A^{-1}))$.

Toën and Vezzosi show that a derived $\mathbb{C}$–scheme $\boldsymbol{X}$ has a *cotangent complex* $\mathbb{L}_{\boldsymbol{X}}$ [36, Section 1.4; 34, Sections 4.2.4–4.2.5] in a stable ∞–category $L_{\mathrm{qcoh}}(\boldsymbol{X})$ defined in [34, Section 3.1.7, Section 4.2.4]. We will be interested in the p^{th} exterior power $\Lambda^p\mathbb{L}_{\boldsymbol{X}}$, and the dual $(\mathbb{L}_{\boldsymbol{X}})^\vee$, which is called the *tangent complex* $\mathbb{T}_{\boldsymbol{X}}$. There is a *de Rham differential* $\mathrm{d}_{\mathrm{dR}}\colon \Lambda^p\mathbb{L}_{\boldsymbol{X}} \to \Lambda^{p+1}\mathbb{L}_{\boldsymbol{X}}$.

Restricted to the classical scheme $X = t_0(\boldsymbol{X})$, the cotangent complex $\mathbb{L}_{\boldsymbol{X}}|_X$ may Zariski locally be modelled as a finite complex of vector bundles

$$[F^{-m} \to F^{1-m} \to \cdots \to F^0]$$

on X in degrees $[-m,0]$ for some $m \geqslant 0$. The (*complex*) *virtual dimension* $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X}$ is $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = \sum_{i=0}^m (-1)^i \operatorname{rank} F^{-i}$. It is a locally constant function $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X}\colon X \to \mathbb{Z}$, so is constant on each connected component of X. We say that $\boldsymbol{X}$ *has* (*complex*) *virtual dimension* $n \in \mathbb{Z}$ if $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n$.

When $\boldsymbol{X} = X$ is a classical scheme, the homotopy category of $L_{\mathrm{qcoh}}(\boldsymbol{X})$ is the triangulated category $D_{\mathrm{qcoh}}(X)$ of complexes of quasicoherent sheaves. These $\mathbb{L}_{\boldsymbol{X}}$, $\mathbb{T}_{\boldsymbol{X}}$ have the usual properties of (co)tangent complexes. For instance, if $\boldsymbol{f}\colon \boldsymbol{X} \to \boldsymbol{Y}$ is a morphism in $\mathbf{dSch}_{\mathbb{C}}$ there is a distinguished triangle

$$\boldsymbol{f}^*(\mathbb{L}_{\boldsymbol{Y}}) \xrightarrow{\mathbb{L}_{\boldsymbol{f}}} \mathbb{L}_{\boldsymbol{X}} \longrightarrow \mathbb{L}_{\boldsymbol{X}/\boldsymbol{Y}} \longrightarrow \boldsymbol{f}^*(\mathbb{L}_{\boldsymbol{Y}})[1],$$

where $\mathbb{L}_{\boldsymbol{X}/\boldsymbol{Y}}$ is the *relative cotangent complex* of $\boldsymbol{f}$.

Now suppose $A^\bullet$ is a cdga over $\mathbb{C}$, and $\boldsymbol{X}$ a derived $\mathbb{C}$–scheme with $\boldsymbol{X} \simeq \mathbf{Spec}\, A^\bullet$ in $\mathbf{dSch}_{\mathbb{C}}$. Then we have an equivalence of triangulated categories $L_{\mathrm{qcoh}}(\boldsymbol{X}) \simeq D(\operatorname{mod} A)$, which identifies cotangent complexes $\mathbb{L}_{\boldsymbol{X}} \simeq \mathbb{L}_{A^\bullet}$. If also $A^\bullet$ is of standard form then $\mathbb{L}_{A^\bullet} \simeq \Omega^1_{A^\bullet}$, so $\mathbb{L}_{\boldsymbol{X}} \simeq \Omega^1_{A^\bullet}$.

Bussi, Brav and Joyce [6, Theorem 4.1] prove:

Theorem 2.5 *Suppose $\boldsymbol{X}$ is a derived $\mathbb{C}$–scheme (as always, assumed locally finitely presented), and $x \in \boldsymbol{X}$. Then there exists a standard form cdga $A^\bullet$ over $\mathbb{C}$ and a Zariski open inclusion $\boldsymbol{\alpha}\colon \mathbf{Spec}\, A^\bullet \hookrightarrow \boldsymbol{X}$ with $x \in \operatorname{Im} \boldsymbol{\alpha}$.*

See Remark 2.2 on the difference in definitions of "standard form". Bussi et al also explain [6, Theorem 4.2] how to compare two such standard form charts $\mathbf{Spec}\, A^\bullet \hookrightarrow \boldsymbol{X}$, $\mathbf{Spec}\, B^\bullet \hookrightarrow \boldsymbol{X}$ on their overlap in $\boldsymbol{X}$, using a third chart. We will need the following conditions on derived $\mathbb{C}$–schemes and their morphisms.

Definition 2.6 A derived $\mathbb{C}$–scheme $\boldsymbol{X}$ is called *separated*, or *proper*, or *quasicompact*, if the classical $\mathbb{C}$–scheme $X = t_0(\boldsymbol{X})$ is separated, or proper, or quasicompact, respectively, in the classical sense, as in Hartshorne [16, pages 80, 96, 100]. Proper implies separated. A morphism of derived schemes $\boldsymbol{f}\colon \boldsymbol{X} \to \boldsymbol{Y}$ is *proper* if $t_0(\boldsymbol{f})\colon t_0(\boldsymbol{X}) \to t_0(\boldsymbol{Y})$ is proper in the classical sense [16, page 100].

We will need the following nontrivial fact about the relation between classical and derived $\mathbb{C}$–schemes. As in Toën [35, Section 2.2, page 186], a derived $\mathbb{C}$–scheme $\boldsymbol{X}$ is affine if and only if the classical $\mathbb{C}$–scheme $X = t_0(\boldsymbol{X})$ is affine.

Recall that a morphism $\alpha\colon X \to Y$ in $\mathbf{Sch}_{\mathbb{C}}$ (or $\boldsymbol{\alpha}\colon \boldsymbol{X} \to \boldsymbol{Y}$ in $\mathbf{dSch}_{\mathbb{C}}$) is *affine* if whenever $\beta\colon U \to Y$ is a Zariski open inclusion with U affine (or $\boldsymbol{\beta}\colon \boldsymbol{U} \to \boldsymbol{Y}$ is Zariski open with $\boldsymbol{U}$ affine), the fibre product $X \times_{\alpha,Y,\beta} U$ in $\mathbf{Sch}_{\mathbb{C}}$ (or homotopy fibre product $\boldsymbol{X} \times^h_{\boldsymbol{\alpha},\boldsymbol{Y},\boldsymbol{\beta}} \boldsymbol{U}$ in $\mathbf{dSch}_{\mathbb{C}}$) is also affine. Since $\boldsymbol{X}$ is affine if and only if $X = t_0(\boldsymbol{X})$ is affine, we see that a morphism $\boldsymbol{\alpha}\colon \boldsymbol{X} \to \boldsymbol{Y}$ in $\mathbf{dSch}_{\mathbb{C}}$ is affine if and only if $t_0(\boldsymbol{\alpha})\colon t_0(\boldsymbol{X}) \to t_0(\boldsymbol{Y})$ is affine.

Now let $\boldsymbol{X}$ be a separated derived $\mathbb{C}$–scheme. Then $X = t_0(\boldsymbol{X})$ is a separated classical $\mathbb{C}$–scheme, so [16, page 96] the diagonal morphism $\Delta_X\colon X \to X \times X$ is a closed immersion. But closed immersions are affine, and $\Delta_X = t_0(\Delta_{\boldsymbol{X}})$ for $\Delta_{\boldsymbol{X}}\colon \boldsymbol{X} \to \boldsymbol{X} \times \boldsymbol{X}$ the derived diagonal morphism, so $\Delta_{\boldsymbol{X}}$ is also affine. That is, $\boldsymbol{X}$ has *affine diagonal*. Therefore if $\boldsymbol{U}_1, \boldsymbol{U}_2 \hookrightarrow \boldsymbol{X}$ are Zariski open inclusions with $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ affine, then $\boldsymbol{U}_1 \times^h_{\boldsymbol{X}} \boldsymbol{U}_2 \hookrightarrow \boldsymbol{X}$ is also Zariski open with $\boldsymbol{U}_1 \times^h_{\boldsymbol{X}} \boldsymbol{U}_2$ affine. Thus, *finite intersections of open affine derived $\mathbb{C}$–subschemes in a separated derived $\mathbb{C}$–scheme $\boldsymbol{X}$ are affine.*

2.3 The shifted symplectic geometry of Pantev, Toën, Vaquié and Vezzosi

Next we summarize parts of the theory of shifted symplectic geometry, as developed by Pantev, Toën, Vaquié and Vezzosi in [31]. We explain them for derived $\mathbb{C}$–schemes $\boldsymbol{X}$, although Pantev et al work more generally with derived stacks.

Given a (locally finitely presented) derived $\mathbb{C}$–scheme $\boldsymbol{X}$ and given $p \geqslant 0$, $k \in \mathbb{Z}$, Pantev et al [31] define complexes of *k–shifted p–forms* $\mathcal{A}^p_{\mathbb{C}}(\boldsymbol{X}, k)$ and *k–shifted closed p–forms* $\mathcal{A}^{p,\mathrm{cl}}_{\mathbb{C}}(\boldsymbol{X}, k)$. These are defined first for affine derived $\mathbb{C}$–schemes $\boldsymbol{Y} = \mathbf{Spec}\, A^\bullet$ for $A^\bullet$ a cdga over $\mathbb{C}$, and shown to satisfy étale descent. Then for general $\boldsymbol{X}$, k–shifted (closed) p–forms are defined as a mapping stack; basically, a k–shifted (closed) p–form ω on $\boldsymbol{X}$ is the functorial choice for all $\boldsymbol{Y}$, $\boldsymbol{f}$ of a k–shifted (closed) p–form $\boldsymbol{f}^*(\omega)$ on $\boldsymbol{Y}$ whenever $\boldsymbol{Y} = \mathbf{Spec}\, A^\bullet$ is affine and $\boldsymbol{f}\colon \boldsymbol{Y} \to \boldsymbol{X}$ is a morphism.

Definition 2.7 Let $\boldsymbol{Y} \simeq \mathbf{Spec}\, A^\bullet$ be an affine derived $\mathbb{C}$–scheme, for $A^\bullet$ a cdga over $\mathbb{C}$. A *k–shifted p–form* on $\boldsymbol{Y}$ for $k \in \mathbb{Z}$ is an element $\omega_{A^\bullet} \in (\Lambda^p \mathbb{L}_{A^\bullet})^k$ with $\mathrm{d}\omega_{A^\bullet} = 0$ in $(\Lambda^p \mathbb{L}_{A^\bullet})^{k+1}$, so that $\omega_{A^\bullet}$ defines a cohomology class $[\omega_{A^\bullet}] \in H^k(\Lambda^p \mathbb{L}_{A^\bullet})$. When $p = 2$, we call $\omega_{A^\bullet}$ *nondegenerate*, or a *k–shifted presymplectic form*, if the induced morphism $\mathbb{T}_{A^\bullet} \xrightarrow{\omega_{A^\bullet}\cdot} \mathbb{L}_{A^\bullet}[k]$ is a quasi-isomorphism.

A *k–shifted closed p–form* on $\boldsymbol{Y}$ is a sequence $\omega^*_{A^\bullet} = (\omega^0_{A^\bullet}, \omega^1_{A^\bullet}, \omega^2_{A^\bullet}, \dots)$ such that $\omega^m_{A^\bullet} \in (\Lambda^{p+m} \mathbb{L}_{A^\bullet})^{k-m}$ for $m \geqslant 0$, with $\mathrm{d}\omega^0_{A^\bullet} = 0$ and $\mathrm{d}\omega^{1+m}_{A^\bullet} + \mathrm{d}_{\mathrm{dR}}\omega^m_{A^\bullet} = 0$ in $(\Lambda^{p+m+1} \mathbb{L}_{A^\bullet})^{k-m}$ for all $m \geqslant 0$. Note that if $\omega^*_{A^\bullet} = (\omega^0_{A^\bullet}, \omega^1_{A^\bullet}, \dots)$ is a k–shifted closed p–form then $\omega^0_{A^\bullet}$ is a k–shifted p–form.

When $p = 2$, we call a k–shifted closed 2–form $\omega^*_{A^\bullet}$ a *k–shifted symplectic form* if the associated 2–form $\omega^0_{A^\bullet}$ is nondegenerate (presymplectic).

If $\boldsymbol{X}$ is a general derived $\mathbb{C}$–scheme, then Pantev et al [31, Section 1.2] define *k–shifted 2–forms* $\omega_{\boldsymbol{X}}$, which may be *nondegenerate* (*presymplectic*), and *k–shifted closed 2–forms* $\omega^*_{\boldsymbol{X}}$, which have an associated k–shifted 2–form $\omega^0_{\boldsymbol{X}}$, and where $\omega^*_{\boldsymbol{X}}$ is called a *k–shifted symplectic form* if $\omega^0_{\boldsymbol{X}}$ is nondegenerate (presymplectic). We will not go into the details of this definition for general $\boldsymbol{X}$.

The important thing for us is this: if $\boldsymbol{Y} \subseteq \boldsymbol{X}$ is a Zariski open affine derived $\mathbb{C}$–subscheme with $\boldsymbol{Y} \simeq \mathbf{Spec}\, A^\bullet$ then a k–shifted 2–form $\omega_{\boldsymbol{X}}$ (or a k–shifted closed 2–form $\omega^*_{\boldsymbol{X}}$) on $\boldsymbol{X}$ induces a k–shifted 2–form $\omega_{A^\bullet}$ (or a k–shifted closed 2–form $\omega^*_{A^\bullet}$) on $\boldsymbol{Y}$ in the sense above, where $\omega_{A^\bullet}$ is unique up to cohomology in the complex $((\Lambda^2 \mathbb{L}_{A^\bullet})^*, \mathrm{d})$, and $\omega_{\boldsymbol{X}}$ nondegenerate/presymplectic implies $\omega_{A^\bullet}$ nondegenerate/presymplectic (or where $\omega^*_{A^\bullet}$ is unique up to cohomology in the complex $\bigl(\prod_{m \geqslant 0} (\Lambda^{2+m} \mathbb{L}_{A^\bullet})^{*-m}, \mathrm{d} + \mathrm{d}_{\mathrm{dR}}\bigr)$, and $\omega^*_{\boldsymbol{X}}$ symplectic implies $\omega^*_{A^\bullet}$ symplectic).

It is easy to show that if $\boldsymbol{X}$ is a derived $\mathbb{C}$–scheme with a k–shifted symplectic or presymplectic form, then $k \leqslant 0$, and the complex virtual dimension $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X}$ satisfies $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X} = 0$ if k is odd, and $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X}$ is even if $k \equiv 0 \bmod 4$ (which includes classical complex symplectic schemes when $k = 0$), and $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X} \in \mathbb{Z}$ if $k \equiv 2 \bmod 4$. In particular, in the case $k = -2$ of interest in this paper, $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X}$ can take any value in $\mathbb{Z}$.

The main examples we have in mind come from Pantev et al [31, Section 2.1]:

Theorem 2.8 *Suppose Y is a Calabi–Yau m–fold over $\mathbb{C}$, and $\boldsymbol{\mathcal{M}}$ a derived moduli stack of coherent sheaves (or complexes of coherent sheaves) on Y. Then $\boldsymbol{\mathcal{M}}$ has a natural $(2-m)$–shifted symplectic form $\omega_{\boldsymbol{\mathcal{M}}}$.*

In particular, derived moduli schemes and stacks on a Calabi–Yau 4–fold Y are -2–shifted symplectic.

Bussi, Brav and Joyce [6] prove "Darboux theorems" for k–shifted symplectic derived $\mathbb{C}$–schemes $(\boldsymbol{X},\omega_{\boldsymbol{X}})$ for $k<0$, which give explicit Zariski local models for $(\boldsymbol{X},\omega_{\boldsymbol{X}})$. We will explain their main result for $k=-2$. The next definition is taken from [6, Example 5.16] (with notation changed, $2q_js_j$ in place of s_j).

Definition 2.9 A pair $(A^\bullet,\omega_{A^\bullet})$ is called in -2–*Darboux form* if $A^\bullet$ is a standard form cdga over $\mathbb{C}$, and $\omega_{A^\bullet}\in(\Lambda^2\mathbb{L}_{A^\bullet})^{-2}=(\Lambda^2\Omega^1_{A^\bullet})^{-2}$ with $\mathrm{d}\omega_{A^\bullet}=0$ in $(\Lambda^2\mathbb{L}_{A^\bullet})^{-1}$ and $\mathrm{d}_{\mathrm{dR}}\omega_{A^\bullet}=0$ in $(\Lambda^3\mathbb{L}_{A^\bullet})^{-2}$, so that $\omega^*_{A^\bullet}:=(\omega_{A^\bullet},0,0,\dots)$ is a -2–shifted closed 2–form on $A^\bullet$, such that:

(i) A^0 is a smooth $\mathbb{C}$–algebra of dimension m, and there exist $x_1,\dots,x_m$ in A^0 forming an étale coordinate system on $V=\operatorname{Spec}A^0$.

(ii) The commutative graded algebra A^* is freely generated over A^0 by elements $y_1,\dots,y_n$ of degree -1 and $z_1,\dots,z_m$ of degree -2.

(iii) There are invertible elements $q_1,\dots,q_n$ in A^0 such that

$$\omega_{A^\bullet}=\mathrm{d}_{\mathrm{dR}}z_1\,\mathrm{d}_{\mathrm{dR}}x_1+\cdots+\mathrm{d}_{\mathrm{dR}}z_m\,\mathrm{d}_{\mathrm{dR}}x_m+\mathrm{d}_{\mathrm{dR}}(q_1y_1)\,\mathrm{d}_{\mathrm{dR}}y_1+\cdots+\mathrm{d}_{\mathrm{dR}}(q_ny_n)\,\mathrm{d}_{\mathrm{dR}}y_n. \tag{1}$$

(iv) There are elements $s_1,\dots,s_n\in A^0$ satisfying

$$q_1(s_1)^2+\cdots+q_n(s_n)^2=0\quad\text{in } A^0, \tag{2}$$

such that the differential d on $A^\bullet=(A^*,\mathrm{d})$ is given by

$$\mathrm{d}x_i=0,\quad \mathrm{d}y_j=s_j,\quad \mathrm{d}z_i=\sum_{j=1}^n y_j\left(2q_j\frac{\partial s_j}{\partial x_i}+s_j\frac{\partial q_j}{\partial x_i}\right). \tag{3}$$

Here the only assumptions are that A^0, $x_1,\dots,x_m$ are as in (i) and we are given $q_1,\dots,q_n$, $s_1,\dots,s_n$ in A^0 satisfying (2), and everything else follows from these. Defining A^* as in (ii) and d as in (3), then $A^\bullet=(A^*,\mathrm{d})$ is a standard form cdga over $\mathbb{C}$, where to show that $\mathrm{d}\circ\mathrm{d}z_i=0$ we apply $\partial/\partial x_i$ to (2). Clearly $\mathrm{d}_{\mathrm{dR}}\omega_{A^\bullet}=0$, as $\mathrm{d}_{\mathrm{dR}}\circ\mathrm{d}_{\mathrm{dR}}=0$. We have

$$\begin{aligned}
\mathrm{d}\omega_{A^\bullet} &= \sum_{i=1}^m (\mathrm{d}\circ \mathrm{d}_{\mathrm{dR}} z_i)\, \mathrm{d}_{\mathrm{dR}} x_i + \sum_{j=1}^n (\mathrm{d}\circ \mathrm{d}_{\mathrm{dR}}(q_j y_j))\, \mathrm{d}_{\mathrm{dR}} y_j + (\mathrm{d}\circ \mathrm{d}_{\mathrm{dR}} y_j)\, \mathrm{d}_{\mathrm{dR}}(q_j y_j) \\
&= -\mathrm{d}_{\mathrm{dR}} \sum_{i=1}^m \mathrm{d}z_i\, \mathrm{d}_{\mathrm{dR}} x_i - \mathrm{d}_{\mathrm{dR}} \sum_{j=1}^n [\mathrm{d}(q_j y_j)\, \mathrm{d}_{\mathrm{dR}} y_j + \mathrm{d}y_j\, \mathrm{d}_{\mathrm{dR}}(q_j y_j)] \\
&= -\mathrm{d}_{\mathrm{dR}} \sum_{i=1}^m \sum_{j=1}^n y_j \left(2q_j \frac{\partial s_j}{\partial x_i} + s_j \frac{\partial q_j}{\partial x_i}\right) \mathrm{d}_{\mathrm{dR}} x_i - \mathrm{d}_{\mathrm{dR}} \sum_{j=1}^n [q_j s_j\, \mathrm{d}_{\mathrm{dR}} y_j + s_j\, \mathrm{d}_{\mathrm{dR}}(q_j y_j)] \\
&= -\mathrm{d}_{\mathrm{dR}} \circ \mathrm{d}_{\mathrm{dR}} \sum_{j=1}^n [(q_j s_j) y_j + s_j (q_j y_j)] = 0,
\end{aligned}$$

using (1) and $\mathrm{d}\circ \mathrm{d}_{\mathrm{dR}} x_i = 0$ for degree reasons in the first step, $\mathrm{d}\circ \mathrm{d}_{\mathrm{dR}} = -\mathrm{d}_{\mathrm{dR}} \circ \mathrm{d}$ and $\mathrm{d}_{\mathrm{dR}} \circ \mathrm{d}_{\mathrm{dR}} = 0$ in the second, (3) in the third, $\mathrm{d}s_j = \sum_{i=1}^n (\partial s_j / \partial x_i)\, \mathrm{d}_{\mathrm{dR}} x_i$ and similarly for q_j in the fourth, and $\mathrm{d}_{\mathrm{dR}} \circ \mathrm{d}_{\mathrm{dR}} = 0$ in the fifth. Hence $\omega^*_{A^\bullet}$ is a -2–shifted closed 2–form on $A^\bullet$.

The action $\mathbb{T}_{A^\bullet} \xrightarrow{\omega_{A^\bullet}\cdot} \mathbb{L}_{A^\bullet}[-2]$ is given by

$$\omega_{A^\bullet} \cdot \frac{\partial}{\partial x_i} = -\mathrm{d}_{\mathrm{dR}} z_i + \sum_{j=1}^n \frac{\partial q_j}{\partial x_i} y_j\, \mathrm{d}_{\mathrm{dR}} y_j,$$

$$\omega_{A^\bullet} \cdot \frac{\partial}{\partial y_j} = 2q_j\, \mathrm{d}_{\mathrm{dR}} y_j - \sum_{i=1}^m y_j \frac{\partial q_j}{\partial x_i}\, \mathrm{d}_{\mathrm{dR}} x_i, \qquad \omega_{A^\bullet} \cdot \frac{\partial}{\partial z_i} = \mathrm{d}_{\mathrm{dR}} x_i.$$

By writing this as an upper triangular matrix with invertible diagonal (since the q_j are invertible), we see that $\omega_{A^\bullet}\cdot$ is actually an isomorphism of complexes, so a quasi-isomorphism, and $\omega^*_{A^\bullet}$ is a -2–shifted symplectic form on $A^\bullet$.

The main result of Bussi, Brav and Joyce [6, Theorem 5.18] when $k = -2$ yields:

Theorem 2.10 *Suppose $(\boldsymbol{X}, \omega^*_{\boldsymbol{X}})$ is a -2–shifted symplectic derived $\mathbb{C}$–scheme. Then for each $x \in X = t_0(\boldsymbol{X})$ there exists a pair $(A^\bullet, \omega_{A^\bullet})$ in -2–Darboux form and a Zariski open inclusion $\boldsymbol{\alpha}\colon \mathbf{Spec}\, A^\bullet \hookrightarrow \boldsymbol{X}$ such that $x \in \operatorname{Im} \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*(\omega^*_{\boldsymbol{X}}) \simeq \omega_{A^\bullet}$ in $\mathcal{A}^{2,\mathrm{cl}}_{\mathbb{C}}(\mathbf{Spec}\, A^\bullet, -2)$. Furthermore, we can choose $A^\bullet$* ***minimal*** *at x, in the sense that $m = \dim H^0(\mathbb{T}_{\boldsymbol{X}}|_x)$ and $n = \dim H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ in Definition 2.9.*

2.4 Orientations on k–shifted symplectic derived schemes

If $\boldsymbol{X}$ is a derived $\mathbb{C}$–scheme (always assumed locally finitely presented), with classical $\mathbb{C}$–scheme $X = t_0(\boldsymbol{X})$, the cotangent complex $\mathbb{L}_{\boldsymbol{X}}|_X$ restricted to X is a perfect complex, so it has a determinant line bundle $\det(\mathbb{L}_{\boldsymbol{X}}|_X)$ on X.

The following notion is important for -1–shifted symplectic derived schemes, 3–Calabi–Yau moduli spaces, and generalizations of Donaldson–Thomas theory:

Definition 2.11 Let $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ be a -1–shifted symplectic derived $\mathbb{C}$–scheme (or more generally k–shifted symplectic, for $k<0$ odd). An *orientation* for $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ is a choice of square root line bundle $\det(\mathbb{L}_{\boldsymbol{X}}|_X)^{1/2}$ for $\det(\mathbb{L}_{\boldsymbol{X}}|_X)$.

Writing $X_{\rm an}$ for the complex analytic topological space of X, the obstruction to existence of orientations for $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ lies in $H^2(X_{\rm an};\mathbb{Z}_2)$, and if the obstruction vanishes, the set of orientations is a torsor for $H^1(X_{\rm an};\mathbb{Z}_2)$.

This notion of orientation, and its analogue for "d-critical loci", are used by Ben-Bassat, Brav, Bussi, Dupont, Joyce, Meinhardt and Szendrői in a series of papers [2; 5; 6; 7; 22]. They use orientations on $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ to define natural perverse sheaves, $\mathcal{D}$–modules, mixed Hodge modules, and motives on X. A similar idea first appeared in Kontsevich and Soibelman [26, Section 5] as "orientation data" needed to define motivic Donaldson–Thomas invariants of Calabi–Yau 3–folds.

This paper concerns -2–shifted symplectic derived schemes, and 4–Calabi–Yau moduli spaces. It turns out that there is a parallel notion of orientation in the -2–shifted case, needed to construct virtual cycles.

To define this, note that determinant line bundles $\det(E^\bullet)$ of perfect complexes $\mathcal{E}^\bullet$ satisfy $\det[(E^\bullet)^\vee]\cong[\det(E^\bullet)]^{-1}$, and $\det(E^\bullet[k])\cong[\det(E^\bullet)]^{(-1)^k}$. If $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ is a k–shifted symplectic derived $\mathbb{C}$–scheme, then $\mathbb{T}_{\boldsymbol{X}}\simeq\mathbb{L}_{\boldsymbol{X}}[k]$, where $\mathbb{T}_{\boldsymbol{X}}\simeq(\mathbb{L}_{\boldsymbol{X}})^\vee$. Restricting to X and taking determinant line bundles gives $\det(\mathbb{L}_{\boldsymbol{X}}|_X)^{-1}\cong\det(\mathbb{L}_{\boldsymbol{X}}|_X)^{(-1)^k}$. If k is odd this is trivial, but for k even, this gives a canonical isomorphism of line bundles on X:

$$\iota_{\boldsymbol{X},\omega_{\boldsymbol{X}}^*}\colon [\det(\mathbb{L}_{\boldsymbol{X}}|_X)]^{\otimes^2}\to\mathcal{O}_X\cong\mathcal{O}_X^{\otimes^2}. \tag{4}$$

The next definition is new, so far as the authors know.

Definition 2.12 Let $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme (or more generally k–shifted symplectic, for $k<0$ with $k\equiv 2 \bmod 4$). An *orientation* for $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ is a choice of isomorphism $o\colon\det(\mathbb{L}_{\boldsymbol{X}}|_X)\to\mathcal{O}_X$ such that $o\otimes o=\iota_{\boldsymbol{X},\omega_{\boldsymbol{X}}^*}$, for $\iota_{\boldsymbol{X},\omega_{\boldsymbol{X}}^*}$ as in (4).

Writing $X_{\rm an}$ for the complex analytic topological space of X, the obstruction to existence of orientations for $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ lies in $H^1(X_{\rm an};\mathbb{Z}_2)$, and if the obstruction vanishes, the set of orientations is a torsor for $H^0(X_{\rm an};\mathbb{Z}_2)$.

This definition makes sense for k–shifted symplectic derived $\mathbb{C}$–schemes with k even, but when $k\equiv 0 \bmod 4$ (including the classical symplectic case $k=0$) there is a natural choice of orientation o, so we restrict to $k\equiv 2 \bmod 4$.

At a point $x \in X_{\mathrm{an}}$, we have a canonical isomorphism

$$\det(\mathbb{L}_{\boldsymbol{X}}|_x) \cong \Lambda^{\mathrm{top}} H^0(\mathbb{L}_{\boldsymbol{X}}|_x) \otimes [\Lambda^{\mathrm{top}} H^{-1}(\mathbb{L}_{\boldsymbol{X}}|_x)]^* \otimes \Lambda^{\mathrm{top}} H^{-2}(\mathbb{L}_{\boldsymbol{X}}|_x).$$

Now $H^{-1}(\mathbb{L}_{\boldsymbol{X}}|_x) \cong H^1(\mathbb{T}_{\boldsymbol{X}}|_x)^*$, and $\omega^0_{\boldsymbol{X}}|_x$ gives $H^0(\mathbb{L}_{\boldsymbol{X}}|_x) \cong H^{-2}(\mathbb{L}_{\boldsymbol{X}}|_x)^*$, so we see that $\Lambda^{\mathrm{top}} H^0(\mathbb{L}_{\boldsymbol{X}}|_x) \cong [\Lambda^{\mathrm{top}} H^{-2}(\mathbb{L}_{\boldsymbol{X}}|_x)]^*$. Thus we have a canonical isomorphism

$$\det(\mathbb{L}_{\boldsymbol{X}}|_x) \cong \Lambda^{\mathrm{top}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x). \tag{5}$$

Write Q_x for the nondegenerate, symmetric $\mathbb{C}$–bilinear pairing

$$H^1(\mathbb{T}_{\boldsymbol{X}}|_x) \times H^1(\mathbb{T}_{\boldsymbol{X}}|_x) \xrightarrow{Q_x := \omega^0_{\boldsymbol{X}}|_x \cdot} \mathbb{C}. \tag{6}$$

The determinant $\det Q_x$ is an isomorphism $[\Lambda^{\mathrm{top}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x)]^{\otimes^2} \to \mathbb{C}$, and $\det Q_x$ corresponds to $\iota_{\boldsymbol{X},\omega^*_{\boldsymbol{X}}}|_x$ under the isomorphism (5). There is a natural bijection

$$\{\text{orientations on } (\boldsymbol{X},\omega^*_{\boldsymbol{X}}) \text{ at } x\} \cong \{\mathbb{C}\text{–orientations on } (H^1(\mathbb{T}_{\boldsymbol{X}}|_x), Q_x)\}. \tag{7}$$

To see this, note that if $(e_1,\dots,e_n)$ is an orthonormal basis for $(H^1(\mathbb{T}_{\boldsymbol{X}}|_x), Q_x)$ then $e_1 \wedge \dots \wedge e_n$ lies in $\Lambda^{\mathrm{top}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ with $\det Q_x \colon [e_1 \wedge \dots \wedge e_n]^{\otimes^2} \mapsto 1$. Orientations for $(\boldsymbol{X},\omega^*_{\boldsymbol{X}})$ at x give isomorphisms $\lambda\colon \Lambda^{\mathrm{top}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x) \to \mathbb{C}$ with $\lambda^2 = \det Q_x$, and these correspond to orientations for $(H^1(\mathbb{T}_{\boldsymbol{X}}|_x), Q_x)$ such that $\lambda\colon e_1 \wedge \dots \wedge e_n \mapsto 1$ if $(e_1,\dots,e_n)$ is an oriented orthonormal basis.

2.5 Kuranishi atlases

We now define our notion of *Kuranishi atlases* on a topological space X. These are a simplification of m-Kuranishi spaces in [21, Section 4.7], which in turn are based on the "Kuranishi spaces" of Fukaya, Oh, Ohta and Ono [14; 15].

Definition 2.13 Let X be a topological space. A *Kuranishi neighbourhood* on X is a quadruple (V, E, s, ψ) such that:

(a) V is a smooth manifold.

(b) $\pi\colon E \to V$ is a real vector bundle over V, called the *obstruction bundle*.

(c) $s\colon V \to E$ is a smooth section of E, called the *Kuranishi section*.

(d) ψ is a homeomorphism from $s^{-1}(0)$ to an open subset $R = \operatorname{Im}\psi$ in X, where $\operatorname{Im}\psi = \{\psi(x) \mid x \in s^{-1}(0)\}$ is the image of ψ.

If $S \subseteq X$ is open, by a *Kuranishi neighbourhood over* S, we mean a Kuranishi neighbourhood (V, E, s, ψ) on X with $S \subseteq \operatorname{Im}\psi \subseteq X$.

Definition 2.14 Let (V_J, E_J, s_J, ψ_J), (V_K, E_K, s_K, ψ_K) be Kuranishi neighbourhoods on a topological space X, and $S \subseteq \operatorname{Im}\psi_J \cap \operatorname{Im}\psi_K \subseteq X$ be open. A *coordinate change* $\Phi_{JK}\colon (V_J, E_J, s_J, \psi_J) \to (V_K, E_K, s_K, \psi_K)$ *over* S is a triple $\Phi_{JK} = (V_{JK}, \phi_{JK}, \hat{\phi}_{JK})$ satisfying:

(a) V_{JK} is an open neighbourhood of $\psi_J^{-1}(S)$ in V_J.

(b) $\phi_{JK}\colon V_{JK} \to V_K$ is a smooth map.

(c) $\hat{\phi}_{JK}\colon E_J|_{V_{JK}} \to \phi_{JK}^*(E_K)$ is a morphism of vector bundles on V_{JK}.

(d) $\hat{\phi}_{JK}(s_J|_{V_{JK}}) = \phi_{JK}^*(s_K)$.

(e) $\psi_J = \psi_K \circ \phi_{JK}$ on $s_J^{-1}(0) \cap V_{JK}$.

(f) If $x \in S$, and we set $v_J = \psi_J^{-1}(x) \in V_J$ and $v_K = \psi_K^{-1}(x) \in V_K$, then the following is an exact sequence of real vector spaces:

$$0 \to T_{v_J}V_J \xrightarrow{\mathrm{d}s_J|_{v_J} \oplus \mathrm{d}\phi_{JK}|_{v_J}} E_J|_{v_J} \oplus T_{v_K}V_K \xrightarrow{-\hat{\phi}_{JK}|_{v_J} \oplus \mathrm{d}s_K|_{v_K}} E_K|_{v_K} \to 0. \tag{8}$$

We can *compose coordinate changes*: if

$$\Phi_{JK} = (V_{JK}, \phi_{JK}, \hat{\phi}_{JK})\colon (V_J, E_J, s_J, \psi_J) \to (V_K, E_K, s_K, \psi_K),$$
$$\Phi_{KL} = (V_{KL}, \phi_{KL}, \hat{\phi}_{KL})\colon (V_K, E_K, s_K, \psi_K) \to (V_L, E_L, s_L, \psi_L)$$

are coordinate changes over S_{JK}, S_{KL}, then

$$\Phi_{KL} \circ \Phi_{JK} := (V_{JK} \cap \phi_{JK}^{-1}(V_{KL}), \phi_{KL} \circ \phi_{JK}|_{\dots}, \phi_{JK}^*(\hat{\phi}_{KL}) \circ \hat{\phi}_{JK}|_{\dots})\colon$$
$$(V_J, E_J, s_J, \psi_J) \to (V_L, E_L, s_L, \psi_L)$$

is a coordinate change over $S_{JK} \cap S_{KL}$.

Definition 2.15 A *Kuranishi atlas* $\mathcal{K}$ *of virtual dimension* n on a topological space X is data $\mathcal{K} = (A, \prec, (V_J, E_J, s_J, \psi_J)_{J \in A}, \Phi_{JK,\, J \prec K \in A})$, where:

(a) A is an indexing set (not necessarily finite).

(b) $\prec$ is a partial order on A, where by convention $J \prec K$ only if $J \neq K$.

(c) (V_J, E_J, s_J, ψ_J) is a Kuranishi neighbourhood on X for each $J \in A$, with $\dim V_J - \operatorname{rank} E_J = n$.

(d) The images $\operatorname{Im}\psi_J \subseteq X$ for $J \in A$ have the property that if $J, K \in A$ with $J \neq K$ and $\operatorname{Im}\psi_J \cap \operatorname{Im}\psi_K \neq \emptyset$ then either $J \prec K$ or $K \prec J$.

(e) $\Phi_{JK} = (V_{JK}, \phi_{JK}, \hat{\phi}_{JK})\colon (V_J, E_J, s_J, \psi_J) \to (V_K, E_K, s_K, \psi_K)$ is a coordinate change for all $J, K \in A$ with $J \prec K$, over $S = \operatorname{Im}\psi_J \cap \operatorname{Im}\psi_K$.

(f) $\Phi_{KL} \circ \Phi_{JK} = \Phi_{JL}$ for all $J, K, L \in A$ with $J \prec K \prec L$.

(g) $\bigcup_{J \in A} \operatorname{Im}\psi_J = X$.

We call $\mathcal{K}$ a *finite* Kuranishi atlas if the indexing set A is finite.

If X has a Kuranishi atlas then it is locally compact. In applications we invariably impose extra global topological conditions on X, for instance X might be assumed to be compact and Hausdorff; or Hausdorff and second countable; or metrizable; or Hausdorff and paracompact.

We will also need a relative version of Kuranishi atlas in Section 3.7. Suppose Z is a manifold, and $\pi\colon X \to Z$ a continuous map. A *relative Kuranishi atlas* for $\pi\colon X \to Z$ is a Kuranishi atlas $\mathcal{K}$ on X as above, together with smooth maps $\varpi_J\colon V_J \to Z$ for $J \in A$, such that $\varpi_J|_{s_J^{-1}(0)} = \pi \circ \psi_J\colon s_J^{-1}(0) \to Z$ for all $J \in A$, and $\varpi_J|_{V_{JK}} = \varpi_K \circ \phi_{JK}\colon V_{JK} \to Z$ for all $J \prec K$ in A.

Definition 2.16 Let X be a topological space with a Kuranishi atlas $\mathcal{K}$ (Definition 2.15). For each $J \in A$ we can form the C^∞ real line bundle $\Lambda^{\mathrm{top}} T^* V_J \otimes \Lambda^{\mathrm{top}} E_J$ over V_J, where $\Lambda^{\mathrm{top}}(\cdots)$ means the top exterior power. Thus we can form the restriction

$$(\Lambda^{\mathrm{top}} T^* V_J \otimes \Lambda^{\mathrm{top}} E_J)|_{s_J^{-1}(0)} \to s_J^{-1}(0),$$

considered as a topological real line bundle over the topological space $s_J^{-1}(0)$.

If $J \prec K$ in A then for each v_J in $s_J^{-1}(0) \cap V_{JK}$ with $\phi_{JK}(v_J) = v_K$ in $s_K^{-1}(0)$ we have an exact sequence (8). Taking top exterior powers in (8) (and using a suitable orientation convention) gives an isomorphism

$$\Lambda^{\mathrm{top}} T^*_{v_J} V_J \otimes \Lambda^{\mathrm{top}} E_J|_{v_J} \cong \Lambda^{\mathrm{top}} T^*_{v_K} V_K \otimes \Lambda^{\mathrm{top}} E_K|_{v_K}.$$

This depends continuously on v_J, v_K, and so induces an isomorphism of topological line bundles on $s_J^{-1}(0) \cap V_{JK}$

$$(\Phi_{JK})_*\colon (\Lambda^{\mathrm{top}} T^* V_J \otimes \Lambda^{\mathrm{top}} E_J)|_{s_J^{-1}(0) \cap V_{JK}} \to \phi_{JK}|_{\dots}^*(\Lambda^{\mathrm{top}} T^* V_K \otimes \Lambda^{\mathrm{top}} E_K).$$

If $J \prec K \prec L$ in A then as $\Phi_{KL} \circ \Phi_{JK} = \Phi_{JL}$ by Definition 2.15(f), we see that $(\Phi_{KL})_* \circ (\Phi_{JK})_* = (\Phi_{JL})_*$ in topological line bundles over $s_J^{-1}(0) \cap V_{JK} \cap V_{JL}$.

An *orientation* on $(X, \mathcal{K})$ is a choice of orientation on the fibres of the topological real line bundle $(\Lambda^{\mathrm{top}} T^* V_J \otimes \Lambda^{\mathrm{top}} E_J)|_{s_J^{-1}(0)}$ on $s_J^{-1}(0)$ for all $J \in A$, such that $(\Phi_{JK})_*$ is orientation-preserving on $s_J^{-1}(0) \cap V_{JK}$ for all $J \prec K$ in A.

An equivalent way to think about this is that there is a natural topological real line bundle $K_X \to X$ called the *canonical bundle* with given isomorphisms

$$\iota_J\colon (\Lambda^{\mathrm{top}} T^* V_J \otimes \Lambda^{\mathrm{top}} E_J)|_{s_J^{-1}(0)} \to \psi_J^*(K_X)$$

for $J \in A$, such that $\iota_J|_{s_J^{-1}(0) \cap V_{JK}} = \phi_{JK}^*(\iota_K) \circ (\Phi_{JK})_*$ for all $J \prec K$ in A, and an orientation on $(X, \mathcal{K})$ is an orientation on the fibres of K_X.

Remark 2.17 (a) Our Kuranishi atlases are based on Joyce's "m-Kuranishi spaces" [21, Section 4.7]. They are similar to Fukaya, Oh, Ohta and Ono's "good coordinate systems" [14, Lemma A1.11; 15, Definition 6.1], and McDuff and Wehrheim's "Kuranishi atlases" [28; 29]. Our orientations are based on [15, Definition 5.8] and [14, Definition A1.17].

There are two important differences with [14; 15; 28; 29]. Firstly, [14; 15; 28; 29] use Kuranishi neighbourhoods (V, E, Γ, s, ψ), where Γ is a finite group acting equivariantly on V, E, s and ψ maps $s^{-1}(0)/\Gamma \to X$. This is because their Kuranishi spaces are a kind of derived orbifolds, not derived manifolds.

Secondly, [14; 15; 28; 29] each use a more restrictive notion of coordinate change $\Phi_{JK} = (V_{JK}, \phi_{JK}, \hat{\phi}_{JK})$, in which $\phi_{JK}\colon V_{JK} \hookrightarrow V_K$ must be an embedding, and $\hat{\phi}_{JK}\colon E_J|_{V_{JK}} \hookrightarrow \phi_{JK}^*(E_K)$ an embedding of vector bundles, so that $\dim V_J \leqslant \dim V_K$ and $\operatorname{rank} E_J \leqslant \operatorname{rank} E_K$. In the Kuranishi atlases we construct later, $\phi_{JK}\colon V_{JK} \to V_K$ will be a submersion, and $\hat{\phi}_{JK}\colon E_J|_{V_{JK}} \to \phi_{JK}^*(E_K)$ will be surjective, so that $\dim V_J \geqslant \dim V_K$ and $\operatorname{rank} E_J \geqslant \operatorname{rank} E_K$. That is, our coordinate changes actually go *the opposite way* to those in [14; 15; 28; 29].

(b) Similar structures to Kuranishi atlases are studied [14; 15; 21; 28; 29] because it is natural to construct them on many differential-geometric moduli spaces. Broadly speaking, any moduli space of solutions of a smooth nonlinear elliptic PDE on a compact manifold should admit a Kuranishi atlas. References [14; 15; 28; 29] concern moduli spaces of J–holomorphic curves in symplectic geometry.

2.6 Derived smooth manifolds and virtual classes

Readers of this paper do not need to know what a derived manifold is. Here is a brief summary of the points relevant to this paper:

- "Derived manifolds" are derived versions of smooth manifolds, where "derived" is in the sense of derived algebraic geometry.

- There are several different versions, due to Spivak [32], Borisov and Noel [3; 4] and Joyce [18; 19; 20; 21], which form ∞–categories or 2–categories. They all include ordinary manifolds **Man** as a full subcategory.

- All these versions are roughly equivalent. There are natural one-to-one correspondences between equivalence classes of derived manifolds in each theory.

- Much of classical differential geometry generalizes nicely to derived manifolds: submersions, orientations, transverse fibre products, ...

- Given a Hausdorff, second countable topological space X with a Kuranishi atlas $\mathcal{K}$ of dimension n, we can construct a derived manifold $\boldsymbol{X}$ with topological space X and dimension $\operatorname{vdim} \boldsymbol{X} = n$, unique up to equivalence. Orientations on $(X, \mathcal{K})$ are in one-to-one correspondence with orientations on $\boldsymbol{X}$.

- Compact, oriented derived manifolds $\boldsymbol{X}$ have *virtual classes* $[\boldsymbol{X}]_{\mathrm{virt}}$ in homology or bordism, generalizing the fundamental class $[X] \in H_{\dim X}(X; \mathbb{Z})$ of a compact oriented manifold X.

- These virtual classes are used to define enumerative invariants such as Gromov–Witten, Donaldson, and Donaldson–Thomas invariants. Such invariants are unchanged under deformations of the underlying geometry.

- Given a compact Hausdorff topological space X with an oriented Kuranishi atlas $\mathcal{K}$, we could construct the virtual class $[\boldsymbol{X}]_{\mathrm{virt}}$ directly from $(X, \mathcal{K})$, as in [14; 15; 28; 29], without going via the derived manifold $\boldsymbol{X}$.

Readers who do not want to know more details can now skip forward to Section 3.

2.6.1 Different definitions of derived manifold The earliest reference to derived differential geometry we are aware of is a short final paragraph by Jacob Lurie [27, Section 4.5]. Broadly following [27, Section 4.5], Lurie's student David Spivak [32] constructed an ∞–category $\mathbf{DerMan}_{\mathrm{Spi}}$ of "derived manifolds". Borisov and Noël [4] gave a simplified version, an ∞–category $\mathbf{DerMan}_{\mathrm{BoNo}}$, and showed that $\mathbf{DerMan}_{\mathrm{Spi}} \simeq \mathbf{DerMan}_{\mathrm{BoNo}}$.

Joyce [18; 19; 20] defined 2–categories $\mathbf{dMan}$ of "d-manifolds" (a kind of derived manifold), and $\mathbf{dOrb}$ of "d-orbifolds" (a kind of derived orbifold), and also strict 2–categories of d-manifolds and d-orbifolds with boundary $\mathbf{dMan}^{\mathrm{b}}$, $\mathbf{dOrb}^{\mathrm{b}}$ and with corners $\mathbf{dMan}^{\mathrm{c}}$, $\mathbf{dOrb}^{\mathrm{c}}$, and studied their differential geometry in detail. Borisov [3] constructed a 2–functor $F\colon \pi_2(\mathbf{DerMan}_{\mathrm{BoNo}}) \to \mathbf{dMan}$, where $\pi_2(\mathbf{DerMan}_{\mathrm{BoNo}})$ is the 2–category truncation of $\mathbf{DerMan}_{\mathrm{BoNo}}$, and proved that F is close to being an equivalence of 2–categories.

All of [3; 4; 18; 19; 20; 27; 32] use "C^∞–algebraic geometry", as in Joyce [17], a version of (derived) algebraic geometry in which rings are replaced by "C^∞–rings", and define derived manifolds to be special kinds of "derived C^∞–schemes".

In [21; 23; 24], Joyce gave an alternative approach to derived differential geometry based on the work of Fukaya et al [14; 15]. He defined 2–categories of "m-Kuranishi spaces" $\mathbf{mKur}$, a kind of derived manifold, and "Kuranishi spaces" $\mathbf{Kur}$, a kind of derived orbifold. Here m-Kuranishi spaces are similar to a pair $(X, \mathcal{K})$ of a Hausdorff, second countable topological space X and a Kuranishi atlas $\mathcal{K}$ in the sense of Section 2.5.

Joyce [24] will define equivalences of 2–categories **dMan** $\simeq$ **mKur** and **dOrb** $\simeq$ **Kur**, showing that the two approaches to derived differential geometry of [18; 19; 20] and [21] are essentially the same.

2.6.2 Orientations on derived manifolds Derived manifolds have a good notion of *orientation*, which behaves much like orientations on ordinary manifolds. Some references are Joyce [20, Section 4.8; 19, Section 4.8; 18, Section 4.6] for d-manifolds, Joyce [24] for m-Kuranishi spaces, and Fukaya, Oh, Ohta and Ono [15, Section 5; 14, Section A1.1] for Kuranishi spaces in their sense.

For any kind of derived manifold $\boldsymbol{X}$, we can define a (topological or C^∞) real line bundle $K_{\boldsymbol{X}}$ over the topological space X called the *canonical bundle*. It is the determinant line bundle of the cotangent complex $\mathbb{L}_{\boldsymbol{X}}$. For each $x \in \boldsymbol{X}$ we can define a *tangent space* $T_x\boldsymbol{X}$ and *obstruction space* $O_x\boldsymbol{X}$, and then

$$K_{\boldsymbol{X}}|_x \cong \Lambda^{\mathrm{top}} T_x^*\boldsymbol{X} \otimes_{\mathbb{R}} \Lambda^{\mathrm{top}} O_x\boldsymbol{X}.$$

An *orientation* on $\boldsymbol{X}$ is an orientation on the fibres of $K_{\boldsymbol{X}}$. In a similar way to (7), at a single point $x \in X$ we have a natural bijection

$$\{\text{orientations on } \boldsymbol{X} \text{ at } x\} \cong \{\text{orientations on } T_x^*\boldsymbol{X} \oplus O_x\boldsymbol{X}\}. \tag{9}$$

If (V, E, s, ψ) is a Kuranishi neighbourhood on $\boldsymbol{X}$ and $v \in s^{-1}(0) \subseteq V$ with $\psi(v) = x \in \boldsymbol{X}$, then there is a natural exact sequence

$$0 \longrightarrow T_x\boldsymbol{X} \longrightarrow T_vV \xrightarrow{\mathrm{d}s|_v} E|_v \longrightarrow O_x\boldsymbol{X} \longrightarrow 0. \tag{10}$$

Taking top exterior powers in (10) gives an isomorphism

$$K_{\boldsymbol{X}}|_x \cong \Lambda^{\mathrm{top}} T_x^*\boldsymbol{X} \otimes_{\mathbb{R}} \Lambda^{\mathrm{top}} O_x\boldsymbol{X} \cong \Lambda^{\mathrm{top}} T_v^*V \otimes_{\mathbb{R}} \Lambda^{\mathrm{top}} E|_v,$$

and thus, with a suitable orientation convention, a natural bijection

$$\{\text{orientations on } \boldsymbol{X} \text{ at } x\} \cong \{\text{orientations on } T_v^*V \oplus E|_v\}.$$

2.6.3 Kuranishi atlases and derived manifolds The next theorem relates topological spaces with Kuranishi atlases to derived manifolds. The assumption that X is Hausdorff and second countable is just to match the global topological assumptions in [4; 18; 19; 20; 21; 32]. For the last part we restrict to (a) and (b) as orientations have not been written down for the theories of (c) and (d), although this would not be very difficult.

Theorem 2.18 *Let X be a Hausdorff, second countable topological space with a Kuranishi atlas $\mathcal{K}$ of dimension n in the sense of Section 2.5. Then we can construct*

(a) *an m-Kuranishi space $\boldsymbol{X}$ in the sense of Joyce* [21, Section 4.7]*;*
(b) *a d-manifold $\boldsymbol{X}$ in the sense of Joyce* [18; 19; 20]*;*

(c) *a derived manifold in the sense of Borisov and Noël* [4]*; and*

(d) *a derived manifold in the sense of Spivak* [32].

In each case $\boldsymbol{X}$ *has topological space* X *and dimension* $\operatorname{vdim}\boldsymbol{X} = n$*, and* $\boldsymbol{X}$ *is canonical up to equivalence in the* 2*–categories* **mKur**, **dMan** *or* ∞*–categories* $\mathbf{DerMan}_{\mathrm{BoNo}}$, $\mathbf{DerMan}_{\mathrm{Spi}}$. *In cases* (a) *and* (b) *there is a natural one-to-one correspondence between orientations on* $\mathcal{K}$*, and orientations on* $\boldsymbol{X}$ *in Joyce* [18; 19; 20; 24].

If also Z *is a manifold,* $\pi\colon X \to Z$ *is continuous, and* $(\mathcal{K}, \{\varpi_J \mid J \in A\})$ *is a relative Kuranishi atlas for* $\pi\colon X \to Z$*, then we can construct a morphism of derived manifolds* $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$*, canonical up to* 2*–isomorphism, with continuous map* π.

Proof Part (a) follows from [21, Theorem 4.67] in the m-Kuranishi space case, and part (b) from [20, Theorem 4.16], in each case with topological space X, and $\operatorname{vdim}\boldsymbol{X} = n$, and $\boldsymbol{X}$ canonical up to equivalence in **mKur**, **dMan**. Part (c) then follows from (b) and Borisov [3], and part (d) from (c) and Borisov and Noël [4]. The one-to-one correspondences of orientations can be proved by comparing Definition 2.16 with Section 2.6.2. The last part also follows from [20, Theorem 4.16]. □

2.6.4 Bordism for derived manifolds We now discuss bordism, following [20, Section 4.10], [19, Section 15] and [18, Section 13].

Definition 2.19 Let Y be a manifold, and $k \in \mathbb{N}$. Consider pairs (X, f), where X is a compact, oriented manifold with $\dim X = k$, and $f\colon X \to Y$ is a smooth map. Define an equivalence relation $\sim$ on such pairs by $(X, f) \sim (X', f')$ if there exists a compact, oriented $(k+1)$–manifold with boundary W, a smooth map $e\colon W \to Y$, and a diffeomorphism of oriented manifolds $j\colon -X \amalg X' \to \partial W$, such that $f \amalg f' = e \circ i_W \circ j$, where $-X$ is X with the opposite orientation, and $i_W\colon \partial W \hookrightarrow W$ is the inclusion map.

Write $[X, f]$ for the $\sim$–equivalence class (*bordism class*) of a pair (X, f). Define the *bordism group* $B_k(Y)$ of Y to be the set of all such bordism classes $[X, f]$ with $\dim X = k$. It is an abelian group, with zero $0_Y = [\varnothing, \varnothing]$, addition $[X, f] + [X', f'] = [X \amalg X', f \amalg f']$, and inverses $-[X, f] = [-X, f]$.

Define $\Pi_{\mathrm{bo}}^{\mathrm{hom}}\colon B_k(Y) \to H_k(Y;\mathbb{Z})$ by $\Pi_{\mathrm{bo}}^{\mathrm{hom}}\colon [X, f] \mapsto f_*([X])$, where $H_*(-;\mathbb{Z})$ is singular homology, and $[X] \in H_k(X;\mathbb{Z})$ is the fundamental class.

When Y is the point $*$, the maps $f\colon X \to *$, $e\colon W \to *$ are trivial, and we can omit them, and consider $B_k(*)$ to be the abelian group of bordism classes $[X]$ of compact, oriented, k–dimensional manifolds X.

As in Conner [11, Section I.5], bordism is a generalized homology theory. Results of Thom, Wall and others in [11, Section I.2] compute the bordism groups $B_k(*)$. We define d-manifold bordism by replacing manifolds X in $[X, f]$ by d-manifolds $\boldsymbol{X}$:

Definition 2.20 Let Y be a manifold, and $k \in \mathbb{Z}$. Consider pairs $(\boldsymbol{X}, \boldsymbol{f})$, where $\boldsymbol{X} \in \mathbf{dMan}$ is a compact, oriented d-manifold with $\operatorname{vdim} \boldsymbol{X} = k$, and $\boldsymbol{f}\colon \boldsymbol{X} \to Y$ is a 1–morphism in $\mathbf{dMan}$.

Define an equivalence relation $\sim$ between such pairs by $(\boldsymbol{X}, \boldsymbol{f}) \sim (\boldsymbol{X}', \boldsymbol{f}')$ if there is a compact, oriented d-manifold with boundary $\boldsymbol{W}$ with $\operatorname{vdim} \boldsymbol{W} = k+1$, a 1–morphism $\boldsymbol{e}\colon \boldsymbol{W} \to Y$ in $\mathbf{dMan}^{\mathrm{b}}$, an equivalence of oriented d-manifolds $\boldsymbol{j}\colon -\boldsymbol{X} \amalg \boldsymbol{X}' \to \partial \boldsymbol{W}$, and a 2–morphism $\eta\colon \boldsymbol{f} \amalg \boldsymbol{f}' \Rightarrow \boldsymbol{e} \circ \boldsymbol{i}_{\boldsymbol{W}} \circ \boldsymbol{j}$, where $\boldsymbol{i}_{\boldsymbol{W}}\colon \partial \boldsymbol{W} \to \boldsymbol{W}$ is the natural 1–morphism.

Write $[\boldsymbol{X}, \boldsymbol{f}]$ for the $\sim$–equivalence class (*d-bordism class*) of a pair $(\boldsymbol{X}, \boldsymbol{f})$. Define the *d-bordism group* $dB_k(Y)$ of Y to be the set of d-bordism classes $[\boldsymbol{X}, \boldsymbol{f}]$ with $\operatorname{vdim} \boldsymbol{X} = k$. As for $B_k(Y)$, it is an abelian group, with zero $0_Y = [\boldsymbol{\varnothing}, \boldsymbol{\varnothing}]$, addition $[\boldsymbol{X}, \boldsymbol{f}] + [\boldsymbol{X}', \boldsymbol{f}'] = [\boldsymbol{X} \amalg \boldsymbol{X}', \boldsymbol{f} \amalg \boldsymbol{f}']$, and $-[\boldsymbol{X}, \boldsymbol{f}] = [-\boldsymbol{X}, \boldsymbol{f}]$. Define $\Pi_{\mathrm{bo}}^{\mathrm{dbo}}\colon B_k(Y) \to dB_k(Y)$ for $k \geqslant 0$ by $\Pi_{\mathrm{bo}}^{\mathrm{dbo}}\colon [X, f] \mapsto [X, f]$. When Y is a point $*$, we can omit $\boldsymbol{f}\colon \boldsymbol{X} \to *$, and consider $dB_k(*)$ to be the abelian group of d-bordism classes $[\boldsymbol{X}]$ of compact, oriented d-manifolds $\boldsymbol{X}$.

In [18, Section 13.2] we show that $B_*(Y)$ and $dB_*(Y)$ are isomorphic. See [32, Theorem 2.6] for an analogous (unoriented) result for Spivak's derived manifolds.

Theorem 2.21 *For any manifold Y, we have that $dB_k(Y) = 0$ for $k < 0$ and that $\Pi_{\mathrm{bo}}^{\mathrm{dbo}}\colon B_k(Y) \to dB_k(Y)$ is an isomorphism for $k \geqslant 0$.*

The main idea of the proof of Theorem 2.21 is that (compact, oriented) d-manifolds $\boldsymbol{X}$ can be turned into (compact, oriented) manifolds $\widetilde{X}$ by a small perturbation. By Theorem 2.21, we may define a projection $\Pi_{\mathrm{dbo}}^{\mathrm{hom}}\colon dB_k(Y) \to H_k(Y;\mathbb{Z})$ for $k \geqslant 0$ by $\Pi_{\mathrm{dbo}}^{\mathrm{hom}} = \Pi_{\mathrm{bo}}^{\mathrm{hom}} \circ (\Pi_{\mathrm{bo}}^{\mathrm{dbo}})^{-1}$. We think of $\Pi_{\mathrm{dbo}}^{\mathrm{hom}}$ as a *virtual class map*, and call $[\boldsymbol{X}]_{\mathrm{virt}} = \Pi_{\mathrm{dbo}}^{\mathrm{hom}}([\boldsymbol{X}, \boldsymbol{f}])$ the *virtual class*. Virtual classes are used in several areas of geometry to construct enumerative invariants using moduli spaces, for example in [14, Section A1; 15, Section 6] for Fukaya, Oh, Ohta and Ono's Kuranishi spaces, and in Behrend and Fantechi [1] in algebraic geometry.

2.6.5 Virtual classes for derived manifolds in homology If $\boldsymbol{X}$ is a compact, oriented derived manifold of dimension $k \in \mathbb{Z}$ we can also define a virtual class $[\boldsymbol{X}]_{\mathrm{virt}}$ in the homology $H_k(X;\mathbb{Z})$ of the underlying topological space X, for a suitable homology theory. By [20, Corollary 4.30] or [19, Corollary 4.31] or [18, Theorem 4.29], we can choose an embedding $\boldsymbol{f}\colon \boldsymbol{X} \hookrightarrow \mathbb{R}^n$ for $n \gg 0$. If Y is an open neighbourhood of $f(X)$ in $\mathbb{R}^n$ then Section 2.6.4 defines $\Pi_{\rm dbo}^{\rm hom}([\boldsymbol{X},\boldsymbol{f}])$ in $H_k(Y;\mathbb{Z})$. We also have a pushforward map $f_*\colon H_k(X;\mathbb{Z})\to H_k(Y;\mathbb{Z})$.

If X is a Euclidean neighbourhood retract (ENR), we can choose Y so that it retracts onto $f(X)$, and then $f_*\colon H_k(X;\mathbb{Z})\to H_k(Y;\mathbb{Z})$ is an isomorphism, so we can define the virtual class $[\boldsymbol{X}]_{\rm virt}=(f_*)^{-1}\circ\Pi_{\rm dbo}^{\rm hom}([\boldsymbol{X},\boldsymbol{f}])$ in ordinary homology $H_k(X;\mathbb{Z})$. This $[\boldsymbol{X}]_{\rm virt}$ is independent of the choices of $\boldsymbol{f}$, n, Y.

General derived manifolds may not be ENRs. In this case we use a trick that the authors learned from McDuff and Wehrheim [29, Section 7.5]. Choose a sequence $\mathbb{R}^n\supseteq Y_1\supseteq Y_2\supseteq\cdots$ of open neighbourhoods of $f(X)$ in $\mathbb{R}^n$ with $f(X)=\bigcap_{i\geqslant 1}Y_i$. Now *Steenrod homology* $H_*^{\rm St}(-;\mathbb{Z})$ (see Milnor [30]) is a homology theory with the nice properties that (i) $H_*^{\rm St}(Y_i;\mathbb{Z})\cong H_*(Y_i;\mathbb{Z})$ as Y_i is a manifold and (ii) as $f(X)=\bigcap_{i\geqslant 1}Y_i$ there is an isomorphism with the inverse limit:

$$H_k^{\rm St}(f(X);\mathbb{Z})\cong \varprojlim{}_{i\geqslant 1}\, H_k^{\rm St}(Y_i;\mathbb{Z}). \tag{11}$$

Čech homology $\check H_*(-;\mathbb{Q})$ over $\mathbb{Q}$ (the dual $\mathbb{Q}$–vector spaces to Čech cohomology $\check H^*(-;\mathbb{Q})$) has the same limiting property. Then writing $\boldsymbol{f}_i=\boldsymbol{f}\colon\boldsymbol{X}\to Y_i$, so that $\Pi_{\rm dbo}^{\rm hom}([\boldsymbol{X},\boldsymbol{f}_i])\in H_k(Y_i;\mathbb{Z})\cong H_k^{\rm St}(Y_i;\mathbb{Z})$, using (11) we may form the inverse limit $\varprojlim{}_{i\geqslant 1}\Pi_{\rm dbo}^{\rm hom}([\boldsymbol{X},\boldsymbol{f}_i])$ in $H_k^{\rm St}(f(X);\mathbb{Z})$, so that

$$[\boldsymbol{X}]_{\rm virt}:=(f_*)^{-1}\big[\varprojlim{}_{i\geqslant 1}\Pi_{\rm dbo}^{\rm hom}([\boldsymbol{X},\boldsymbol{f}_i])\big]$$

is a virtual class in $H_k^{\rm St}(X;\mathbb{Z})$, or similarly in $\check H_k(X;\mathbb{Q})$. Here $[\boldsymbol{X}]_{\rm virt}$ is independent of the choices of $\boldsymbol{f}$, n, Y_i.

For the examples in this paper, X is the complex analytic topological space of a proper $\mathbb{C}$–scheme, and therefore an ENR. Then $H_k^{\rm St}(X;\mathbb{Z})\cong H_k(X;\mathbb{Z})$ and $\check H_k(X;\mathbb{Q})\cong H_k(X;\mathbb{Q})$, and the virtual class lives in ordinary homology.

3 The main results

We now give our main results. We begin in Section 3.1 with a general existence result for a special kind of atlas for $\boldsymbol{\pi}\colon\boldsymbol{X}\to Z$, where $\boldsymbol{X}$ is a separated derived $\mathbb{C}$–scheme and Z a smooth affine classical $\mathbb{C}$–scheme, an atlas in which the charts are spectra of standard form cdgas, the coordinate changes are quasifree, and composition of coordinate changes is strictly associative.

Sections 3.2–3.5 build up to our primary goal, Theorems 3.15 and 3.16 in Section 3.5, which show that to a separated, -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ with $\mathrm{vdim}_{\mathbb{C}}\boldsymbol{X}=n$ and complex analytic topological space $X_{\rm an}$, we can build a

Kuranishi atlas $\mathcal{K}$ on $X_{\rm an}$, and so construct a derived manifold $\boldsymbol{X}_{\rm dm}$ with topological space $X_{\rm an}$, with $\mathrm{vdim}_{\mathbb{R}} \boldsymbol{X}_{\rm dm} = n$. In Section 3.6 we show that orientations on $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ and on $(X_{\rm an}, \mathcal{K})$ and on $\boldsymbol{X}_{\rm dm}$ correspond, and prove that for $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ proper and oriented, the bordism class $[\boldsymbol{X}_{\rm dm}] \in dB_n(*)$ is a "virtual cycle" independent of choices.

Section 3.7 extends Sections 3.2–3.6 to families $(\boldsymbol{\pi}\colon \boldsymbol{X} \to Z, [\omega_{\boldsymbol{X}/Z}])$ over a connected base $\mathbb{C}$–scheme Z, and shows that the bordism class $[\boldsymbol{X}_{\rm dm}^z] \in dB_n(*)$ associated to a fibre $\boldsymbol{\pi}^{-1}(z)$ is independent of $z \in Z_{\rm an}$. Finally, Sections 3.8–3.9 discuss applying our results to define Donaldson–Thomas style invariants "counting" coherent sheaves on Calabi–Yau 4–folds, and motivation from gauge theory.

3.1 Zariski homotopy atlases on derived schemes

Derived schemes and stacks, discussed in Section 2.2, are very abstract objects, and difficult to do computations with. But standard form cdgas $A^\bullet$, $B^\bullet$ and quasifree morphisms $\Phi\colon A^\bullet \to B^\bullet$ in Section 2.1 are easy to work with explicitly. Our first main result, proved in Section 4, constructs well-behaved homotopy atlases for a derived scheme $\boldsymbol{X}$, built from standard form cdgas and quasifree morphisms.

Theorem 3.1 *Let $\boldsymbol{X}$ be a separated derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ be a smooth classical affine $\mathbb{C}$–scheme for B a smooth $\mathbb{C}$–algebra of pure dimension, and $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ be a morphism. Suppose we are given data $\{(A_i^\bullet, \boldsymbol{\alpha}_i, \beta_i) \mid i \in I\}$, where I is an indexing set and for each $i \in I$, $A_i^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ is a standard form cdga, and $\boldsymbol{\alpha}_i\colon \mathbf{Spec}\, A_i^\bullet \hookrightarrow \boldsymbol{X}$ is a Zariski open inclusion in $\mathbf{dSch}_{\mathbb{C}}$, and $\beta_i\colon B \to A_i^0$ is a smooth morphism of classical $\mathbb{C}$–algebras such that the following diagram homotopy commutes in $\mathbf{dSch}_{\mathbb{C}}$:*

$$\begin{tikzcd} \mathbf{Spec}\, A_i^\bullet \arrow[rr, "\boldsymbol{\alpha}_i"] \arrow[dr, "\mathbf{Spec}\,\beta_i"'] & & \boldsymbol{X} \arrow[d, "\boldsymbol{\pi}"] \\ & & \operatorname{Spec} B = Z \end{tikzcd} \tag{12}$$

Here we regard β_i as a morphism $B \to A_i^\bullet$. Then we can construct the following data:

(i) *For all finite subsets $\emptyset \neq J \subseteq I$, a standard form cdga $A_J^\bullet \in \mathbf{cdga}_{\mathbb{C}}$, a Zariski open inclusion $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$, with image $\operatorname{Im} \boldsymbol{\alpha}_J = \bigcap_{i \in J} \operatorname{Im} \boldsymbol{\alpha}_i$, and a smooth morphism of classical $\mathbb{C}$–algebras $\beta_J\colon B \to A_J^0$, such that the following diagram homotopy commutes in $\mathbf{dSch}_{\mathbb{C}}$:*

$$\begin{tikzcd} \mathbf{Spec}\, A_J^\bullet \arrow[rr, "\boldsymbol{\alpha}_J"] \arrow[dr, "\mathbf{Spec}\,\beta_J"'] & & \boldsymbol{X} \arrow[d, "\boldsymbol{\pi}"] \\ & & \operatorname{Spec} B = Z \end{tikzcd} \tag{13}$$

When $J = \{i\}$ for $i \in I$ we have $A_{\{i\}}^\bullet = A_i^\bullet$, $\boldsymbol{\alpha}_{\{i\}} = \boldsymbol{\alpha}_i$, and $\beta_{\{i\}} = \beta_i$.

(ii) *For all inclusions of finite subsets $\varnothing \neq K \subseteq J \subseteq I$, a quasifree morphism of standard form cdgas $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ with $\beta_J = \Phi_{JK} \circ \beta_K\colon B \to A_J^0$, such that the following diagram homotopy commutes in* $\mathbf{dSch}_{\mathbb{C}}$:

$$\begin{tikzcd} \mathbf{Spec}\, A_J^\bullet \arrow[rr, "\mathbf{Spec}\,\Phi_{JK}"] \arrow[drr, "\boldsymbol{\alpha}_J"'] & & \mathbf{Spec}\, A_K^\bullet \arrow[d, "\boldsymbol{\alpha}_K"] \\ & & \boldsymbol{X} \end{tikzcd} \tag{14}$$

If $\varnothing \neq L \subseteq K \subseteq J \subseteq I$ then $\Phi_{JL} = \Phi_{JK} \circ \Phi_{KL}\colon A_L^\bullet \to A_J^\bullet$.

3.2 Interpreting Zariski atlases using complex geometry

Given a -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ satisfying certain conditions, we will construct a derived manifold structure $\boldsymbol{X}_{\mathrm{dm}}$ on the complex analytic topological space X_{an} underlying $\boldsymbol{X}$. To do this, we need a *change of language*: we have to pass from talking about derived schemes $\boldsymbol{X}$, cdgas $A^\bullet$, etc, to talking about smooth manifolds V, vector bundles $E \to V$, smooth sections $s\colon V \to E$, as $\boldsymbol{X}_{\mathrm{dm}}$ will be built by gluing together such local Kuranishi models (V, E, s).

Therefore we now rewrite part of the output $A_J^\bullet$, $\beta_J\colon B \to A_J^0$, $\Phi_{JK}\colon A_J^\bullet \to A_K^\bullet$ of Theorem 3.1 in terms of complex manifolds V, holomorphic vector bundles $E \to V$, and holomorphic sections $s\colon V \to E$. In Section 3.5 we will pass to certain real vector bundles $E^+ = E/E^-$ to define $\boldsymbol{X}_{\mathrm{dm}}$.

First we interpret standard form cdgas $A^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ using holomorphic data. We discuss only data from degrees $0, -1, -2$ in $A^\bullet$, as this is all we need, but one could also define vector bundles $G, H, \dots$ over V corresponding to $M^{-3}, M^{-4}, \dots$, and many vector bundle morphisms, satisfying certain equations.

Definition 3.2 Let $A^\bullet = (\cdots \to A^{-2} \xrightarrow{\mathrm{d}} A^{-1} \xrightarrow{\mathrm{d}} A^0)$ be a standard form cdga over $\mathbb{C}$, as in Section 2.1. Then A^0 is a finitely generated smooth $\mathbb{C}$–algebra, so $V^{\mathrm{alg}} := \operatorname{Spec} A^0$ is a smooth affine $\mathbb{C}$–scheme, assumed of pure dimension, as in Section 2.1. Now any $\mathbb{C}$–scheme S has an underlying complex analytic space S_{an}, which is a complex manifold if S is smooth and of pure dimension.

Write V for the complex manifold $(V^{\mathrm{alg}})_{\mathrm{an}}$ associated to $V^{\mathrm{alg}} = \operatorname{Spec} A^0$.

As $A^\bullet$ is of standard form, the graded $\mathbb{C}$–algebra A^* is freely generated over A^0 by a series of finitely generated free A^0–modules $M^{-1} \subseteq A^{-1}$, $M^{-2} \subseteq A^{-2}, \dots$. Thus $A^{-1} \cong M^{-1}$, $A^{-2} \cong M^{-2} \oplus \Lambda_{A^0}^2 M^{-1}$, and so on, giving

$$M^{-1} = A^{-1}, \quad M^{-2} \cong A^{-2}/\Lambda_{A^0}^2 A^{-1}, \quad \dots. \tag{15}$$

Hence, the M^i are determined by A^* as A^0–modules up to canonical isomorphism, although for $i \leqslant -2$ the inclusions $M^i \hookrightarrow A^i$ involve an arbitrary choice.

Now finitely generated free A^0–modules M are those of the form $M \cong H^0(C^{\mathrm{alg}})$ for $C^{\mathrm{alg}} \to V^{\mathrm{alg}} = \operatorname{Spec} A^0$ a trivial algebraic vector bundle. Write $E^{\mathrm{alg}} \to V^{\mathrm{alg}}$, $F^{\mathrm{alg}} \to V^{\mathrm{alg}}$ for the trivial algebraic vector bundles (unique up to canonical isomorphism) with $M^{-1} \cong H^0((E^{\mathrm{alg}})^*)$, $M^{-2} \cong H^0((F^{\mathrm{alg}})^*)$. That is, we set $E^{\mathrm{alg}} = \operatorname{Spec}\operatorname{Sym}^*_{A^0}(M^{-1})$, and so on. Write $E \to V$, $F \to V$ for the holomorphic vector bundles corresponding to E^{alg}, F^{alg}.

We now have isomorphisms

$$
\begin{aligned}
A^0 &\cong H^0(\mathcal{O}_{V^{\mathrm{alg}}}),\\
A^{-1} &\cong H^0((E^{\mathrm{alg}})^*),\\
A^{-2} &\cong H^0((F^{\mathrm{alg}})^*) \oplus H^0(\Lambda^2(E^{\mathrm{alg}})^*).
\end{aligned} \tag{16}
$$

Thus d: $A^{-1} \to A^0$ is identified with an A^0–module morphism $H^0((E^{\mathrm{alg}})^*) \to H^0(\mathcal{O}_{V^{\mathrm{alg}}})$, that is, a morphism $(E^{\mathrm{alg}})^* \to \mathcal{O}_{V^{\mathrm{alg}}}$ of algebraic vector bundles, which is dual to a morphism $\mathcal{O}_{V^{\mathrm{alg}}} \cong \mathcal{O}^*_{V^{\mathrm{alg}}} \to E^{\mathrm{alg}}$, ie a section $s^{\mathrm{alg}} \in H^0(E^{\mathrm{alg}})$ of E^{alg}. Write $s \in H^0(E)$ for the corresponding holomorphic section.

Similarly, write $t^{\mathrm{alg}}\colon E^{\mathrm{alg}} \to F^{\mathrm{alg}}$ for the algebraic vector bundle morphism dual to the component of d: $A^{-2} \to A^{-1}$ mapping $H^0((F^{\mathrm{alg}})^*) \to H^0((E^{\mathrm{alg}})^*)$ under (16), and write $t\colon E \to F$ for the corresponding morphism of holomorphic vector bundles. Then $\mathrm{d} \circ \mathrm{d} = 0$ implies that $t \circ s = 0\colon \mathcal{O}_V \to F$.

We should also consider how this data E, F, s, t depends on the choice of inclusion $M^{-2} \hookrightarrow A^{-2}$. Here E, F are independent of choices up to canonical isomorphism, and s is independent of choices. Changing the inclusion $M^{-2} \hookrightarrow A^{-2}$ is equivalent to choosing an algebraic vector bundle morphism $\gamma^{\mathrm{alg}}\colon \Lambda^2 E^{\mathrm{alg}} \to F^{\mathrm{alg}}$ and identifying M^{-2} with the image of $\mathrm{id} \oplus (\gamma^{\mathrm{alg}})^*\colon H^0((F^{\mathrm{alg}})^*) \hookrightarrow H^0((F^{\mathrm{alg}})^*) \oplus H^0(\Lambda^2(E^{\mathrm{alg}})^*)$. Writing $\gamma\colon \Lambda^2 E \to F$ for the corresponding holomorphic morphism, this changes t to $\tilde{t}$, where

$$
\tilde{t} = t + \gamma \circ (- \wedge s). \tag{17}
$$

Notice that $t|_v\colon E|_v \to F|_v$ is independent of choices at $v \in V$ with $s(v) = 0$.

Next suppose $\boldsymbol{X}$ is a derived $\mathbb{C}$–scheme and $\boldsymbol{\alpha}\colon \mathbf{Spec}\, A^\bullet \hookrightarrow \boldsymbol{X}$ a Zariski open inclusion. Write $X = t_0(\boldsymbol{X})$ for the classical $\mathbb{C}$–scheme, and X_{an} for the set of $\mathbb{C}$–points of X equipped with the complex analytic topology. (One can give X_{an} the structure of a complex analytic space, but we will not use this.) Then $t_0(\mathbf{Spec}\, A^\bullet)$ is the $\mathbb{C}$–subscheme $(s^{\mathrm{alg}})^{-1}(0) \subseteq V^{\mathrm{alg}}$, so $\alpha = t_0(\boldsymbol{\alpha})$ is a Zariski open inclusion

$(s^{\mathrm{alg}})^{-1}(0) \hookrightarrow X$. Write $\psi\colon s^{-1}(0) \hookrightarrow X_{\mathrm{an}}$ for the corresponding map of $\mathbb{C}$–points. Then ψ is a homeomorphism with an open set $R = \mathrm{Im}\,\psi \subseteq X_{\mathrm{an}}$. Note that (V, E, s, ψ) is a Kuranishi neighbourhood on X_{an}, in the sense of Section 2.5.

As we explained in Sections 2.1–2.2, if $A^\bullet$ is a standard form cdga then it is easy to compute the cotangent complex $\mathbb{L}_{A^\bullet} \simeq \Omega^1_{A^\bullet}$, and this also can be identified with the cotangent complex $\mathbb{L}_{\mathbf{Spec}\,A^\bullet}$ of the derived scheme $\mathbf{Spec}\,A^\bullet$. Let $v \in s^{-1}(0) \subseteq V$ with $\psi(v) = x \in X_{\mathrm{an}}$. Then v is a $\mathbb{C}$–point of $\mathbf{Spec}\,A^\bullet$ and x a $\mathbb{C}$–point of $\boldsymbol{X}$ with $\boldsymbol{\alpha}(v) = x$, so $\mathbb{L}_{\boldsymbol{\alpha}}|_v\colon \mathbb{L}_{\boldsymbol{X}}|_x \to \mathbb{L}_{\mathbf{Spec}\,A^\bullet}|_v$ is a quasi-isomorphism, and induces an isomorphism on cohomology. One can show that $\mathbb{L}_{\mathbf{Spec}\,A^\bullet}|_v$ is represented by the complex of $\mathbb{C}$–vector spaces

$$\cdots \to F|_v^* \xrightarrow{t|_v^*} E|_v^* \xrightarrow{\mathrm{d}s|_v^*} T_v^*V \to 0, \tag{18}$$

with T_v^*V in degree 0. Dualizing to tangent complexes and taking cohomology, we get canonical isomorphisms

$$H^0(\mathbb{T}_{\boldsymbol{\alpha}}|_v)\colon \mathrm{Ker}(\mathrm{d}s|_v\colon T_vV \to E|_v) \to H^0(\mathbb{T}_{\boldsymbol{X}}|_x), \tag{19}$$

$$H^1(\mathbb{T}_{\boldsymbol{\alpha}}|_v)\colon \frac{\mathrm{Ker}(t|_v\colon E|_v \to F|_v)}{\mathrm{Im}(\mathrm{d}s|_v\colon T_vV \to E|_v)} \to H^1(\mathbb{T}_{\boldsymbol{X}}|_x). \tag{20}$$

Now suppose that $Z = \mathrm{Spec}\,B$ is a smooth classical affine $\mathbb{C}$–scheme of pure dimension, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ is a morphism, and $\beta\colon B \to A^0$ is a smooth morphism of $\mathbb{C}$–algebras, such that as for (12)–(13) the following homotopy commutes:

$$\begin{tikzcd} \mathbf{Spec}\,A^\bullet \arrow[rr, "\boldsymbol{\alpha}"] \arrow[drr, "\mathbf{Spec}\,\beta"'] & & \boldsymbol{X} \arrow[d, "\boldsymbol{\pi}"] \\ & & \mathrm{Spec}\,B = Z \end{tikzcd} \tag{21}$$

Then Z_{an} is a complex manifold, and $\tau^{\mathrm{alg}} := \mathrm{Spec}\,\beta\colon V^{\mathrm{alg}} \to Z$ is a smooth morphism of $\mathbb{C}$–schemes, and $\tau := (\tau^{\mathrm{alg}})_{\mathrm{an}}\colon V \to Z_{\mathrm{an}}$ is a holomorphic submersion of complex manifolds. We can form the relative cotangent complexes $\mathbb{L}_{\boldsymbol{X}/Z}$, $\mathbb{L}_{\mathbf{Spec}\,A^\bullet/Z}$ and dual relative tangent complexes $\mathbb{T}_{\boldsymbol{X}/Z}$, $\mathbb{T}_{\mathbf{Spec}\,A^\bullet/Z}$, and (21) gives morphisms $\mathbb{L}_{\boldsymbol{\alpha}}\colon \mathbb{L}_{\boldsymbol{X}/Z} \to \mathbb{L}_{\mathbf{Spec}\,A^\bullet/Z}$, $\mathbb{T}_{\boldsymbol{\alpha}}\colon \mathbb{T}_{\mathbf{Spec}\,A^\bullet/Z} \to \mathbb{T}_{\boldsymbol{X}/Z}$.

Write $T(V/Z_{\mathrm{an}}) = \mathrm{Ker}(\mathrm{d}\tau\colon TV \to \tau^*(TZ_{\mathrm{an}}))$ for the *relative tangent bundle* of V/Z_{an}. It is a holomorphic vector subbundle of TV of rank $\dim V - \dim Z$, as τ is a holomorphic submersion. Let $v \in s^{-1}(0) \subseteq V$ with $\psi(v) = x \in X_{\mathrm{an}}$ and $\tau(v) = \pi(x) = z \in Z_{\mathrm{an}}$. Then as in (18), $\mathbb{L}_{\mathbf{Spec}\,A^\bullet/Z}|_v$ is represented by the complex of $\mathbb{C}$–vector spaces

$$\cdots \to F|_v^* \xrightarrow{t|_v^*} E|_v^* \xrightarrow{\mathrm{d}s|_v^*} T_v^*(V/Z_{\mathrm{an}}) \to 0,$$

with $T_v^*(V/Z_{\mathrm{an}})$ in degree 0. As for (19)–(20) we get canonical isomorphisms

$$H^0(\mathbb{T}_{\boldsymbol{\alpha}}|_v)\colon \mathrm{Ker}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v) \to H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x), \tag{22}$$

$$H^1(\mathbb{T}_{\boldsymbol{\alpha}}|_v)\colon \frac{\mathrm{Ker}(t|_v\colon E|_v \to F|_v)}{\mathrm{Im}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v)} \to H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x). \tag{23}$$

Example 3.3 Suppose $(A^\bullet, \omega_{A^\bullet})$ is in -2–Darboux form, in the sense of Definition 2.9, with coordinates $x_1, \dots, x_m$, $y_1, \dots, y_n$, $z_1, \dots, z_m$, and 2–form $\omega_{A^\bullet}$ in (1), depending on invertible functions $q_1, \dots, q_n \in A^0$.

Let V, E, F, s, t be as in Definition 3.2. Then V is a smooth $\mathbb{C}$–scheme of dimension m, with étale coordinates $(x_1, \dots, x_m)$, so that TV is a trivial vector bundle with basis of sections $\partial/\partial x_1, \dots, \partial/\partial x_m$. Also E is a trivial vector bundle of rank n, with basis $e_1 := \partial/\partial y_1, \dots, e_n := \partial/\partial y_n$, and F is trivial of rank m, with basis $\partial/\partial z_1, \dots, \partial/\partial z_m$. Using the first line of $\omega_{A^\bullet}$ in (1), it is natural to identify $F \cong T^*V$ by identifying $\partial/\partial z_i \cong \mathrm{d}_{\mathrm{dR}} x_i$ for $i = 1, \dots, m$.

The natural section $s \in H^0(E)$ is $s = s_1 e_1 + \cdots + s_n e_n$. Write $\epsilon^1, \dots, \epsilon^n$ for the basis of sections of E^* dual to $e_1, \dots, e_n$, so that $\epsilon^j \cong \mathrm{d}_{\mathrm{dR}} y_j$. Motivated by the second line of $\omega_{A^\bullet}$ in (1), define $Q = q_1 \epsilon^1 \otimes \epsilon^1 + \cdots + q_n \epsilon^n \otimes \epsilon^n$ in $H^0(S^2 E^*)$. Then Q is a natural nondegenerate quadratic form on the fibres of E, and (2) implies that $Q(s,s) = 0$.

Identifying $F = T^*V$, from (3) we see that $t\colon E \to F$ is given by

$$t(e_j) = \sum_{i=1}^m \left(2q_j \frac{\partial s_j}{\partial x_i} + s_j \frac{\partial q_j}{\partial x_i} \right) \mathrm{d}_{\mathrm{dR}} x_i = 2q_j \, \mathrm{d}_{\mathrm{dR}} s_j + s_j \, \mathrm{d}_{\mathrm{dR}} q_j \tag{24}$$

for $j = 1, \dots, n$. Then $t \circ s = 0$ follows from applying d_{dR} to $Q(s,s) = 0$.

What will matter later is that we have a complex manifold V, a holomorphic vector bundle $E \to V$, a section $s \in H^0(E)$, and a nondegenerate holomorphic quadratic form $Q \in H^0(S^2 E^*)$ with $Q(s,s) = 0$, such that the classical complex analytic topological space $(\mathrm{Spec}\, H^0(A^\bullet))_{\mathrm{an}}$ is $s^{-1}(0) \subseteq V$.

Next we interpret quasifree morphisms of standard form cdgas $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$, as in Theorem 3.1(ii), in terms of complex geometry.

Definition 3.4 Let $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ be a quasifree morphism of standard form cdgas over $\mathbb{C}$, as in Section 2.1. Let V_J^{alg}, E_J^{alg}, F_J^{alg}, s_J^{alg}, t_J^{alg}, V_J, E_J, F_J, s_J, t_J be as in Definition 3.2 for $A_J^\bullet$, and let V_K^{alg}, $E_K^{\mathrm{alg}}, \dots, t_K$ be as for $A_K^\bullet$.

Then $\phi_{JK}^{\mathrm{alg}} := \mathrm{Spec}\, \Phi_{JK}^0\colon V_J^{\mathrm{alg}} = \mathrm{Spec}\, A_J^0 \to V_K^{\mathrm{alg}} = \mathrm{Spec}\, A_K^0$ is a $\mathbb{C}$–scheme morphism. Write $\phi_{JK}\colon V_J \to V_K$ for the corresponding holomorphic map. The quasifree condition on Φ_{JK} implies $\mathrm{d}\phi_{JK}^{\mathrm{alg}}\colon (\phi_{JK}^{\mathrm{alg}})^*(T^*V_K^{\mathrm{alg}}) \to T^*V_J^{\mathrm{alg}}$ is injective, and thus $\mathrm{d}\phi_{JK}\colon \phi_{JK}^*(T^*V_K) \to T^*V_J$ is injective, that is, $\phi_{JK}\colon V_J \to V_K$ is a submersion of complex manifolds.

Now $\Phi_{JK}^{-1}\colon A_K^{-1} \to A_J^{-1}$ induces an A_J^0–linear map

$$(\Phi_{JK}^{-1})_*\colon A_K^{-1} \otimes_{A_K^0} A_J^0 \to A_J^{-1},$$

which under (16) corresponds to an algebraic vector bundle morphism

$$(\phi_{JK}^{\mathrm{alg}})^*((E_K^{\mathrm{alg}})^*) \to (E_J^{\mathrm{alg}})^*.$$

Write $\chi_{JK}^{\mathrm{alg}}\colon E_J^{\mathrm{alg}} \to (\phi_{JK}^{\mathrm{alg}})^*(E_K^{\mathrm{alg}})$ for the dual morphism, and $\chi_{JK}\colon E_J \to \phi_{JK}^*(E_K)$ for the corresponding morphism of holomorphic vector bundles. It is surjective, as Φ_{JK} is quasifree. Then $\mathrm{d} \circ \Phi_{JK}^{-1} = \Phi_{JK}^0 \circ \mathrm{d}$ implies that

$$\chi_{JK}(s_J) = \phi_{JK}^*(s_K) \in H^0(\phi_{JK}^*(E_K)). \tag{25}$$

By (15) we have a natural composition of morphisms

$$H^0((F_K^{\mathrm{alg}})^*) \cong M_K^{-2} \cong A_K^{-2}/\Lambda^2_{A_K^0} A_K^{-1} \xrightarrow{(\Phi_{JK}^{-2})_*} A_J^{-2}/\Lambda^2_{A_J^0} A_J^{-1} \cong M_J^{-2} \cong H^0((F_J^{\mathrm{alg}})^*).$$

The induced A_J^0–linear map corresponds to a natural algebraic vector bundle morphism $(\phi_{JK}^{\mathrm{alg}})^*((F_K^{\mathrm{alg}})^*) \to (F_J^{\mathrm{alg}})^*$. Write $\xi_{JK}^{\mathrm{alg}}\colon F_J^{\mathrm{alg}} \to (\phi_{JK}^{\mathrm{alg}})^*(F_K^{\mathrm{alg}})$ for the dual morphism, and $\xi_{JK}\colon F_J \to \phi_{JK}^*(F_K)$ for the corresponding morphism of holomorphic vector bundles. It is surjective, as Φ_{JK} is quasifree.

These ξ_{JK}^{alg}, ξ_{JK} are independent of choices, as they depend on the canonical isomorphism $M^{-2} \cong A^{-2}/\Lambda^2_{A^0} A^{-1}$ rather than on the noncanonical inclusion $M^{-2} \hookrightarrow A^{-2}$ in Definition 3.2. However, Φ_{JK}^{-2} need not map $M_K^{-2} \subseteq A_K^{-2}$ to $M_J^{-2} \subseteq A_J^{-2}$, and so under the isomorphisms (16) need not map $H^0((F_K^{\mathrm{alg}})^*) \to H^0((F_J^{\mathrm{alg}})^*)$. Write $\delta_{JK}^{\mathrm{alg}}\colon \Lambda^2 E_J^{\mathrm{alg}} \to (\phi_{JK}^{\mathrm{alg}})^*(F_K^{\mathrm{alg}})$ for the algebraic vector bundle morphism dual to the component of Φ_{JK}^{-2} mapping $H^0((F_K^{\mathrm{alg}})^*) \to H^0(\Lambda^2(E_J^{\mathrm{alg}})^*)$, and $\delta_{JK}\colon \Lambda^2 E_J \to \phi_{JK}^*(F_K)$ for the corresponding morphism of vector bundles. Then $\mathrm{d} \circ \Phi_{JK}^{-2} = \Phi_{JK}^{-1} \circ \mathrm{d}$ implies that

$$\xi_{JK} \circ t_J + \delta_{JK} \circ (- \wedge s_J) = \phi_{JK}^*(t_K) \circ \chi_{JK}\colon E_J \to \phi_{JK}^*(F_K). \tag{26}$$

Therefore χ_{JK}, ξ_{JK} do not strictly commute with t_J, t_K, which is not surprising, since t_J, t_K depend on arbitrary choices as in (17). But notice that $\xi_{JK}|_v \circ t_J|_v = t_K|_{\phi_{JK}(v)} \circ \chi_{JK}|_v$ at $v \in V_J$ with $s_J(v) = 0$.

Next suppose that we are given Zariski open inclusions $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$ and $\boldsymbol{\alpha}_K\colon \mathbf{Spec}\, A_K^\bullet \hookrightarrow \boldsymbol{X}$ into a derived $\mathbb{C}$–scheme $\boldsymbol{X}$, such that (14) homotopy commutes, and let

$$\psi_J\colon s_J^{-1}(0)\hookrightarrow X_{\mathrm{an}},\quad \psi_K\colon s_K^{-1}(0)\hookrightarrow X_{\mathrm{an}}$$

be as in Definition 3.2. As the classical truncation of (14) commutes, we see that

$$\psi_J=\psi_K\circ\phi_{JK}|_{s_J^{-1}(0)}\colon s_J^{-1}(0)\to X_{\mathrm{an}}. \tag{27}$$

Suppose $v_J\in s_J^{-1}(0)\subseteq V_J$ with $\phi_{JK}(v_J)=v_K\in s_K^{-1}(0)\subseteq V_K$ and $\psi_J(v_J)=\psi_K(v_K)=x\in X_{\mathrm{an}}$. As (14) homotopy commutes, the corresponding morphisms of tangent complexes $\mathbb{T}_{\mathbf{Spec}\,A_J^\bullet}$, $\mathbb{T}_{\mathbf{Spec}\,A_K^\bullet}$, $\mathbb{T}_{\boldsymbol{X}}$ commute up to homotopy, so restricting to v_J, v_K, x and taking homology gives strictly commuting diagrams. Thus using (19)–(20), we see that the following diagrams commute:

$$\begin{tikzcd}
\mathrm{Ker}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}V_J\to E_J|_{v_J}) \arrow[d, "(\mathrm{d}\phi_{JK}|_{v_J})|_{\mathrm{Ker}(\cdots)}"'] \arrow[dr, "H^0(\mathbb{T}_{\boldsymbol{\alpha}_J}|_{v_J})"] & \\
\mathrm{Ker}(\mathrm{d}s_K|_{v_K}\colon T_{v_K}V_K\to E_K|_{v_K}) \arrow[r, "H^0(\mathbb{T}_{\boldsymbol{\alpha}_K}|_{v_K})"] & H^0(\mathbb{T}_{\boldsymbol{X}}|_x)
\end{tikzcd} \tag{28}$$

$$\begin{tikzcd}
\dfrac{\mathrm{Ker}(t_J|_{v_J}\colon E_J|_{v_J}\to F_J|_{v_J})}{\mathrm{Im}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}V_J\to E_J|_{v_K})} \arrow[d, "(\chi_{JK}|_{v_J})_*"'] \arrow[dr, "H^1(\mathbb{T}_{\boldsymbol{\alpha}_J}|_{v_J})"] & \\
\dfrac{\mathrm{Ker}(t_K|_{v_K}\colon E_K|_{v_K}\to F_K|_{v_K})}{\mathrm{Im}(\mathrm{d}s_K|_{v_K}\colon T_{v_K}V_K\to E_K|_{v_K})} \arrow[r, "H^1(\mathbb{T}_{\boldsymbol{\alpha}_K}|_{v_K})"] & H^1(\mathbb{T}_{\boldsymbol{X}}|_x)
\end{tikzcd} \tag{29}$$

Now suppose that $Z=\mathrm{Spec}\,B$ is a smooth classical affine $\mathbb{C}$–scheme of pure dimension, $\boldsymbol{\pi}\colon \boldsymbol{X}\to Z$ is a morphism, and $\beta_J\colon B\to A_J^0$, $\beta_K\colon B\to A_K^0$ are smooth morphisms of $\mathbb{C}$–algebras, such that (13) homotopy commutes for J, K, and $\beta_J=\Phi_{JK}\circ\beta_K$. As in Definition 3.2 we have holomorphic submersions $\tau_J\colon V_J\to Z_{\mathrm{an}}$, $\tau_K\colon V_K\to Z_{\mathrm{an}}$, with $\tau_J=\tau_K\circ\phi_{JK}\colon V_J\to Z_{\mathrm{an}}$ as $\beta_J=\Phi_{JK}\circ\beta_K$. Let $v_J\in s_J^{-1}(0)\subseteq V_J$ with $\phi_{JK}(v_J)=v_K\in s_K^{-1}(0)\subseteq V_K$, and $\psi_J(v_J)=\psi_K(v_K)=x\in X_{\mathrm{an}}$, and $\tau_J(v_J)=\tau_K(v_K)=\pi(x)=z\in Z_{\mathrm{an}}$. Then using (22)–(23), we see that the following diagrams commute:

$$\begin{tikzcd}
\mathrm{Ker}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}(V_J/Z_{\mathrm{an}})\to E_J|_{v_J}) \arrow[d, "(\mathrm{d}\phi_{JK}|_{v_J})|_{\mathrm{Ker}(\cdots)}"'] \arrow[dr, "H^0(\mathbb{T}_{\boldsymbol{\alpha}_J}|_{v_J})"] & \\
\mathrm{Ker}(\mathrm{d}s_K|_{v_K}\colon T_{v_K}(V_K/Z_{\mathrm{an}})\to E_K|_{v_K}) \arrow[r, "H^0(\mathbb{T}_{\boldsymbol{\alpha}_K}|_{v_K})"] & H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x)
\end{tikzcd} \tag{30}$$

$$
\begin{tikzcd}
\dfrac{\mathrm{Ker}(t_J|_{v_J}\colon E_J|_{v_J}\to F_J|_{v_J})}{\mathrm{Im}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}(V_J/Z_{\mathrm{an}})\to E_J|_{v_K})} \arrow[d, "(\chi_{JK}|_{v_J})_*"] \arrow[drr, "H^1(\mathbb{T}_{\boldsymbol{\alpha}_J}|_{v_J})"] & & \\
\dfrac{\mathrm{Ker}(t_K|_{v_K}\colon E_K|_{v_K}\to F_K|_{v_K})}{\mathrm{Im}(\mathrm{d}s_K|_{v_K}\colon T_{v_K}(V_K/Z_{\mathrm{an}})\to E_K|_{v_K})} \arrow[rr, "H^1(\mathbb{T}_{\boldsymbol{\alpha}_K}|_{v_K})"] & & H^1(\mathbb{T}_{\boldsymbol{X}/\boldsymbol{Z}}|_x)
\end{tikzcd}
\tag{31}
$$

Applying Definitions 3.2 and 3.4 to the conclusions of Theorem 3.1 yields:

Corollary 3.5 *In the situation of Theorem 3.1, write X_{an} for the set of $\mathbb{C}$–points of $X=t_0(\boldsymbol{X})$, regarded as a topological space with the complex analytic topology. Then we obtain the following data in complex geometry:*

(i) *For all finite subsets $\varnothing\neq J\subseteq I$, a complex manifold V_J, a holomorphic submersion $\tau_J\colon V_J\to Z_{\mathrm{an}}$, holomorphic vector bundles $E_J,F_J\to V_J$, a holomorphic section $s_J\colon V_J\to E_J$, and a homeomorphism $\psi_J\colon s_J^{-1}(0)\to R_J\subseteq X_{\mathrm{an}}$, where $R_J\subseteq X_{\mathrm{an}}$ is open, with $\pi\circ\psi_J=\tau_J|_{s_J^{-1}(0)}\colon s_J^{-1}(0)\to Z_{\mathrm{an}}$. These image subsets satisfy $R_J=\bigcap_{i\in J}R_{\{i\}}$.*

By making an additional arbitrary choice we also obtain a morphism of holomorphic vector bundles $t_J\colon E_J\to F_J$, with $t_J\circ s_J=0$. Different choices $t_J,\tilde t_J$ are related by (17). The restrictions $t_J|_{v_J}\colon E_J|_{v_J}\to F_J|_{v_J}$ for $v_J\in s_J^{-1}(0)$ are independent of choices. For each $v_J\in s_J^{-1}(0)$ with $\psi_J(v_J)=x\in X_{\mathrm{an}}$, there are canonical isomorphisms (19)–(20) writing $H^i(\mathbb{T}_{\boldsymbol{X}}|_x)$ for $i=0,1$ and (22)–(23) writing $H^i(\mathbb{T}_{\boldsymbol{X}/\boldsymbol{Z}}|_x)$ for $i=0,1$ in terms of V_J, E_J, F_J, s_J, t_J, τ_J at v_J.

(ii) *For all inclusions of finite subsets $\varnothing\neq K\subseteq J\subseteq I$, a holomorphic submersion $\phi_{JK}\colon V_J\to V_K$, and surjective morphisms of holomorphic vector bundles $\chi_{JK}\colon E_J\to\phi_{JK}^*(E_K)$ and $\xi_{JK}\colon F_J\to\phi_{JK}^*(F_K)$. These satisfy $\tau_J=\tau_K\circ\phi_{JK}\colon V_J\to Z_{\mathrm{an}}$, and $\chi_{JK}(s_J)=\phi_{JK}^*(s_K)$, and $\psi_J=\psi_K\circ\phi_{JK}|_{s_J^{-1}(0)}\colon s_J^{-1}(0)\to X_{\mathrm{an}}$.*

If t_J, t_K are possible choices in (i) *then χ_{JK}, ξ_{JK}, t_J, t_K are related as in (26). If $v_J\in s_J^{-1}(0)$ with $\phi_{JK}(v_J)=v_K\in s_K^{-1}(0)$, this implies that*

$$\xi_{JK}|_{v_J}\circ t_J|_{v_J}=t_K|_{v_K}\circ\chi_{JK}|_{v_J}\colon E_J|_{v_J}\to F_K|_{v_K}.$$

If $v_J\in s_J^{-1}(0)\subseteq V_J$ with $\phi_{JK}(v_J)=v_K\in s_K^{-1}(0)\subseteq V_K$ and $\psi_J(v_J)=\psi_K(v_K)=x\in X_{\mathrm{an}}$, then (28)–(31) commute.

If $\varnothing\neq L\subseteq K\subseteq J\subseteq I$ then $\phi_{JL}=\phi_{KL}\circ\phi_{JK}$, $\chi_{JL}=\phi_{JK}^*(\chi_{KL})\circ\chi_{JK}$, and $\xi_{JL}=\phi_{JK}^*(\xi_{KL})\circ\xi_{JK}$.

3.3 Subbundles $E^- \subseteq E$ and Kuranishi neighbourhoods

Throughout Sections 3.3–3.6, when we apply Theorem 3.1 we take $B = \mathbb{C}$, so that Z is the point $* = \operatorname{Spec}\mathbb{C}$, and the data $\boldsymbol{\pi}$, β_i, β_J, τ_J is trivial, so we omit it.

Suppose $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ is a -2–shifted symplectic derived $\mathbb{C}$–scheme, $A^\bullet$ a standard form cdga over $\mathbb{C}$, and $\boldsymbol{\alpha}\colon \mathbf{Spec}\, A^\bullet \to \boldsymbol{X}$ a Zariski open inclusion. Then Definition 3.2 defines complex geometric data V, E, F, s, t, ψ, R, such that (V, E, s, ψ) is a Kuranishi neighbourhood on the topological space X_{an} of $\boldsymbol{X}$.

However these are not the Kuranishi neighbourhoods we want: they depend only on $\boldsymbol{X}$, not on $\omega_{\boldsymbol{X}}^*$, and in general two such neighbourhoods (V_J, E_J, s_J, ψ_J) and (V_K, E_K, s_K, ψ_K) are not compatible over their intersection $R_J \cap R_K$ in X_{an} (eg the virtual dimensions $\dim_{\mathbb{R}} V_J - \operatorname{rank}_{\mathbb{R}} E_J$ and $\dim_{\mathbb{R}} V_K - \operatorname{rank}_{\mathbb{R}} E_K$ may be different), so we cannot glue them to make X_{an} into a derived manifold.

The basic problem is that the rank of E may be too large; for instance, we can modify $A^\bullet$ to replace E, F, s, t by $\tilde{E} = E \oplus G$, $\tilde{F} = F \oplus G$, $\tilde{s} = s \oplus 0$, $\tilde{t} = t \oplus \mathrm{id}_G$ for some holomorphic vector bundle $G \to V$. Our solution is to choose a real vector subbundle $E^- \subseteq E$ satisfying some conditions involving $\omega_{\boldsymbol{X}}^*$, and set $E^+ = E/E^-$ to be the quotient bundle and $s^+ = s + E^-$ in $C^\infty(E^+)$ to be the quotient section. The conditions on E^- imply that $s^{-1}(0) = (s^+)^{-1}(0)$, so (V, E^+, s^+, ψ^+) is also a Kuranishi neighbourhood on X_{an}. Under good conditions we can make two such $(V_J, E_J^+, s_J^+, \psi_J^+)$, $(V_K, E_K^+, s_K^+, \psi_K^+)$ compatible over $R_J \cap R_K$, and glue these local models to make X_{an} into a derived manifold.

We define the class of subbundles $E^- \subseteq E$ we are interested in:

Definition 3.6 Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme with virtual dimension $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n$, and suppose $A^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ is of standard form and $\boldsymbol{\alpha}\colon A^\bullet \hookrightarrow \boldsymbol{X}$ is a Zariski open inclusion. Define complex geometric data V, E, F, s, t and $\psi\colon s^{-1}(0) \xrightarrow{\cong} R \subseteq X_{\mathrm{an}}$ as in Definition 3.2, and suppose $R \neq \varnothing$. Then for each $v \in s^{-1}(0)$ with $\psi(v) = x \in X_{\mathrm{an}}$, (20) gives an isomorphism from a vector space depending on V, E, F, s, t at v to $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$.

Equation (6) defined a quadratic form Q_x on $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$. Define

$$\tilde{Q}_v\colon \frac{\operatorname{Ker}(t|_v\colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v\colon T_vV \to E|_v)} \times \frac{\operatorname{Ker}(t|_v\colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v\colon T_vV \to E|_v)} \to \mathbb{C} \tag{32}$$

to be the nondegenerate complex quadratic form identified with Q_x in (6) by the isomorphism $H^1(\mathbb{T}_{\boldsymbol{\alpha}}|_v)$ in (20).

Consider pairs (U, E^-), where $U \subseteq V$ is open and E^- is a real vector subbundle of $E|_U$. Given such (U, E^-), we write $E^+ = E|_U / E^-$ for the quotient vector bundle over U, and $s^+ \in C^\infty(E^+)$ for the image of $s|_U$ under the projection $E|_U \to E^+$, and $\psi^+ := \psi|_{s^{-1}(0)\cap U} \colon s^{-1}(0) \cap U \to X_{\mathrm{an}}$. We say that (U, E^-) *satisfies condition* $(*)$ if:

$(*)$ For each $v \in s^{-1}(0) \cap U$, we have

$$\operatorname{Im}(\mathrm{d}s|_v \colon T_v V \to E|_v) \cap E^-|_v = \{0\} \qquad \text{in } E|_v, \tag{33}$$

$$t|_v(E^-|_v) = t|_v(E|_v) \quad \text{in } F|_v, \tag{34}$$

and the natural real linear map

$$\Pi_v \colon E^-|_v \cap \operatorname{Ker}(t|_v \colon E|_v \to F|_v) \to \frac{\operatorname{Ker}(t|_v \colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v \colon T_v V \to E|_v)}, \tag{35}$$

which is injective by (33), has image $\operatorname{Im} \Pi_v$ a real vector subspace of dimension exactly half the real dimension of $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)$, and the real quadratic form $\operatorname{Re} \widetilde{Q}_v$ on $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)$ from (32) restricts to a negative definite real quadratic form on $\operatorname{Im} \Pi_v$.

We say (U, E^-) *satisfies condition* $(\dagger)$ if

$(\dagger)$ (U, E^-) satisfies condition $(*)$ and $s^{-1}(0) \cap U = (s^+)^{-1}(0) \subseteq U$.

In this case, (U, E^+, s^+, ψ^+) is a Kuranishi neighbourhood on X_{an}.

Observe that if $v \in s^{-1}(0) \cap U$ with $\psi(v) = x \in X_{\mathrm{an}}$ then using (19)–(20) and (33)–(35) we find there is an exact sequence

$$0 \to H^l(\mathbb{T}_{\boldsymbol{X}}|_x) \to T_v U \to E^+|_v \to H^l(\mathbb{T}_{\boldsymbol{X}}|_r)/\operatorname{Im} \Pi_v \to 0. \tag{36}$$

Hence

$$\begin{aligned} \dim_{\mathbb{R}} U - \operatorname{rank}_{\mathbb{R}} E^+ &= \dim_{\mathbb{R}} H^0(\mathbb{T}_{\boldsymbol{X}}|_x) - \dim_{\mathbb{R}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x) + \dim_{\mathbb{R}} \operatorname{Im} \Pi_v \\ &= 2 \dim_{\mathbb{C}} H^0(\mathbb{T}_{\boldsymbol{X}}|_x) - \dim_{\mathbb{C}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x) \\ &= \dim_{\mathbb{C}} H^0(\mathbb{T}_{\boldsymbol{X}}|_x) - \dim_{\mathbb{C}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x) + \dim_{\mathbb{C}} H^2(\mathbb{T}_{\boldsymbol{X}}|_x) \\ &= \operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n. \end{aligned} \tag{37}$$

Here in the second step we use $\dim_{\mathbb{R}} \Pi_v = \frac{1}{2} \dim_{\mathbb{R}} H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ by $(*)$ and (20), in the third that $H^0(\mathbb{T}_{\boldsymbol{X}}|_x) \cong H^2(\mathbb{T}_{\boldsymbol{X}}|_x)^*$ as $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ is -2–shifted symplectic (or -2–shifted presymplectic will do), and in the fourth that $\mathbb{T}_{\boldsymbol{X}}$ is perfect in the interval $[0, 2]$ as $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ is -2–shifted symplectic (or presymplectic).

Equation (37) says that the Kuranishi neighbourhood (U, E^+, s^+, ψ^+) has real virtual dimension $\dim U - \operatorname{rank} E^+ = n = \operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = \frac{1}{2} \operatorname{vdim}_{\mathbb{R}} \boldsymbol{X}$. Note that this is half the virtual dimension we might have expected, and the real virtual dimension can be odd, even though X, V, E, $s,\dots$ are all complex.

Here are some important properties of such U, E^-, E^+, s^+, proved in Section 5.

Theorem 3.7 *In the situation of Definition 3.6, with $\boldsymbol{X}$, $\omega_{\boldsymbol{X}}^*$, $A^\bullet$, $\boldsymbol{\alpha}$, V, E, F, s, t, ψ fixed, we have:*

(a) *If the conditions in $(*)$ hold at some $v\in s^{-1}(0)\cap U$, then they also hold for all v' in an open neighbourhood of v in $s^{-1}(0)\cap U$.*

(b) *Suppose $C\subseteq V$ is closed, and (U,E^-) satisfies condition $(*)$ with $C\subseteq U\subseteq V$. (We allow $C=U=\varnothing$.) Then there exists $(\tilde U,\tilde E^-)$ satisfying $(*)$ with $C\cup s^{-1}(0)\subseteq\tilde U\subseteq V$, and an open neighbourhood U' of C in $U\cap\tilde U$ such that $E^-|_{U'}=\tilde E^-|_{U'}$.*

(c) *If (U,E^-) satisfies $(*)$, the closed subsets $s^{-1}(0)\cap U$ and $(s^+)^{-1}(0)$ in $U\subseteq V$ coincide in an open neighbourhood U' of $s^{-1}(0)\cap U$ in U. Hence $(U',E^-|_{U'})$ satisfies condition $(\dagger)$, and $(U',E^+|_{U'},s^+|_{U'},\psi^+)$ is a Kuranishi neighbourhood on X_{an}. Thus, we can make (U,E^-) satisfying $(*)$ also satisfy $(\dagger)$ by shrinking U, without changing $R=\operatorname{Im}\psi$ in X_{an}.*

The next example proves Theorem 3.7(c) near $v\in s^{-1}(0)\cap U$ in a special case, when $(A^\bullet,\omega_{A^\bullet})$ is in -2–Darboux form and minimal at v. The general case in Section 5.3 is proved by reducing to Example 3.8.

Example 3.8 Suppose that $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ is a -2–shifted symplectic derived $\mathbb{C}$–scheme and that $x\in X_{\mathrm{an}}$. Then Theorem 2.10 gives a pair $(A^\bullet,\omega_{A^\bullet})$ in -2–Darboux form and a Zariski open inclusion $\boldsymbol{\alpha}\colon \mathbf{Spec}\,A^\bullet\hookrightarrow\boldsymbol{X}$ which is minimal at $x\in\operatorname{Im}\boldsymbol{\alpha}$, with $\boldsymbol{\alpha}^*(\omega_{\boldsymbol{X}}^*)\simeq\omega_{A^\bullet}$ in $\mathcal{A}^{2,\mathrm{cl}}_{\mathbb{C}}(\mathbf{Spec}\,A^\bullet,-2)$.

Example 3.3 describes the data V, E, F, s, t associated to $A^\bullet$ in Section 3.2, and defines a nondegenerate quadratic form $Q\in H^0(S^2E^*)$ with $Q(s,s)=0$ using $\omega_{A^\bullet}$. As $x\in\operatorname{Im}\boldsymbol{\alpha}$ there is $v\in s^{-1}(0)\subseteq V$ with $\boldsymbol{\alpha}(v)=x$, and $(A^\bullet,\boldsymbol{\alpha})$ minimal at x means that $\mathrm{d}s|_v=0$, so that $t|_v=0$ by (24). Thus in (20) we have $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)=E|_v$, identified with $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$. Since $\boldsymbol{\alpha}^*(\omega_{\boldsymbol{X}}^*)\simeq\omega_{A^\bullet}$, the quadratic form $\tilde Q_v$ on $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)=E|_v$ in (32) is $Q|_v$.

Given a pair (U,E^-) as in Definition 3.6 with $v\in U$, the map Π_v in (35) is just the inclusion $E^-|_v\hookrightarrow E|_v$. So $(*)$ at v says that $E^-|_v$ is a real vector subspace of $E|_v$ with $\dim_{\mathbb{R}}E^-|_v=\frac12\dim_{\mathbb{R}}E|_v=\dim_{\mathbb{C}}E|_v$, such that $\operatorname{Re}Q|_v$ is negative definite on $E^-|_v$.

As this is an open condition, there exists an open neighbourhood U' of v in U such that $\mathrm{Re}\, Q|_{U'}$ is negative definite on $E^-|_{U'}$. Define a real vector subbundle $\widetilde{E}^+$ of $E|_{U'}$ to be the orthogonal subbundle of $E^-|_{U'}$ with respect to the nondegenerate real quadratic form $\mathrm{Re}\, Q|_{U'}$. Then $E|_{U'} = \widetilde{E}^+ \oplus E^-|_{U'}$, so we can write $s|_{U'} = \widetilde{s}^+ \oplus s^-$, for $\widetilde{s}^+ \in C^\infty(\widetilde{E}^+)$ and $s^- \in C^\infty(E^-|_{U'})$. The projection $E|_{U'} \to E^+|_{U'} = E|_{U'}/E^-|_{U'}$ restricts to an isomorphism $\widetilde{E}^+ \to E^+|_{U'}$, which maps $\widetilde{s}^+ \mapsto s^+|_{U'}$.

Because $\mathrm{Re}\, Q$ is the real part of a complex form, it has the same number of positive as negative eigenvalues. Thus $\mathrm{Re}\, Q|_{U'}$ is positive definite on $\widetilde{E}^+$. Now

$$0 = \mathrm{Re}\, Q(s,s)|_{U'} = \mathrm{Re}\, Q(\widetilde{s}^+ + s^-, \widetilde{s}^+ + s^-) = \mathrm{Re}\, Q(\widetilde{s}^+, \widetilde{s}^+) + \mathrm{Re}\, Q(s^-, s^-), \tag{38}$$

using $\mathrm{Re}\, Q(\widetilde{s}^+, s^-) = 0$ as $\widetilde{E}^+$, $E^-|_{U'}$ are orthogonal with respect to $\mathrm{Re}\, Q|_{U'}$.

For each $u \in U'$, we now have

$$\begin{aligned} s^+(u) = 0 \iff{}& \quad \widetilde{s}^+(u) = 0 \quad \iff \mathrm{Re}\, Q(\widetilde{s}^+, \widetilde{s}^+)|_u = 0 \\ \iff{}& \mathrm{Re}\, Q(s^-, s^-)|_u = 0 \iff \widetilde{s}^+(u) = s^-(u) = 0 \iff s(u) = 0, \end{aligned}$$

using $\widetilde{E}^+ \to E^+|_{U'}$ an isomorphism mapping $\widetilde{s}^+ \mapsto s^+|_{U'}$ in the first step, $\mathrm{Re}\, Q$ positive definite on $\widetilde{E}^+$ in the second, (38) in the third, $\mathrm{Re}\, Q$ negative definite on $E^-|_{U'}$ in the fourth, and $s|_{U'} = \widetilde{s}^+ \oplus s^-$ in the fifth.

This proves there exists an open neighbourhood U' of v in U such that $s^{-1}(0) \cap U' = (s^+)^{-1}(0) \cap U'$, which is Theorem 3.7(c), except that U' is a neighbourhood of v rather than of $s^{-1}(0) \cap U$.

Remark 3.9 Pairs (U, E^-) satisfying (†) will be used to prove our main result, constructing a derived manifold structure X_{dm} on the complex analytic topological space X_{an} of a -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$.

Our construction apparently uses less than the full -2–shifted symplectic structure $\omega_{\boldsymbol{X}}^*$ on $\boldsymbol{X}$. In particular, conditions $(*)$ and (†) only involve the nondegenerate pairings $\omega_{\boldsymbol{X}}^0|_x$ on $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ in (6), which depend only on the presymplectic structure $\omega_{\boldsymbol{X}}^0$, not the symplectic structure $\omega_{\boldsymbol{X}}^* = (\omega_{\boldsymbol{X}}^0, \omega_{\boldsymbol{X}}^1, \dots)$. The proofs of Theorem 3.7(a),(b) in Sections 5.1–5.2 also use only $\omega_{\boldsymbol{X}}^0$ rather than $\omega_{\boldsymbol{X}}^*$.

However, the proof of Theorem 3.7(c) in Section 5.3 involves $\omega_{\boldsymbol{X}}^*$, as it uses the existence of a minimal -2–Darboux form presentation for $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ near each $x \in X_{\mathrm{an}}$, as in Theorem 2.10. The authors do not know whether Theorem 3.7(c) holds for -2–shifted presymplectic $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^0)$ which are not symplectic.

3.4 Comparing (U_J, E_J^-), (U_K, E_K^-) under Φ_{JK}

Section 3.3 discussed how to use standard form charts $\boldsymbol{\alpha}$: $\mathbf{Spec}\, A^\bullet \to \boldsymbol{X}$ on $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ to choose pairs (U, E^-), and so define Kuranishi neighbourhoods (U, E^+, s^+, ψ^+) on X_{an}. We now explain how to pull back such pairs (U_K, E_K^-) along a quasifree morphism Φ_{JK}: $A_K^\bullet \to A_J^\bullet$, and construct coordinate changes between the Kuranishi neighbourhoods $(U_J, E_J^+, s_J^+, \psi_J^+)$, $(U_K, E_K^+, s_K^+, \psi_K^+)$.

Definition 3.10 Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme with $\mathrm{vdim}_{\mathbb{C}}\, \boldsymbol{X} = n$, and suppose Φ_{JK}: $A_K^\bullet \to A_J^\bullet$ is a quasifree morphism of standard form cdgas over $\mathbb{C}$ and $\boldsymbol{\alpha}_J$: $\mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$, $\boldsymbol{\alpha}_K$: $\mathbf{Spec}\, A_K^\bullet \hookrightarrow \boldsymbol{X}$ are Zariski open inclusions such that (14) homotopy commutes. Define complex geometric data V_J, E_J, F_J, s_J, t_J, ψ_J, R_J, V_K, E_K, F_K, s_K, t_K, ψ_K, R_K, ϕ_{JK}, χ_{JK}, ξ_{JK} in Definitions 3.2 and 3.4, and suppose $R_J \neq \varnothing$, so $R_K \neq \varnothing$ as $R_J \subseteq R_K \subseteq X_{\mathrm{an}}$.

Consider pairs (U_J, E_J^-) for $A_J^\bullet$ and (U_K, E_K^-) for $A_K^\bullet$ satisfying condition $(*)$ in Definition 3.6. We say that (U_J, E_J^-) and (U_K, E_K^-) are *compatible* if $\phi_{JK}(U_J) \subseteq U_K$ and $\chi_{JK}|_{U_J}(E_J^-) \subseteq \phi_{JK}|_{U_J}^*(E_K^-) \subseteq \phi_{JK}|_{U_J}^*(E_K)$.

For compatible pairs (U_J, E_J^-) and (U_K, E_K^-), define a vector bundle morphism χ_{JK}^+: $E_J^+ \to \phi_{JK}|_{U_J}^*(E_K^+)$ on U_J by the commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & E_J^- \arrow[r] \arrow[d, "\chi_{JK}|_{E_J^-}"] & E_J|_{U_J} \arrow[r] \arrow[d, "\chi_{JK}|_{U_J}"] & E_J^+ \arrow[r] \arrow[d, dotted, "\chi_{JK}^+"] & 0 \\
0 \arrow[r] & \phi_{JK}|_{U_J}^*(E_K^-) \arrow[r] & \phi_{JK}|_{U_J}^*(E_K) \arrow[r] & \phi_{JK}|_{U_J}^*(E_K^+) \arrow[r] & 0
\end{tikzcd}$$

Let $v_J \in s_J^{-1}(0) \subseteq U_J \subseteq V_J$ with $\phi_{JK}(v_J) = v_K \in s_K^{-1}(0) \subseteq U_K \subseteq V_K$ and $\psi_J(v_J) = \psi_K(v_K) = x \in X_{\mathrm{an}}$. Consider the diagram, with rows (36) for (U_J, E_J^-), v_J and (U_K, E_K^-), v_K:

$$\begin{tikzcd}
0 \arrow[r] & H^0(\mathbb{T}_{\boldsymbol{X}}|_x) \arrow[r] \arrow[d, "\mathrm{id}"] & T_{v_J}U_J \arrow[r, "\mathrm{d}s_J^+|_{v_J}"'] \arrow[d, "\mathrm{d}\phi_{JK}|_{v_J}"'] & E_J^+|_{v_J} \arrow[r] \arrow[d, "\chi_{JK}^+|_{v_J}"] & H^1(\mathbb{T}_{\boldsymbol{X}}|_x)/\operatorname{Im}\Pi_{v_J} \arrow[r] \arrow[d, "\mathrm{id}"] & 0 \\
0 \arrow[r] & H^0(\mathbb{T}_{\boldsymbol{X}}|_x) \arrow[r] & T_{v_K}U_K \arrow[r, "\mathrm{d}s_K^+|_{v_K}"] & E_K^+|_{v_K} \arrow[r] & H^1(\mathbb{T}_{\boldsymbol{X}}|_x)/\operatorname{Im}\Pi_{v_K} \arrow[r] & 0
\end{tikzcd} \tag{39}$$

Here if we regard $\operatorname{Im}\Pi_{v_J}$, $\operatorname{Im}\Pi_{v_K}$ from (35) as subspaces of $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ using (20), compatibility $\chi_{JK}(E_J^-|_{v_J}) \subseteq E_K^-|_{v_K}$ and (29) imply that $\operatorname{Im}\Pi_{v_J} \subseteq \operatorname{Im}\Pi_{v_K}$, so $\operatorname{Im}\Pi_{v_J} = \operatorname{Im}\Pi_{v_K}$ as they have the same dimension by $(*)$, and the right-hand column of (39) makes sense. From (25), (28) and (29) we see that (39) commutes. Elementary linear algebra then gives an exact sequence

$$(40)\quad 0\to T_{v_J}U_J \xrightarrow{\mathrm{d}s_J^+|_{v_J}\oplus \mathrm{d}\phi_{JK}|_{v_J}} E_J^+|_{v_J}\oplus T_{v_K}U_K \xrightarrow{-\chi_{JK}^+|_{v_J}\oplus \mathrm{d}s_K^+|_{v_K}} E_K^+|_{v_K}\to 0.$$

From (40) and Definition 2.14, we deduce:

Corollary 3.11 *In the situation of Definition 3.10, if (U_J,E_J^-) and (U_K,E_K^-) are compatible and satisfy (†) then, in the sense of Section 2.5,*

$$(U_J,\phi_{JK}|_{U_J},\chi_{JK}^+)\colon (U_J,E_J^+,s_J^+,\psi_J)\to (U_K,E_K^+,s_K^+,\psi_K)$$

is a coordinate change of Kuranishi neighbourhoods on X_{an}.

Lemma 3.12 *In the situation of Definition 3.10, fix (U_K,E_K^-) satisfying $(*)$ for $A_K^\bullet$, $\boldsymbol{\alpha}_K$. Set $U'_{JK}=\phi_{JK}^{-1}(U_K)\subseteq V_J$. Then $E'_{JK}:=\chi_{JK}|_{U'_{JK}}^{-1}(E_K^-)$ is a vector subbundle of $E_J|_{U'_{JK}}$, as χ_{JK} is surjective. Choose a complementary real vector subbundle E''_{JK}, so that $E_J|_{U'_{JK}}=E'_{JK}\oplus E''_{JK}$.*

Choose a connection ∇ on E_J, so that $\nabla s_J\colon TV_J\to E_J$ is a vector bundle morphism. Now $\mathrm{Ker}(\mathrm{d}\phi_{JK}\colon TV_J\to \phi_{JK}^*(TV_K))$ is a vector subbundle of TV_J, as $\mathrm{d}\phi_{JK}$ is surjective, and ∇s_J is injective on $\mathrm{Ker}\,\mathrm{d}\phi_{JK}$ near $s_J^{-1}(0)$, so $E'''_{JK}:=(\nabla s_J)[\mathrm{Ker}\,\mathrm{d}\phi_{JK}]$ is a vector subbundle of E_J near $s_J^{-1}(0)$ in V_J.

Then (U_J,E_J^-) satisfies $(*)$ for $A_J^\bullet$, $\boldsymbol{\alpha}_J$ and is compatible with (U_K,E_K^-) if and only if U_J is open in U'_{JK}, and E_{JK}^- is a vector subbundle of $E'_{JK}|_{U_J}$ satisfying $E_J|_{U_J}=E_{JK}^-\oplus E''_{JK}|_{U_J}\oplus E'''_{JK}|_{U_J}$ near $s_J^{-1}(0)\cap U_J$ in U_J. Alternatively, identifying E'_{JK} with $E_J|_{U'_{JK}}/E''_{JK}$, this condition may be written as $E'_{JK}|_{U_J}=E_{JK}^-\oplus[(E''_{JK}\oplus E'''_{JK})/E''_{JK}]|_{U_J}$ near $s_J^{-1}(0)\cap U_J$.

Proof We deduce ∇s_J is injective on $\mathrm{Ker}\,\mathrm{d}\phi_{JK}$ at $v_J\in s_J^{-1}(0)$ using (28), check that $(*)$ for U_J, E_J^- is equivalent to $E_J=E_{JK}^-\oplus E''_{JK}\oplus E'''_{JK}$ at each $v_J\in s_J^{-1}(0)$, and note that both are open conditions. □

Lemma 3.12 shows we can always *pull back* (U_K,E_K^-) satisfying $(*)$ along submersions $\phi_{JK}\colon V_J\to V_K$: we just have to choose a complement E_J^- to $(E''_{JK}\oplus E'''_{JK})/E''_{JK}$ in E'_{JK} on some small open neighbourhood U_J of $s_J^{-1}(0)$ in U'_{JK}, for instance, the orthogonal complement with respect to any metric on E'_{JK}. By Theorem 3.7(c), making U_J smaller, we can suppose (U_J,E_J^-) satisfies (†).

3.5 Constructing Kuranishi atlases and derived manifolds

Let $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme with $\mathrm{vdim}_{\mathbb{C}}\,\boldsymbol{X}=n$ in $\mathbb{Z}$, and write X_{an} for the complex analytic topological space. Suppose $\boldsymbol{X}$ is separated and

X_{an} is a paracompact topological space. (Paracompactness is automatic if $\boldsymbol{X}$ is proper, or quasicompact, or of finite type, or if X_{an} is second countable.) We will construct a Kuranishi atlas on X_{an}, in the sense of Section 2.5.

First choose a family $\{(A_i^\bullet, \boldsymbol{\alpha}_i) \mid i \in I\}$, where $A_i^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ is a standard form cdga, and $\boldsymbol{\alpha}_i \colon \mathbf{Spec}\, A_i^\bullet \hookrightarrow \boldsymbol{X}$ a Zariski open inclusion in $\mathbf{dSch}_{\mathbb{C}}$ for each i in I, an indexing set, such that $\{R_i := (\mathrm{Im}\, \boldsymbol{\alpha}_i)_{\mathrm{an}} \mid i \in I\}$ is an open cover of the complex analytic topological space X_{an}. This is possible by Theorem 2.5. If $\boldsymbol{X}$ is quasicompact (since $\boldsymbol{X}$ is locally of finite type, this is equivalent to $\boldsymbol{X}$ being of finite type) then we can take I to be finite.

Apply Theorem 3.1 to get data $A_J^\bullet \in \mathbf{cdga}_{\mathbb{C}}$, $\boldsymbol{\alpha}_J \colon \mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$ for finite $\varnothing \neq J \subseteq I$ and quasifree $\Phi_{JK} \colon A_K^\bullet \to A_J^\bullet$, for all finite $\varnothing \neq K \subseteq J \subseteq I$.

Use the notation of Section 3.2 to rewrite $A_J^\bullet$, Φ_{JK} in terms of complex geometry. As in Corollary 3.5, this gives data V_J, E_J, F_J, s_J, t_J, ψ_J, R_J for all finite $\varnothing \neq J \subseteq I$, and ϕ_{JK}, χ_{JK}, ξ_{JK} for all finite $\varnothing \neq K \subseteq J \subseteq I$.

For brevity we write $A = \{J \mid \varnothing \neq J \subseteq I \text{ and } J \text{ is finite}\}$. The proof of the next result in Section 6.1 is based on McDuff and Wehrheim [29, Lemma 7.1.7].

Proposition 3.13 *Suppose Z is a paracompact, Hausdorff topological space and $\{R_i \mid i \in I\}$ an open cover of Z. Then we can choose closed subsets $C_J \subseteq Z$ for all finite $\varnothing \neq J \subseteq I$, satisfying:*

(i) $C_J \subseteq \bigcap_{i \in J} R_i$ *for all* J.

(ii) *Each $z \in Z$ has an open neighbourhood $U_z \subseteq Z$ with $U_z \cap C_J \neq \varnothing$ for only finitely many J.*

(iii) $C_J \cap C_K \neq \varnothing$ *only if* $J \subseteq K$ *or* $K \subseteq J$.

(iv) $\bigcup_{\varnothing \neq J \subseteq I \text{ finite}} C_J = Z$.

In our case, X_{an} is Hausdorff and second countable. It is also locally compact, as it is locally homeomorphic to closed subsets $s_J^{-1}(0)$ of complex manifolds V_J. But Hausdorff, locally compact and second countable imply that X is paracompact and normal. Thus Proposition 3.13 applies to $Z = X_{\mathrm{an}}$ with the open cover $\{R_i \mid i \in I\}$, and we can choose closed subsets $C_J \subseteq R_J = \bigcap_{i \in J} R_i \subseteq X_{\mathrm{an}}$ for all $J \in A$ satisfying conditions (i)–(iv).

The next proposition, proved in Section 6.2 using Theorem 3.7 and Lemma 3.12, chooses pairs (U_J, E_J^-) satisfying (†), as in Section 3.3, with (U_J, E_J^-), (U_K, E_K^-) compatible near $C_J \cap C_K$ under the quasifree morphism $\Phi_{JK} \colon A_K^\bullet \to A_J^\bullet$.

Proposition 3.14 *In the situation above, we can choose (U_J, E_J^-) satisfying condition (†) for $V_J, E_J, \dots$ for each $J \in A$, such that $\psi_J^{-1}(C_J) \subseteq U_J \subseteq V_J$, and setting $S_J = \psi_J(s_J^{-1}(0) \cap U_J)$ so that S_J is an open neighbourhood of C_J in X_{an}, then for all $J, K \in A$, we have $S_J \cap S_K \neq \varnothing$ only if $J \subseteq K$ or $K \subseteq J$, and if $K \subsetneq J$ then there exists open $U_{JK} \subseteq U_J$ with $s_J^{-1}(0) \cap U_{JK} = \psi_J^{-1}(S_J \cap S_K)$ such that $(U_{JK}, E_J^-|_{U_{JK}})$ is compatible with (U_K, E_K^-), in the sense of Section 3.4.*

We can now prove two of the central results of this paper.

Theorem 3.15 *Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme with complex virtual dimension $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X} = n$ in $\mathbb{Z}$, and write X_{an} for the set of $\mathbb{C}$–points of $X = t_0(\boldsymbol{X})$ with the complex analytic topology. Suppose that $\boldsymbol{X}$ is separated, and X_{an} is a paracompact topological space. Then we can construct a Kuranishi atlas $\mathcal{K}$ on X_{an} of real dimension n, in the sense of Section 2.5. If $\boldsymbol{X}$ is quasicompact (equivalently, of finite type) then we can take $\mathcal{K}$ to be finite.*

Proof In the discussion from the beginning of Section 3.5 up to Proposition 3.14, we have the following:

(i) A Hausdorff, paracompact topological space X_{an}.

(ii) An indexing set I, where we write $A = \{J \mid \varnothing \neq J \subseteq I \text{ and } J \text{ is finite}\}$.

(iii) An open cover $\{S_J \mid J \in A\}$ of X_{an}, such that $S_J \cap S_K \neq \varnothing$ for $J, K \in A$ only if $J \subseteq K$ or $K \subseteq J$.

(iv) For each $J \in A$, a Kuranishi neighbourhood $(U_J, E_J^+, s_J^+, \psi_J^+)$ on X_{an} with $\dim U_J - \mathrm{rank}\, E_J^+ = n$, constructed as in Section 3.3 from (U_J, E_J^-) satisfying (†), with $\mathrm{Im}\, \psi_J^+ = S_J \subseteq X_{\mathrm{an}}$.

(v) For all $J, K \in A$ with $K \subsetneq J$, a coordinate change of Kuranishi neighbourhoods over $S_J \cap S_K$, as in Corollary 3.11,

$$(U_{JK}, \phi_{JK}|_{U_{JK}}, \chi_{JK}^+): (U_J, E_J^+, s_J^+, \psi_J^+) \to (U_K, E_K^+, s_K^+, \psi_K^+),$$

since $(U_{JK}, E_J^-|_{U_{JK}})$ is compatible with (U_K, E_K^-).

(vi) For all $J, K, L \in A$ with $L \subsetneq K \subsetneq J$, Corollary 3.5 implies that $\phi_{JL} = \phi_{KL} \circ \phi_{JK}$ and $\chi_{JL}^+ = \phi_{JK}^*(\chi_{KL}^+) \circ \chi_{JK}^+$ on $U_{JK} \cap U_{JL} \cap \phi_{JK}^{-1}(U_{KL})$.

This is a Kuranishi atlas $\mathcal{K}$ in the sense of Definition 2.15, where the partial order $\prec$ on A is $J \prec K$ if $K \subsetneq J$. If $\boldsymbol{X}$ is quasicompact then we can take I finite, so A and $\mathcal{K}$ are finite. □

Combining Theorems 2.18 and 3.15 yields:

Theorem 3.16 *Let $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ be a -2–shifted symplectic derived $\mathbb{C}$–scheme with complex virtual dimension $\mathrm{vdim}_{\mathbb{C}}\,\boldsymbol{X}=n$ in $\mathbb{Z}$, and write X_{an} for the set of $\mathbb{C}$–points of $X=t_0(\boldsymbol{X})$ with the complex analytic topology. Suppose that $\boldsymbol{X}$ is separated, so that X_{an} is Hausdorff, and also that X_{an} is a second countable topological space, which holds if and only if X admits a Zariski open cover $\{X_c \mid c\in C\}$ with C countable and each X_c a finite type $\mathbb{C}$–scheme.*

Then we can make the topological space X_{an} into a derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ with real virtual dimension $\mathrm{vdim}_{\mathbb{R}}\,\boldsymbol{X}_{\mathrm{dm}}=n$, in any of the senses (a) *Joyce's m-Kuranishi spaces* **mKur** [21, Section 4.7], (b) *Joyce's d-manifolds* **dMan** [18; 19; 20], (c) *Borisov and Noël's derived manifolds* $\mathbf{DerMan}_{\mathrm{BoNo}}$ [3; 4], *or* (d) *Spivak's derived manifolds* $\mathbf{DerMan}_{\mathrm{Spi}}$ [32], *all discussed in Section 2.6.*

We will discuss the dependence of $\boldsymbol{X}_{\mathrm{dm}}$ on choices made in the constructions in Section 3.6. Note that $\boldsymbol{X}_{\mathrm{dm}}$ in Theorem 3.16 has dimension $\mathrm{vdim}_{\mathbb{R}}\,\boldsymbol{X}_{\mathrm{dm}}=\mathrm{vdim}_{\mathbb{C}}\,\boldsymbol{X}=\frac{1}{2}\,\mathrm{vdim}_{\mathbb{R}}\,\boldsymbol{X}$, which is exactly half what we might have expected.

3.6 Orientations, bordism classes and virtual classes

Work in the situation of Theorems 3.15 and 3.16, so that we have a -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ with complex analytic topological space X_{an}, a Kuranishi atlas $\mathcal{K}$ on X_{an}, and a derived manifold $\boldsymbol{X}_{\mathrm{dm}}$. The next proposition, proved in Section 6.3, justifies our notions of orientation in Sections 2.4–2.6.

Proposition 3.17 *In the situation of Theorems 3.15 and 3.16, there are canonical one-to-one correspondences between*

(a) *orientations on $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ in the sense of Section 2.4;*

(b) *orientations on $(X_{\mathrm{an}},\mathcal{K})$ in the sense of Section 2.5; and*

(c) *orientations on $\boldsymbol{X}_{\mathrm{dm}}$ in the sense of Section 2.6.2.*

Next we consider how the derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ in Theorem 3.16 depends on choices made in the construction. Once we have chosen the Kuranishi atlas $\mathcal{K}$ in Theorem 3.15, Theorem 2.18 shows that $\boldsymbol{X}_{\mathrm{dm}}$ is determined uniquely up to equivalence in its 2–category or ∞–category. However, constructing $\mathcal{K}$ involves many arbitrary choices, and the next proposition, proved in Section 6.4 using the material of Section 3.7, explains how $\boldsymbol{X}_{\mathrm{dm}}$ depends on these.

Proposition 3.18 *In the situation of Theorem 3.16, for $(\boldsymbol{X},\omega_{\boldsymbol{X}}^*)$ and n fixed, the derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ depends on choices made in the construction only up to bordisms of derived manifolds which fix the underlying topological space X_{an}.*

That is, if $\boldsymbol{X}_{\mathrm{dm}}$, $\boldsymbol{X}'_{\mathrm{dm}}$ *are possible derived manifolds in Theorem 3.16, then we can construct a derived manifold with boundary* $\boldsymbol{W}_{\mathrm{dm}}$ *with topological space* $X_{\mathrm{an}}\times[0,1]$ *and* $\operatorname{vdim}\boldsymbol{W}_{\mathrm{dm}}=n+1$, *and an equivalence of derived manifolds* $\partial\boldsymbol{W}_{\mathrm{dm}}\simeq\boldsymbol{X}_{\mathrm{dm}}\amalg\boldsymbol{X}'_{\mathrm{dm}}$, *topologically identifying* $\boldsymbol{X}_{\mathrm{dm}}$ *with* $X_{\mathrm{an}}\times\{0\}$ *and* $\boldsymbol{X}'_{\mathrm{dm}}$ *with* $X_{\mathrm{an}}\times\{1\}$. *We regard* $\boldsymbol{W}_{\mathrm{dm}}$ *as a bordism from* $\boldsymbol{X}_{\mathrm{dm}}$ *to* $\boldsymbol{X}'_{\mathrm{dm}}$.

This bordism $\boldsymbol{W}_{\mathrm{dm}}$ *is compatible with orientations in Proposition 3.17. That is, given an orientation on* $(\boldsymbol{X},\omega^*_{\boldsymbol{X}})$, *we get natural orientations on* $\boldsymbol{X}_{\mathrm{dm}}$, $\boldsymbol{X}'_{\mathrm{dm}}$, $\boldsymbol{W}_{\mathrm{dm}}$, *and an equivalence of oriented derived manifolds* $\partial\boldsymbol{W}_{\mathrm{dm}}\simeq-\boldsymbol{X}_{\mathrm{dm}}\amalg\boldsymbol{X}'_{\mathrm{dm}}$, *where* $-\boldsymbol{X}_{\mathrm{dm}}$ *is* $\boldsymbol{X}_{\mathrm{dm}}$ *with the opposite orientation.*

Combining this with material in Sections 2.6.4–2.6.5 yields:

Corollary 3.19 *Suppose* $(\boldsymbol{X},\omega^*_{\boldsymbol{X}})$ *is a proper* -2*–shifted symplectic derived* $\mathbb{C}$*–scheme, with* $\operatorname{vdim}_{\mathbb{C}}\boldsymbol{X}=n$, *and with an orientation in the sense of Section 2.4. Then Theorem 3.16 constructs a compact derived manifold* $\boldsymbol{X}_{\mathrm{dm}}$ *with* $\operatorname{vdim}_{\mathbb{R}}\boldsymbol{X}_{\mathrm{dm}}=n$, *and Proposition 3.17 defines an orientation on* $\boldsymbol{X}_{\mathrm{dm}}$.

Although $\boldsymbol{X}_{\mathrm{dm}}$ *depends on arbitrary choices, the d-bordism class* $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{dbo}}$ *in* $B_n(*)$ *from Section 2.6.4 and the virtual class* $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{virt}}$ *in* $H_n(X_{\mathrm{an}};\mathbb{Z})$ *from Section 2.6.5 are independent of these, and depend only on* $(\boldsymbol{X},\omega^*_{\boldsymbol{X}})$ *and its orientation.*

3.7 Working relative to a smooth base $\mathbb{C}$–scheme Z

Let $Z=\operatorname{Spec}B$ be a smooth classical affine $\mathbb{C}$–scheme, which we now assume is connected. Then the set Z_{an} of $\mathbb{C}$–points of Z is a complex manifold, and hence a real manifold. In this section we will show that all of Sections 3.1–3.6 also works relatively over the base Z. To do this, we will need a notion of a family $(\boldsymbol{\pi}\colon\boldsymbol{X}\to Z,\omega_{\boldsymbol{X}/Z})$ of -2–shifted symplectic derived $\mathbb{C}$–schemes over the base Z.

To understand the next definition, recall from Remark 3.9 that if $(\boldsymbol{X},\omega^*_{\boldsymbol{X}})$ is -2–shifted symplectic, then the derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ constructed in Section 3.5 does not depend on the whole sequence $\omega^*_{\boldsymbol{X}}=(\omega^0_{\boldsymbol{X}},\omega^1_{\boldsymbol{X}},\dots)$, but only on the nondegenerate pairings $\omega^0_{\boldsymbol{X}}|_x$ on $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ for $x\in X_{\mathrm{an}}$, and therefore only on the cohomology class $[\omega^0_{\boldsymbol{X}}]\in H^{-2}(\mathbb{L}_{\boldsymbol{X}})$. We require that choices of $\omega^1_{\boldsymbol{X}},\omega^2_{\boldsymbol{X}},\dots$ should exist (they are needed to apply Theorem 2.10, which is used in the proof of Theorem 3.7(c)), but $\boldsymbol{X}_{\mathrm{dm}}$ does not depend on them.

Definition 3.20 Let $\boldsymbol{X}$ be a derived $\mathbb{C}$–scheme, $Z=\operatorname{Spec}B$ a smooth, connected, classical affine $\mathbb{C}$–scheme, and $\boldsymbol{\pi}\colon\boldsymbol{X}\to Z$ a morphism. A *family of* -2*–shifted symplectic structures on* $\boldsymbol{X}/Z$ is $[\omega_{\boldsymbol{X}/Z}]\in H^{-2}(\mathbb{L}_{\boldsymbol{X}/Z})$, such that if $z\in Z_{\mathrm{an}}$, writing

$\boldsymbol{X}^z = \boldsymbol{\pi}^{-1}(z) = \boldsymbol{X} \times^h_{\boldsymbol{\pi},Z,z} *$ for the fibre of $\boldsymbol{\pi}$ over z and $[\omega_{\boldsymbol{X}/Z}]|_{\boldsymbol{X}^z} \in H^{-2}(\mathbb{L}_{\boldsymbol{X}^z})$ for the restriction of $[\omega_{\boldsymbol{X}/Z}]$ to $\boldsymbol{X}^z$, then there should exist a -2–shifted symplectic structure $\omega^*_{\boldsymbol{X}^z} = (\omega^0_{\boldsymbol{X}^z}, \omega^1_{\boldsymbol{X}^z}, \dots)$ on $\boldsymbol{X}^z$ such that $[\omega_{\boldsymbol{X}/Z}]|_{\boldsymbol{X}^z} = [\omega^0_{\boldsymbol{X}^z}]$ in $H^{-2}(\mathbb{L}_{\boldsymbol{X}^z})$.

That is, a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$ is a -2–shifted relative 2–form $[\omega_{\boldsymbol{X}/Z}]$ on $\boldsymbol{X}/Z$, which on each fibre $\boldsymbol{X}^z$ extends to a closed 2–form which is -2–shifted symplectic. We will explain how to extend the arguments of Sections 3.3–3.6 to the relative case. Here is the analogue of Definition 3.6:

Definition 3.21 Let $\boldsymbol{X}$ be a derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ a smooth, classical, affine $\mathbb{C}$–scheme of pure dimension, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ a morphism, and $[\omega_{\boldsymbol{X}/Z}]$ in $H^{-2}(\mathbb{L}_{\boldsymbol{X}/Z})$ a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$. Write $\dim_{\mathbb{C}} Z = k$ and $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n + k$. Suppose $A^\bullet \in \mathbf{cdga}_{\mathbb{C}}$ is of standard form, $\boldsymbol{\alpha}\colon A^\bullet \hookrightarrow \boldsymbol{X}$ is a Zariski open inclusion, and $\beta\colon B \to A^0$ is a smooth morphism of $\mathbb{C}$–algebras, such that (21) homotopy commutes. Define complex geometric data V, τ, E, F, s, t and $\psi\colon s^{-1}(0) \xrightarrow{\cong} R \subseteq X_{\mathrm{an}}$ as in Definition 3.2, and suppose $R \neq \varnothing$. Then for each $v \in s^{-1}(0)$ with $\psi(v) = x \in X_{\mathrm{an}}$ and $\tau(v) = \pi(x) = z \in Z_{\mathrm{an}}$, (23) gives an isomorphism from a vector space depending on V, τ, Z_{an}, E, F, s, t, τ at v to $H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x)$.

As in (6), the relative 2–form $[\omega_{\boldsymbol{X}/Z}]$ induces a pairing

$$H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x) \times H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x) \xrightarrow{Q_x := \omega^0_{\boldsymbol{X}/Z}|_x \cdot} \mathbb{C}, \tag{41}$$

which is nondegenerate because Q_x, under the equivalence $\mathbb{T}_{\boldsymbol{X}/Z}|_x \simeq \mathbb{T}_{\boldsymbol{X}^z}|_x$, is identified with the pairing induced by a -2–shifted symplectic form $\omega^*_{\boldsymbol{X}^z}$ on $\boldsymbol{X}^z$, as in Definition 3.20. Define

$$\widetilde{Q}_v\colon \frac{\operatorname{Ker}(t|_v\colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v)} \times \frac{\operatorname{Ker}(t|_v\colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v)} \to \mathbb{C} \tag{42}$$

to be the nondegenerate complex quadratic form identified with Q_x in (41) by the isomorphism $H^1(\mathbb{T}_{\boldsymbol{\alpha}}|_v)$ in (23).

Consider pairs (U, E^-), where $U \subseteq V$ is open and E^- is a real vector subbundle of $E|_U$. Given such (U, E^-), we write $E^+ = E|_U / E^-$ for the quotient vector bundle over U, and $s^+ \in C^\infty(E^+)$ for the image of $s|_U$ under the projection $E|_U \to E^+$, and $\psi^+ := \psi|_{s^{-1}(0)\cap U}\colon s^{-1}(0) \cap U \to X_{\mathrm{an}}$. We say that (U, E^-) *satisfies condition* $(*)$ if

$(*)$ For each $v \in s^{-1}(0) \cap U$, we have

$$\operatorname{Im}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v) \cap E^-|_v = \{0\} \qquad \text{in } E|_v, \tag{43}$$

$$t|_v(E^-|_v) = t|_v(E|_v) \qquad \text{in } F|_v, \tag{44}$$

and the natural real linear map

$$\Pi_v\colon E^-|_v \cap \operatorname{Ker}(t|_v\colon E|_v \to F|_v) \to \frac{\operatorname{Ker}(t|_v\colon E|_v \to F|_v)}{\operatorname{Im}(\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}}) \to E|_v)}, \tag{45}$$

which is injective by (43), has image $\operatorname{Im}\Pi_v$ a real vector subspace of dimension exactly half the real dimension of $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)$, and the real quadratic form $\operatorname{Re}\widetilde{Q}_v$ on $\operatorname{Ker}(t|_v)/\operatorname{Im}(\mathrm{d}s|_v)$ from (42) restricts to a negative definite real quadratic form on $\operatorname{Im}\Pi_v$.

We say (U,E^-) *satisfies condition* $(\dagger)$ if

($\dagger$) (U,E^-) satisfies condition $(*)$ and $s^{-1}(0)\cap U=(s^+)^{-1}(0)\subseteq U$.

In this case, (U,E^+,s^+,ψ^+) is a Kuranishi neighbourhood on X_{an}.

Observe that if $v\in s^{-1}(0)\cap U$ with $\psi(v)=x\in X_{\mathrm{an}}$ then using (22)–(23) and (43)–(45) we find as for (36) that there is an exact sequence

$$0\to H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x)\to T_v(V/Z_{\mathrm{an}})\to E^+|_v\to H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x)/\operatorname{Im}\Pi_v\to 0. \tag{46}$$

Hence as for (37) we have

$$\begin{aligned}
\dim_{\mathbb{R}} U-\dim_{\mathbb{R}} Z_{\mathrm{an}}-\operatorname{rank}_{\mathbb{R}} E^+ &\\
&= \dim_{\mathbb{R}} H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x)-\dim_{\mathbb{R}} H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x)+\dim_{\mathbb{R}}\operatorname{Im}\Pi_v\\
&= 2\dim_{\mathbb{C}} H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x)-\dim_{\mathbb{C}} H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x)\\
&= \dim_{\mathbb{C}} H^0(\mathbb{T}_{\boldsymbol{X}/Z}|_x)-\dim_{\mathbb{C}} H^1(\mathbb{T}_{\boldsymbol{X}/Z}|_x)+\dim_{\mathbb{C}} H^2(\mathbb{T}_{\boldsymbol{X}/Z}|_x)\\
&= \operatorname{vdim}_{\mathbb{C}}\boldsymbol{X}-\dim_{\mathbb{C}} Z=n.
\end{aligned}$$

Thus the Kuranishi neighbourhood (U,E^+,s^+,ψ^+) has virtual dimension

$$\dim U-\operatorname{rank} E^+=n+2k=\tfrac{1}{2}(\operatorname{vdim}_{\mathbb{R}}\boldsymbol{X}-\dim_{\mathbb{R}} Z_{\mathrm{an}})+\dim_{\mathbb{R}} Z_{\mathrm{an}},$$

which is the real dimension of the base Z_{an}, plus half the real virtual dimension of the fibres $\boldsymbol{X}^z$.

Note that essentially the only important difference between Definitions 3.6 and 3.21 is that T_vV in (32), (33) and (35) is replaced by $T_v(V/Z_{\mathrm{an}})$ in (42), (43) and (45).

Theorem 3.22 *Theorem 3.7 holds with Definition 3.21 in place of Definition 3.6.*

Proof In the proofs of Theorem 3.7(a),(b) in Sections 5.1–5.2, we replace $\mathrm{d}s|_v\colon T_vV\to E|_v$ by $\mathrm{d}s|_v\colon T_v(V/Z_{\mathrm{an}})\to E|_v$ throughout, and no other changes are needed.

For part (c), fix $z \in Z_{\mathrm{an}}$, so that Definition 3.20 gives a -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X}^z, \omega^*_{\boldsymbol{X}^z})$ with $[\omega_{\boldsymbol{X}/Z}]|_{\boldsymbol{X}^z} = [\omega^0_{\boldsymbol{X}^z}]$ in $H^{-2}(\mathbb{L}_{\boldsymbol{X}^z})$. Consider the complex submanifolds $V^z = \tau^{-1}(z)$ in V and $U^z = U \cap V^z$ in U, and write E^z, F^z, s^z, t^z for the restrictions of E, F, s, t to V^z, and $E^{\pm z}$, s^{+z}, ψ^{+z} for the restrictions of $E^\pm$, s^+, ψ^+ to U^z. Then $(\boldsymbol{X}^z, \omega^*_{\boldsymbol{X}^z})$, V^z, $E^z, \ldots$ satisfy Definition 3.6, so Theorem 3.7(c) shows $(s^z)^{-1}(0) \cap U^z$ and $(s^{+z})^{-1}(0)$ coincide near $(s^z)^{-1}(0) \cap U^z$ in U^z. Hence $(s^{-1}(0) \cap U) \cap \tau^{-1}(z)$ and $((s^+)^{-1}(0)) \cap \tau^{-1}(z)$ coincide near $(s^{-1}(0) \cap U) \cap \tau^{-1}(z)$ in U. As this holds for all $z \in Z_{\mathrm{an}}$,we have that $s^{-1}(0) \cap U$ and $(s^+)^{-1}(0)$ coincide near $s^{-1}(0) \cap U$ in U, and the theorem follows. $\square$

When we extend Section 3.4 to the relative case, in the analogue of Definition 3.10 we also include data $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z = \operatorname{Spec} B$ and smooth $\beta_J\colon B \to A^0_J$, $\beta_K\colon B \to A^0_K$ with $\beta_J = \Phi_{JK} \circ \beta_K$ and (13) homotopy commuting for J, K. We obtain an analogue of (39) with rows (46) rather than (36), and so as for (40) we get an exact sequence

$$0 \to T_{v_J}(U_J/Z_{\mathrm{an}}) \xrightarrow{\mathrm{d}s^+_J|_{v_J} \oplus \mathrm{d}\phi_{JK}|_{v_J}} E^+_J|_{v_J} \oplus T_{v_K}(U_K/Z_{\mathrm{an}}) \xrightarrow{-\chi^+_{JK}|_{v_J} \oplus \mathrm{d}s^+_K|_{v_K}} E^+_K|_{v_K} \to 0.$$

But by taking the direct sum of this with id: $T_z Z_{\mathrm{an}} \to T_z Z_{\mathrm{an}}$ in the second and third positions, we see that this implies (40) is exact, and the analogue of Corollary 3.11 follows. The relative analogue of Lemma 3.12, in which we replace TV_J, TV_K by $T(V_J/Z_{\mathrm{an}})$, $T(V_K/Z_{\mathrm{an}})$, is immediate.

For Section 3.5, we prove the following relative analogue of Theorem 3.15:

Theorem 3.23 *Let $\boldsymbol{X}$ be a separated derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ a smooth, connected, classical affine $\mathbb{C}$–scheme, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ a morphism, and $[\omega_{\boldsymbol{X}/Z}]$ a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$, with $\dim_{\mathbb{C}} Z = k$ and $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n + k$. Write X_{an}, Z_{an} for the sets of $\mathbb{C}$–points of $X = t_0(\boldsymbol{X})$, Z with the complex analytic topology, and suppose X_{an} is paracompact. Then we can construct a relative Kuranishi atlas $(\mathcal{K}, \{\varpi_J \mid J \in A\})$ for $\pi_{\mathrm{an}}\colon X_{\mathrm{an}} \to Z_{\mathrm{an}}$ of real dimension $n + 2k$, as in Definition 2.15, with $\varpi_J\colon U_J \to Z_{\mathrm{an}}$ a submersion. If X is quasicompact (equivalently, of finite type) then we can take $\mathcal{K}$ to be finite.*

Proof First choose a family $\{(A^\bullet_i, \boldsymbol{\alpha}_i, \beta_i) \mid i \in I\}$, where $A^\bullet_i \in \mathbf{cdga}_{\mathbb{C}}$ is a standard form cdga, and $\boldsymbol{\alpha}_i\colon \mathbf{Spec}\, A^\bullet_i \hookrightarrow \boldsymbol{X}$ is a Zariski open inclusion in $\mathbf{dSch}_{\mathbb{C}}$ for each i in I, an indexing set, and $\beta_i\colon B \to A^0_i$ is a smooth morphism of classical $\mathbb{C}$–algebras such that (12) homotopy commutes, with $\{R_i := (\operatorname{Im} \boldsymbol{\alpha}_i)_{\mathrm{an}} \mid i \in I\}$ an open cover of the complex analytic topological space X_{an}. This is possible by a relative version of Theorem 2.5, easily proved by modifying the proof of [6, Theorem 4.1] to work over the base $Z = \operatorname{Spec} B$. Apply Theorem 3.1 to get data $A^\bullet_J \in \mathbf{cdga}_{\mathbb{C}}$, $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A^\bullet_J \hookrightarrow \boldsymbol{X}$,

$\beta_J\colon B \to A_J^0$ for finite $\varnothing \neq J \subseteq I$ and quasifree morphisms $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$, for all finite $\varnothing \neq K \subseteq J \subseteq I$.

Use the notation of Section 3.2 to rewrite $A_J^\bullet$, β_J, Φ_{JK} in terms of complex geometry. As in Corollary 3.5, this gives data V_J, τ_J, E_J, F_J, s_J, t_J, ψ_J, R_J for all finite $\varnothing \neq J \subseteq I$, and ϕ_{JK}, χ_{JK}, ξ_{JK} for all finite $\varnothing \neq K \subseteq J \subseteq I$. Note that the holomorphic submersions $\tau_J\colon V_J \to Z_{\mathrm{an}}$ with $\tau_J = \tau_K \circ \phi_{JK}$ for $K \subseteq J$ were not used in Sections 3.3–3.6 as there Z_{an} was the point $*$, but now we need them.

Proposition 3.14 now also holds in our relative situation. Its proof in Section 6.2 uses Theorem 3.7 and Lemma 3.12, which as above hold in the relative situation with Definition 3.21 and $T(V_J/Z_{\mathrm{an}})$ in place of Definition 3.6 and TV_J. As in the proof of Theorem 3.15, we have now constructed a Kuranishi atlas $\mathcal{K}$ on X_{an}, with dimension $n+2k$. Setting $\varpi_J := \tau_J|_{U_J}\colon U_J \to Z_{\mathrm{an}}$ for $J \in A$, we see that $(\mathcal{K}, \{\varpi_J \mid J \in A\})$ is a relative Kuranishi atlas for π_{an}, with ϖ_J a submersion. If X is quasicompact we can take I finite, so A and $\mathcal{K}$ are finite. □

We then deduce the following relative analogue of Theorem 3.16:

Theorem 3.24 (i) *Let $\boldsymbol{X}$ be a separated derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ a smooth, connected, classical affine $\mathbb{C}$–scheme, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ a morphism, and $[\omega_{\boldsymbol{X}/Z}]$ a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$, with $\dim_{\mathbb{C}} Z = k$ and $\operatorname{vdim}_{\mathbb{C}} \boldsymbol{X} = n+k$. Write X_{an}, Z_{an} for the sets of $\mathbb{C}$–points of $X = t_0(\boldsymbol{X})$, Z with the complex analytic topology, and suppose X_{an} is second countable.*

Then we can make the topological space X_{an} into a derived manifold $\boldsymbol{X}_{\mathrm{dm}}$ with real virtual dimension $\operatorname{vdim}_{\mathbb{R}} \boldsymbol{X}_{\mathrm{dm}} = n+2k$, in any of the senses (a) *Joyce's m-Kuranishi spaces* **mKur** [21, Section 4.7], (b) *Joyce's d-manifolds* **dMan** [18; 19; 20], (c) *Borisov and Noël's derived manifolds* $\mathbf{DerMan}_{\mathrm{BoNo}}$ [3; 4], *or* (d) *Spivak's derived manifolds* $\mathbf{DerMan}_{\mathrm{Spi}}$ [32], *all discussed in Section 2.6.*

(ii) *We can also define a morphism of derived manifolds $\boldsymbol{\pi}_{\mathrm{dm}}\colon \boldsymbol{X}_{\mathrm{dm}} \to Z_{\mathrm{an}}$, with underlying continuous map $\pi_{\mathrm{an}}\colon X_{\mathrm{an}} \to Z_{\mathrm{an}}$.*

(iii) *For each $z \in Z_{\mathrm{an}}$, the fibre $\boldsymbol{X}_{\mathrm{dm}}^z = \boldsymbol{\pi}_{\mathrm{dm}}^{-1}(z) = \boldsymbol{X}_{\mathrm{dm}} \times_{\boldsymbol{\pi}_{\mathrm{dm}}, Z_{\mathrm{an}}, z} *$ is a derived manifold with $\operatorname{vdim}_{\mathbb{R}} \boldsymbol{X}_{\mathrm{dm}}^z = n$. From Definition 3.20, $\boldsymbol{X}^z = \boldsymbol{\pi}^{-1}(z)$ has a -2–shifted symplectic structure $\omega_{\boldsymbol{X}^z}^*$, and both $\boldsymbol{X}_{\mathrm{dm}}^z$, $\boldsymbol{X}^z$ have (complex analytic) topological space $\pi_{\mathrm{an}}^{-1}(z) \subseteq X_{\mathrm{an}}$. Then $\boldsymbol{X}_{\mathrm{dm}}^z$ is up to equivalence a possible choice for the derived manifold associated to $(\boldsymbol{X}^z, \omega_{\boldsymbol{X}^z}^*)$ in Theorem 3.16.*

Proof Parts (i) and (ii) follow from Theorems 2.18 and 3.23. For (iii), if $z \in Z_{\mathrm{an}}$ then as $\tau_J\colon V_J \to Z_{\mathrm{an}}$ is a holomorphic submersion for $J \in A$, the fibre $V_J^z := \tau_J^{-1}(z)$ is a complex submanifold of V_J. Setting $U_J^z = U_J \cap V_J^z$ and writing E_J^z, F_J^z, s_J^z, t_J^z for the restrictions of E_J, F_J, s_J, t_J to V_J^z, and E_J^{-z}, E_J^{+z}, s_J^{+z}, ψ_J^{+z} for the restrictions of E_J^-, E_J^+, s_J^+, ψ_J^+ to U_J^z, we see I, A, V_J^z, E_J^z, F_J^z, s_J^z, t_J^z, $U_J^z, \dots$ are a possible choice for the data I, A, V_J, $E_J, \dots$ in the application of Theorems 3.15 and 3.16 to $(\boldsymbol{X}^z, \omega^*_{\boldsymbol{X}^z})$. But from facts about fibre products of derived manifolds in [18; 19; 20; 24] we see that the derived manifold $\boldsymbol{X}^z_{\mathrm{dm}} = \boldsymbol{X}_{\mathrm{dm}} \times_{\boldsymbol{\pi}_{\mathrm{dm}}, Z_{\mathrm{an}}, z} *$ may be constructed from the data I, A, U_J^z, E_J^{+z}, s_J^{+z}, $\psi_J^{+z}, \dots$, as above. The theorem follows. □

Next we discuss orientations, generalizing Section 2.4 and Section 3.6 to the relative case. Here is the analogue of Definition 2.12:

Definition 3.25 Let $\boldsymbol{X}$ be a derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ a smooth, connected, classical affine $\mathbb{C}$–scheme, $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z$ a morphism, and $[\omega_{\boldsymbol{X}/Z}] \in H^{-2}(\mathbb{L}_{\boldsymbol{X}/Z})$ a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$. Then as in (4), $[\omega_{\boldsymbol{X}/Z}]$ induces a canonical isomorphism of line bundles on $X = t_0(\boldsymbol{X})$:

$$\iota_{\boldsymbol{X}/Z,\omega_{\boldsymbol{X}/Z}}\colon [\det(\mathbb{L}_{\boldsymbol{X}/Z}|_X)]^{\otimes^2} \to \mathcal{O}_X \cong \mathcal{O}_X^{\otimes^2}.$$

An *orientation* for $(\boldsymbol{\pi}\colon \boldsymbol{X} \to Z, [\omega_{\boldsymbol{X}/Z}])$ is an isomorphism $o\colon \det(\mathbb{L}_{\boldsymbol{X}/Z}|_X) \to \mathcal{O}_X$ such that $o \otimes o = \iota_{\boldsymbol{X}/Z,\omega_{\boldsymbol{X}/Z}}$.

Here is the relative analogue of Proposition 3.17. In parts (b) and (c), we could also use notions of relative orientation for $(X_{\mathrm{an}}, \mathcal{K}) \to Z_{\mathrm{an}}$ and $\boldsymbol{X}_{\mathrm{dm}} \to Z_{\mathrm{an}}$. But as Z_{an} is a complex manifold with a natural orientation, these are equivalent to absolute orientations for $(X_{\mathrm{an}}, \mathcal{K})$, $\boldsymbol{X}_{\mathrm{dm}}$, so we do not bother. The proof is an easy modification of that in Section 6.3.

Proposition 3.26 *In the situation of Theorems 3.23 and 3.24, there are canonical one-to-one correspondences between*

(a) *orientations on $(\boldsymbol{\pi}\colon \boldsymbol{X} \to Z, [\omega_{\boldsymbol{X}/Z}])$ in the sense of Definition 3.25;*

(b) *orientations on $(X_{\mathrm{an}}, \mathcal{K})$ in the sense of Section 2.5; and*

(c) *orientations on $\boldsymbol{X}_{\mathrm{dm}}$ in the sense of Section 2.6.2.*

The relative analogue of Proposition 3.18 does hold, but we will not prove it, as we do not need it. The next theorem says that the virtual classes $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{dbo}}$, $[\boldsymbol{X}_{\mathrm{dm}}]_{\mathrm{virt}}$ of a proper oriented -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X}, \omega^*_{\boldsymbol{X}})$ defined in Corollary 3.19 are unchanged under deformation in families. Note that it is essential that the base $\mathbb{C}$–scheme Z be connected in Theorem 3.27.

Theorem 3.27 *Let X be a separated derived $\mathbb{C}$–scheme, $Z = \operatorname{Spec} B$ a smooth, connected, classical affine $\mathbb{C}$–scheme, $\boldsymbol{\pi}\colon X \to Z$ a proper morphism, and $[\omega_{X/Z}]$ a family of -2–shifted symplectic structures on X/Z, equipped with an orientation, with $\dim_{\mathbb{C}} Z = k$ and $\operatorname{vdim}_{\mathbb{C}} X = n + k$.*

For each $z \in Z_{\mathrm{an}}$ we have a proper, oriented -2–shifted symplectic $\mathbb{C}$–scheme $(X^z, \omega^*_{X^z})$ with $\operatorname{vdim} X^z = n$, and thus Corollary 3.19 defines a d-bordism class $[X^z_{\mathrm{dm}}]_{\mathrm{dbo}} \in dB_n(*)$ and a virtual class $[X^z_{\mathrm{dm}}]_{\mathrm{virt}} \in H_n(X^z_{\mathrm{an}}; \mathbb{Z})$, which depend only on $(X^z, \omega^*_{X^z})$. Then $[X^z_{\mathrm{dm}}]_{\mathrm{dbo}} = [X^{z'}_{\mathrm{dm}}]_{\mathrm{dbo}}$ and $\iota^z_*([X^z_{\mathrm{dm}}]_{\mathrm{virt}}) = \iota^{z'}_*([X^{z'}_{\mathrm{dm}}]_{\mathrm{virt}})$ for all $z, z' \in Z_{\mathrm{an}}$, where $\iota^z_*([X^z_{\mathrm{dm}}]_{\mathrm{virt}}) \in H_n(X_{\mathrm{an}}; \mathbb{Z})$ is the pushforward under the inclusion $\iota^z\colon X^z_{\mathrm{an}} \hookrightarrow X_{\mathrm{an}}$.

Proof Theorem 3.24 constructs a derived manifold X_{dm} with $\operatorname{vdim} X_{\mathrm{dm}} = n + 2k$ and a morphism $\boldsymbol{\pi}_{\mathrm{dm}}\colon X_{\mathrm{dm}} \to Z_{\mathrm{an}}$, which is proper as $\boldsymbol{\pi}$ is proper, and Proposition 3.26 gives an orientation on X_{dm}.

Let $z, z' \in Z_{\mathrm{an}}$. As Z is connected we can choose a smooth map $\gamma\colon [0,1] \to Z_{\mathrm{an}}$ with $\gamma(0) = z$ and $\gamma(1) = z'$. The fibre product

$$W_{\mathrm{dm}} = X_{\mathrm{dm}} \times_{\boldsymbol{\pi}_{\mathrm{dm}}, Z_{\mathrm{an}}, \gamma} [0,1]$$

exists as a derived manifold with boundary by [19, Section 7.5; 18, Section 7.6] and Joyce [24], with $\operatorname{vdim} W_{\mathrm{dm}} = n+1$, and W_{dm} is compact as $[0,1]$ is and $\boldsymbol{\pi}_{\mathrm{dm}}$ is proper, and oriented since X_{dm}, Z_{an}, $[0,1]$ are. As $\partial X_{\mathrm{dm}} = \partial Z_{\mathrm{an}} = \varnothing$, the boundary is

$$\partial W_{\mathrm{dm}} = X_{\mathrm{dm}} \times_{\boldsymbol{\pi}_{\mathrm{dm}}, Z_{\mathrm{an}}, \gamma} \partial[0,1] = X^z_{\mathrm{dm}} \amalg X^{z'}_{\mathrm{dm}},$$

where X^z_{dm}, $X^{z'}_{\mathrm{dm}}$ are the fibres of $\boldsymbol{\pi}_{\mathrm{dm}}\colon X_{\mathrm{dm}} \to Z_{\mathrm{an}}$ at z, z'.

Since $\partial[0,1] = -\{0\} \amalg \{1\}$ in oriented 0–manifolds, we have $\partial W_{\mathrm{dm}} = -X^z_{\mathrm{dm}} \amalg X^{z'}_{\mathrm{dm}}$ in oriented derived manifolds. Therefore Definition 2.20 gives $[X^z_{\mathrm{dm}}]_{\mathrm{dbo}} = [X^{z'}_{\mathrm{dm}}]_{\mathrm{dbo}}$ in $dB_n(*)$. By Theorem 3.22(c), X^z_{dm}, $X^{z'}_{\mathrm{dm}}$ are outcomes of Theorem 3.16 applied to $(X^z, \omega^*_{X^z})$, $(X^{z'}, \omega^*_{X^{z'}})$, so $[X^z_{\mathrm{dm}}]_{\mathrm{dbo}}$, $[X^{z'}_{\mathrm{dm}}]_{\mathrm{dbo}}$ are the d-bordism classes associated to $(X^z, \omega^*_{X^z})$, $(X^{z'}, \omega^*_{X^{z'}})$ in Corollary 3.19. A similar argument works for the homology classes. □

Remark 3.28 The assumptions that Z is smooth, classical and affine, and X is separated, in Theorem 3.27 are easily removed; we can work over a base Z which is a general classical or derived $\mathbb{C}$–scheme, provided it is connected.

To see this, suppose $\boldsymbol{\pi}\colon X \to Z$ is a proper morphism of derived $\mathbb{C}$–schemes with Z connected, and $[\omega_{X/Z}] \in H^{-2}(\mathbb{L}_{X/Z})$ is a family of -2–shifted symplectic structures on X/Z equipped with an orientation, extending Definitions 3.20 and 3.25 to general Z in the obvious way.

Suppose $z, z' \in Z_{\mathrm{an}}$. As $\boldsymbol{Z}$ is connected we can find a sequence $z=z_0, z_1, \dots, z_N=z'$ of points in Z_{an}, and a sequence of smooth, connected, affine curves $C^1, \dots, C^N$ over $\mathbb{C}$ with morphisms $\pi^i\colon C^i \to \boldsymbol{Z}$, such that $\pi^i(C^i)$ contains z_{i-1}, z_i for $i = 1, \dots, N$. Then $\boldsymbol{X}^i = \boldsymbol{X} \times^h_{\boldsymbol{\pi}, \boldsymbol{Z}, \pi^i} C^i$ is a derived $\mathbb{C}$–scheme, and $[\omega_{\boldsymbol{X}/\boldsymbol{Z}}]$ pulls back to a family $[\omega_{\boldsymbol{X}^i/C^i}]$ of oriented -2–shifted symplectic structures on $\boldsymbol{X}^i/C^i$. Applying Theorem 3.27 to $(\boldsymbol{X}^i \to C^i, [\omega_{\boldsymbol{X}^i/C^i}])$ we see $[\boldsymbol{X}_{\mathrm{dm}}^{z_{i-1}}] = [\boldsymbol{X}_{\mathrm{dm}}^{z_i}]$ in $dB_n(*)$ for $i = 1, \dots, N$, so that

$$[\boldsymbol{X}_{\mathrm{dm}}^{z}]_{\mathrm{dbo}} = [\boldsymbol{X}_{\mathrm{dm}}^{z_0}]_{\mathrm{dbo}} = [\boldsymbol{X}_{\mathrm{dm}}^{z_1}]_{\mathrm{dbo}} = \cdots = [\boldsymbol{X}_{\mathrm{dm}}^{z_N}]_{\mathrm{dbo}} = [\boldsymbol{X}_{\mathrm{dm}}^{z'}]_{\mathrm{dbo}}.$$

The same argument works for virtual classes $[\boldsymbol{X}_{\mathrm{dm}}^{z}]_{\mathrm{virt}}$ in homology.

We took Z to be smooth above to avoid defining families $\boldsymbol{\pi}_{\mathrm{dm}}\colon \boldsymbol{X}_{\mathrm{dm}} \to \boldsymbol{Z}$ of derived manifolds over a base $\boldsymbol{Z}$ which is not a (derived) manifold.

3.8 "Holomorphic Donaldson invariants" of Calabi–Yau 4–folds

We now outline how the results of Sections 3.1–3.7 can be used to define new enumerative invariants of (semi)stable coherent sheaves on Calabi–Yau 4–folds Y, which we could call "holomorphic Donaldson invariants", and which should be unchanged under deformations of Y. A related programme using gauge theory has recently been proposed by Cao and Leung [8; 9; 10], which we discuss in Section 3.9.

We begin by discussing *Donaldson–Thomas invariants* $\mathrm{DT}^\alpha(\tau)$ of Calabi–Yau 3–folds, introduced by Thomas [33]. Suppose Z is a Calabi–Yau 3–fold over $\mathbb{C}$ with an ample line bundle $\mathcal{O}_Z(1)$, which defines a Gieseker stability condition τ on coherent sheaves on Z, and $\alpha \in H^{\mathrm{even}}(Z;\mathbb{Q})$. Then one can form coarse moduli $\mathbb{C}$–schemes $\mathcal{M}^\alpha_{\mathrm{st}}(\tau)$, $\mathcal{M}^\alpha_{\mathrm{ss}}(\tau)$ of τ–(semi)stable coherent sheaves on Z of Chern character α, with $\mathcal{M}^\alpha_{\mathrm{st}}(\tau) \subseteq \mathcal{M}^\alpha_{\mathrm{ss}}(\tau)$ Zariski open, and $\mathcal{M}^\alpha_{\mathrm{ss}}(\tau)$ proper.

Thomas [33] showed that $\mathcal{M}^\alpha_{\mathrm{st}}(\tau)$ carries an "obstruction theory" $\phi\colon E^\bullet \to \mathbb{L}_{\mathcal{M}^\alpha_{\mathrm{st}}(\tau)}$ of virtual dimension 0, in the sense of Behrend and Fantechi [1]. Thus, if there are no strictly τ–semistable sheaves in class α, so that $\mathcal{M}^\alpha_{\mathrm{st}}(\tau) = \mathcal{M}^\alpha_{\mathrm{ss}}(\tau)$ and $\mathcal{M}^\alpha_{\mathrm{st}}(\tau)$ is proper, then [1] gives a virtual count $\mathrm{DT}^\alpha(\tau) = [\mathcal{M}^\alpha_{\mathrm{st}}(\tau)]_{\mathrm{virt}} \in \mathbb{Z}$. Thomas proved that $\mathrm{DT}^\alpha(\tau)$ is unchanged under continuous deformations of Z.

Later, Joyce and Song [25] extended the definition of $\mathrm{DT}^\alpha(\tau)$ to invariants $\overline{\mathrm{DT}}^\alpha(\tau) \in \mathbb{Q}$ for all $\alpha \in H^{\mathrm{even}}(Z;\mathbb{Q})$, dropping the condition that there are no strictly τ–semistable sheaves in class α, and proved a wall-crossing formula for $\overline{\mathrm{DT}}^\alpha(\tau)$ under change of stability condition τ. At about the same time, Kontsevich and Soibelman [26] defined a motivic generalization of Donaldson–Thomas invariants (assuming existence of "orientation data" as in Section 2.4), and proved their own wall-crossing formula under change of τ.

Thomas [33] called his invariants $\mathrm{DT}^\alpha(\tau)$ "holomorphic Casson invariants", though they are now generally known as Donaldson–Thomas invariants. Here *Casson invariants* are integer invariants of oriented real 3–manifolds $Z_{\mathbb{R}}$ which are homology 3–spheres, which "count" flat connections on $Z_{\mathbb{R}}$.

This followed a programme of Donaldson and Thomas [13], which starting with some well-known geometry in real dimensions 2, 3 and 4, aimed to find analogues in complex dimensions 2, 3 and 4; so the complex analogues of homology 3–spheres, and flat connections upon them, are Calabi–Yau 3–folds, and holomorphic vector bundles (or coherent sheaves) upon them.

Donaldson invariants [12] are invariants of compact, oriented 4–manifolds $Y_{\mathbb{R}}$, defined by "counting" moduli spaces $\mathcal{M}_{\mathrm{inst}}^\alpha$ of SU(2)–instantons E on $Y_{\mathbb{R}}$ with $c_2(E)=\alpha\in\mathbb{Z}$. In contrast to Casson and Donaldson–Thomas invariants, the (virtual) dimension d^α of $\mathcal{M}_{\mathrm{inst}}^\alpha$ need not be zero. Oversimplifying/lying a bit, one first constructs an orientation on $\mathcal{M}_{\mathrm{inst}}^\alpha$ [12, Section 5.4]. Then we have a virtual class $[\mathcal{M}_{\mathrm{inst}}^\alpha]_{\mathrm{virt}}\in H_{d^\alpha}(\mathcal{M}_{\mathrm{inst}}^\alpha;\mathbb{Z})$. For each $\beta\in H_2(Y_{\mathbb{R}};\mathbb{Z})$ we construct a natural cohomology class $\mu(\beta)\in H^2(\mathcal{M}_{\mathrm{inst}}^\alpha;\mathbb{Z})$, with $\mu(\beta_1+\beta_2)=\mu(\beta_1)+\mu(\beta_2)$. Then if $d^\alpha=2k$, we define *Donaldson invariants* $D^\alpha(\beta_1,\dots,\beta_k)=(\mu(\beta_1)\cup\cdots\cup\mu(\beta_k))\cdot[\mathcal{M}_{\mathrm{inst}}^\alpha]_{\mathrm{virt}}\in\mathbb{Z}$ for all $\beta_1,\dots,\beta_k\in H_2(Y_{\mathbb{R}};\mathbb{Z})$. We can think of D^α as a $\mathbb{Z}$–valued homogeneous degree-k polynomial on $H_2(Y_{\mathbb{R}};\mathbb{Z})$.

We propose, following [13], to define "holomorphic Donaldson invariants" of Calabi–Yau 4–folds. The gauge theory ideas which were the primary focus of [13] will be discussed in Section 3.9; here we work in the world of (derived) algebraic geometry. Suppose Y is a Calabi–Yau 4–fold over $\mathbb{C}$ (ie Y is smooth and projective with $H^i(\mathcal{O}_Y)=\mathbb{C}$ if $i=0,4$ and $H^i(\mathcal{O}_Y)=0$ otherwise), and $\alpha=(\alpha^0,\alpha^2,\alpha^4,\alpha^6,\alpha^8)\in H^{\mathrm{even}}(Y;\mathbb{Q})$. As above we can form coarse moduli $\mathbb{C}$–schemes $\mathcal{M}_{\mathrm{st}}^\alpha(\tau)\subseteq\mathcal{M}_{\mathrm{ss}}^\alpha(\tau)$ of Gieseker (semi)stable coherent sheaves on Y of Chern character α, with $\mathcal{M}_{\mathrm{ss}}^\alpha(\tau)$ proper.

To make contact with the work of Sections 3.1–3.7, we need to show:

Claim 3.29 *There is a -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{\mathcal{M}}_{\mathrm{st}}^\alpha(\tau),\omega^*)$, natural up to equivalence, with classical truncation $t_0(\boldsymbol{\mathcal{M}}_{\mathrm{st}}^\alpha(\tau))=\mathcal{M}_{\mathrm{st}}^\alpha(\tau)$, of virtual dimension $\mathrm{vdim}_{\mathbb{C}}\,\boldsymbol{\mathcal{M}}_{\mathrm{st}}^\alpha(\tau)=d^\alpha:=2-\deg(\alpha\cup\bar\alpha\cup\mathrm{td}(TY))_8$, where $\bar\alpha=(\alpha^0,-\alpha^2,\alpha^4,-\alpha^6,\alpha^8)$, and $\mathrm{td}(-)$ is the Todd class.*

Pantev et al [31, Section 2.1] prove the analogue of Claim 3.29 in the context of (derived) Artin stacks, but we want to reduce to (derived) schemes. Roughly this means factoring out the $\mathbb{C}^*$ stabilizer groups at each point of the τ–stable derived moduli stack. Actually, it should not be difficult to extend Sections 3.1–3.7 to derived algebraic $\mathbb{C}$–spaces rather than derived $\mathbb{C}$–schemes, and then it would be enough to construct $\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)$ as a derived algebraic $\mathbb{C}$–space.

Next we would need to answer:

Question 3.30 *Does $(\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau),\omega^*)$ in Claim 3.29 have a natural orientation, in the sense of Section 2.4, possibly depending on some choice of data on Y?*

Following the argument of Donaldson [12, Section 5.4], Cao and Leung prove an orientability result [10, Theorem 2.2], which should translate to the statement that if the Calabi–Yau 4–fold Y has holonomy $\mathrm{SU}(4)$ with $H_*(Y;\mathbb{Z})$ torsion-free, and $\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)$ is a derived moduli scheme of coherent sheaves on Y, then orientations on $\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)$ exist, though they do not construct a natural choice.

If both these problems are solved, then Theorem 3.16 makes $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}}$ into a derived manifold $\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}$ of real virtual dimension d^{α}, which is oriented by Proposition 3.17. If there are no strictly τ–semistable sheaves in class α then $\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}$ is also compact, and has a d-bordism class $[\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}]_{\mathrm{dbo}}$ in $dB_{d^{\alpha}}(*)$ and virtual class $[\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}]_{\mathrm{virt}}$ in $H_{d^{\alpha}}(\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}};\mathbb{Z})$.

If $d^{\alpha}=0$ then $[\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}]_{\mathrm{dbo}}\in dB_0(*)\cong\mathbb{Z}$ is the virtual count we want. But if $d^{\alpha}>0$ we should aim to find suitable cohomology classes on $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}}$ and integrate them over $[\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}]_{\mathrm{virt}}$, as for Donaldson invariants above.

Claim 3.31 *One can define natural cohomology classes $\mu(\beta)$ on $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}}$ depending on homology classes β on Y, which can be combined with $[\boldsymbol{\mathcal{M}}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{dm}}]_{\mathrm{virt}}$ to give integer invariants, in a similar way to Donaldson invariants.*

If $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)$ is a fine moduli space, there is a universal sheaf $\mathcal{E}$ on $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)\times Y$, with Chern classes $c_i(\mathcal{E})\in H^{2i}(\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}}\times Y;\mathbb{Q})\cong\bigoplus_k H^{2i-k}(\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}};\mathbb{Q})\otimes H^k(Y;\mathbb{Q})$, and we can make $\mu_i(\beta)\in H^{2i-k}(\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)_{\mathrm{an}};\mathbb{Q})$ by contracting $c_i(\mathcal{E})$ with $\beta\in H_k(Y;\mathbb{Q})$. Using the results of Section 3.7, we should be able to prove that the resulting invariants are unchanged under continuous deformations of Y.

This would take us to the same point as Thomas [33] in the Calabi–Yau 3–fold case: we could "count" moduli spaces $\mathcal{M}_{\mathrm{st}}^{\alpha}(\tau)$ for those classes α containing no strictly τ–semistable sheaves, and get a deformation-invariant answer. Many questions would remain, for instance, how to count strictly τ–semistables, wall-crossing formulae as in [25; 26], computation in examples, and so on.

We hope to return to these issues in future work.

3.9 Motivation from gauge theory and "SU(4) instantons"

Finally we discuss some ideas of Donaldson and Thomas [13], which were part of the motivation for this paper, and the work of Cao and Leung [8; 9; 10].

Let Y be a Calabi–Yau 4–fold over $\mathbb{C}$, regarded as a compact real 8–manifold Y with complex structure J, Ricci-flat Kähler metric g, Kähler form ω and holomorphic volume form Ω. Fix a complex vector bundle $E \to Y$ of rank $r > 0$ with Hermitian metric h and Chern character $\mathrm{ch}(E) = \alpha$, and as in [8; 9] assume for simplicity that $c_1(E) = 0$. Consider connections ∇ on E preserving h that have curvature $F \in C^\infty(\mathrm{End}(E) \otimes_{\mathbb{C}} (\Lambda^2 T^*Y \otimes_{\mathbb{R}} \mathbb{C}))$. The splitting

$$\Lambda^2 T^*Y \otimes_{\mathbb{R}} \mathbb{C} = \langle \omega \rangle_{\mathbb{C}} \oplus \Lambda_0^{1,1} T^*Y \oplus \Lambda^{2,0} T^*Y \oplus \Lambda^{0,2} T^*Y$$

induces a corresponding decomposition $F = F^\omega \oplus F_0^{1,1} \oplus F^{2,0} \oplus F^{0,2}$.

We call ∇ a *Hermitian–Einstein connection* if $F^\omega = F^{2,0} = F^{0,2} = 0$. There is a splitting $\nabla = \partial_E \oplus \bar\partial_E$, where $\bar\partial_E$ gives E the structure of a holomorphic vector bundle on (Y, J), as $F^{0,2} = 0$. The *Hitchin–Kobayashi correspondence* says that if $(E, \bar\partial_E)$ is a holomorphic vector bundle and is slope-stable, then $\bar\partial_E$ extends to a unique Hermitian–Einstein connection $\nabla = \partial_E \oplus \bar\partial_E$ preserving h. Also, holomorphic vector bundles on Y are algebraic. Thus, studying moduli spaces $\mathcal{M}^\alpha_{\text{alg-vb}}$ of stable algebraic vector bundles is roughly equivalent to studying moduli spaces $\mathcal{M}^\alpha_{\mathrm{HE}}$ of Hermitian–Einstein connections, modulo gauge.

As a system of PDEs, the Hermitian–Einstein equations are *overdetermined*: there are $8r^2$ unknowns, $13r^2$ equations and r^2 gauge equivalences, with $8r^2 - 13r^2 - r^2 < 0$. Algebraically, this corresponds to the fact that the natural obstruction theory on $\mathcal{M}_{\text{alg-vb}}$ is not perfect, so we cannot form virtual classes.

Using Ω, g we can define real splittings

$$\Lambda^{2,0} T^*Y = \Lambda_+^{2,0} T^*Y \oplus \Lambda_-^{2,0} T^*Y \quad \text{and} \quad \Lambda^{0,2} T^*Y = \Lambda_+^{0,2} T^*Y \oplus \Lambda_-^{0,2} T^*Y$$

and corresponding decompositions

$$F^{2,0} = F_+^{2,0} \oplus F_-^{2,0} \quad \text{and} \quad F^{0,2} = F_+^{0,2} \oplus F_-^{0,2}.$$

Following Donaldson and Thomas [13, Section 3], we call ∇ an $\mathrm{SU}(4)$*–instanton* if $F^\omega = F_+^{2,0} = F_+^{0,2} = 0$. This gives $8r^2$ unknowns, $7r^2$ equations and r^2 gauge equivalences, with $8r^2 - 7r^2 - r^2 = 0$. It is a determined elliptic system, so that we can hope to define virtual classes. This is special to Calabi–Yau 4–folds, a complex analogue of instantons on real 4–manifolds.

Writing $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$ for the moduli space of SU(4)–instantons, we have $\mathcal{M}^\alpha_{\mathrm{HE}} \subseteq \mathcal{M}^\alpha_{\mathrm{SU}(4)}$, as the SU(4) instanton equations are weaker than the Hermitian–Einstein equations. Now $\alpha = \mathrm{ch}(E) \in \bigoplus_{p=0}^4 H^{p,p}(Y)$ if E admits Hermitian–Einstein connections. Conversely, as in [13, page 36], if $\alpha \in \bigoplus_p H^{p,p}(Y)$ then one can use L^2–norms of components of F to show that any SU(4)–instanton is Hermitian–Einstein. Thus, either $\mathcal{M}^\alpha_{\mathrm{HE}} = \mathcal{M}^\alpha_{\mathrm{SU}(4)}$, or $\mathcal{M}^\alpha_{\mathrm{HE}} = \varnothing$.

However, the equality $\mathcal{M}^\alpha_{\mathrm{HE}} = \mathcal{M}^\alpha_{\mathrm{SU}(4)}$ holds only at the level of sets, or topological spaces. Since $\mathcal{M}^\alpha_{\mathrm{HE}}$ is defined by more equations, if we regard $\mathcal{M}^\alpha_{\mathrm{HE}}$, $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$ as (derived) C^∞–schemes, for instance, then $\mathcal{M}^\alpha_{\mathrm{HE}} \subsetneq \mathcal{M}^\alpha_{\mathrm{SU}(4)}$.

In the setting of Sections 3.1–3.6, we should compare $\mathcal{M}^\alpha_{\mathrm{HE}}$ (a Calabi–Yau 4–fold moduli space, without a virtual class, equivalent to an algebraic moduli scheme $\mathcal{M}^\alpha_{\text{alg-vb}}$) with the -2–shifted symplectic derived $\mathbb{C}$–scheme $(\boldsymbol{X}, \omega^*_{\boldsymbol{X}})$, and $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$ (an elliptic moduli space, hopefully with a virtual class, equal to $\mathcal{M}^\alpha_{\mathrm{HE}}$ on the level of topological spaces) with the derived manifold $\boldsymbol{X}_{\mathrm{dm}}$. It was these ideas from Donaldson and Thomas [13] that led the authors to believe that one could modify a -2–shifted symplectic derived $\mathbb{C}$–scheme to get a derived manifold with the same topological space, and so define a virtual class.

Donaldson and Thomas [13] envisaged using gauge theory to define invariants of Calabi–Yau 4–folds "counting" moduli spaces $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$, and also invariants of compact Spin(7)–manifolds "counting" moduli spaces of "Spin(7)–instantons".

This would require finding suitable compactifications $\overline{\mathcal{M}}{}^\alpha_{\mathrm{SU}(4)}$ of the moduli spaces $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$, and giving them a nice enough geometric structure to define virtual classes, which is a formidably difficult problem in gauge theory in dimensions > 4. A *huge advantage of our approach* is that, working in algebraic geometry, with moduli spaces of coherent sheaves rather than vector bundles, *we often get compactness of moduli spaces for free*, without doing any work.

Cao and Leung [8; 9; 10] also aim to define enumerative invariants of Calabi–Yau 4–folds Y, which they call "DT_4–invariants", and their ideas overlap with ours. As for our outline in Section 3.8, their general theory is still rather incomplete, but they prove many partial results, and do computations in examples.

Given a vector bundle moduli space $\mathcal{M}^\alpha_{\text{alg-vb}} \cong \mathcal{M}^\alpha_{\mathrm{HE}} \cong \mathcal{M}^\alpha_{\mathrm{SU}(4)}$ in topological spaces, assuming it is compact, and with an orientation (compare Question 3.30), Cao and Leung [9, Section 5] define a virtual class $[\mathcal{M}^\alpha_{\mathrm{SU}(4)}]_{\mathrm{virt}}$ for $\mathcal{M}^\alpha_{\mathrm{SU}(4)}$, and contract this with some cohomology classes $\mu(\beta)$ (compare Claim 3.31) to get integer invariants, which they prove are unchanged under deformations of Y. All this involves fairly standard material from gauge theory.

They also discuss the case in which one has a compact moduli space of coherent sheaves $\mathcal{M}^{\alpha}_{\text{coh-sh}}$, which contains the vector bundle moduli space $\mathcal{M}^{\alpha}_{\text{alg-vb}}$ as an open subset. They want to define a virtual class for $\mathcal{M}^{\alpha}_{\text{coh-sh}}$, as we want to, and they can do this under the assumptions that either $\mathcal{M}^{\alpha}_{\text{coh-sh}}$ is smooth, or (in our language) that the -2–shifted symplectic derived scheme $(\boldsymbol{\mathcal{M}}^{\alpha}_{\text{coh-sh}}, \omega^*)$ is locally of the form $T^*\boldsymbol{X}[2]$ for $\boldsymbol{X}$ a quasismooth derived $\mathbb{C}$–scheme.

To compare our work with theirs, given $\mathcal{M}^{\alpha}_{\text{alg-vb}} \subset \mathcal{M}^{\alpha}_{\text{coh-sh}}$ as above, assuming Claim 3.29, our Theorem 3.16 gives $\mathcal{M}^{\alpha}_{\text{coh-sh}}$ the structure of a derived manifold, but one depending on arbitrary choices. By topologically identifying $\mathcal{M}^{\alpha}_{\text{alg-vb}} \cong \mathcal{M}^{\alpha}_{\text{SU}(4)}$, in effect Cao and Leung make $\mathcal{M}^{\alpha}_{\text{alg-vb}}$ into a derived manifold, *canonically up to equivalence* (though depending on the Kähler metric g and holomorphic volume form Ω). However, there seems no reason why their derived manifold structure on $\mathcal{M}^{\alpha}_{\text{alg-vb}} \subset \mathcal{M}^{\alpha}_{\text{coh-sh}}$ should extend smoothly to $\mathcal{M}^{\alpha}_{\text{coh-sh}}$. This is a reason why our approach may in the end be more effective.

4 Proof of Theorem 3.1

In this proof we write $\mathbf{cdga}_{\mathbb{C}}$ for the ordinary category of cdgas over $\mathbb{C}$, and $\mathbf{cdga}^{\infty}_{\mathbb{C}}$ for the ∞–category of cdgas over $\mathbb{C}$, defined using the model structure on $\mathbf{cdga}_{\mathbb{C}}$. All objects in $\mathbf{cdga}_{\mathbb{C}}$ are fibrant. A cdga A is cofibrant if it is a retract of a cdga A' which is *almost-free*, that is, free as a graded commutative algebra. If $\phi\colon A \to B$ is a morphism in $\mathbf{cdga}_{\mathbb{C}}$ then $\phi\colon A \to B$ is also a morphism in $\mathbf{cdga}^{\infty}_{\mathbb{C}}$. However, morphisms $\phi\colon A \to B$ in $\mathbf{cdga}^{\infty}_{\mathbb{C}}$ may not correspond to morphisms $A \to B$ in $\mathbf{cdga}_{\mathbb{C}}$ unless A is cofibrant.

The spectrum functor $\mathbf{Spec}$ maps $(\mathbf{cdga}_{\mathbb{C}})^{\text{op}} \to \mathbf{dSch}_{\mathbb{C}}$ and $(\mathbf{cdga}^{\infty}_{\mathbb{C}})^{\text{op}} \to \mathbf{dSch}_{\mathbb{C}}$, and $(\mathbf{cdga}^{\infty}_{\mathbb{C}})^{\text{op}} \to \mathbf{dSch}_{\mathbb{C}}$ is an equivalence with the full ∞–subcategory of $\mathbf{dSch}_{\mathbb{C}}$ with affine objects. So, morphisms $\phi\colon A \to B$ in $\mathbf{cdga}^{\infty}_{\mathbb{C}}$ are essentially the same thing as morphisms $\mathbf{Spec}\, B \to \mathbf{Spec}\, A$ in $\mathbf{dSch}_{\mathbb{C}}$.

Let $\boldsymbol{\pi}\colon \boldsymbol{X} \to Z = \operatorname{Spec} B$ and $\{(A^{\bullet}_i, \boldsymbol{\alpha}_i, \beta_i) \mid i \in I\}$ be as in Theorem 3.1. Our task is to construct a standard form cdga $A^{\bullet}_J = (A^*_J, \mathrm{d})$, a Zariski open inclusion $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A^{\bullet}_J \hookrightarrow \boldsymbol{X}$, and a morphism $\beta_J\colon B \to A^0_J$ for all finite $\varnothing \neq J \subseteq I$, and a quasifree morphism $\Phi_{JK}\colon A^{\bullet}_K \to A^{\bullet}_J$ for all finite $\varnothing \neq K \subseteq J \subseteq I$, satisfying certain conditions. We will do this by induction on increasing $k = |J|$. Here is our inductive hypothesis:

Hypothesis 4.1 Let $k = 1, 2, \dots$ be given. Then:

(a) We are given finite subsets S^n_J for all $\varnothing \neq J \subseteq I$ with $|J| \leqslant k$ and for all $n = -1, -2, \dots$.

(b) For all $\emptyset \neq J \subseteq I$ with $|J| \leqslant k$ we have $A_J^0 = \bigotimes_{i \in J}^{\text{over } B} A_i^0$ as a smooth $\mathbb{C}$–algebra of pure dimension, where the tensor products are over B using $\beta_i \colon B \to A_i^0$ to make A_i^0 into a B–algebra, so that if $J = \{i_1, \dots, i_j\}$ then

$$A_J^0 = A_{i_1} \otimes_B A_{i_2} \otimes_B \cdots \otimes_B A_{i_j}. \tag{47}$$

The morphism $\beta_J \colon B \to A_J^0$ is induced by (47) and the $\beta_i \colon B \to A_i^0$ for $i \in J$, and is smooth as the β_i are.

(c) For all $\emptyset \neq J \subseteq I$ with $|J| \leqslant k$, as a graded $\mathbb{C}$–algebra, A_J^* is freely generated over A_J^0 by generators $\coprod_{\emptyset \neq K \subseteq J} S_K^n$ in degree n for $n = -1, -2, \dots$.

(d) For all $\emptyset \neq K \subseteq J \subseteq I$ with $|J| \leqslant k$, the morphism $\Phi_{JK}^0 \colon A_K^0 \to A_J^0$ in degree 0 is the morphism

$$A_K^0 = \bigotimes_{i \in K} {}_B A_i^0 = \Bigl(\bigotimes_{i \in K} {}_B A_i^0\Bigr) \otimes_B \Bigl(\bigotimes_{i \in J \setminus K} {}_B B\Bigr) \to \bigotimes_{i \in J} {}_B A_i^0 = A_J^0$$

induced by the morphisms $\mathrm{id} \colon A_i^0 \to A_i^0$ for $i \in K$ and $\beta_i \colon B \to A_i^0$ for $i \in J \setminus K$. Then $\Phi_{JK} \colon A_K^* \to A_J^*$ is the unique morphism of graded $\mathbb{C}$–algebras acting by Φ_{JK}^0 in degree 0, and mapping $\Phi_{JK} \colon \gamma \mapsto \gamma$ for each $\gamma \in S_L^n$ for $\emptyset \neq L \subseteq K \subseteq J \subseteq I$ and $n = -1, -2, \dots$, so that γ is a free generator of both A_K^* over A_K^0 and A_J^* over A_J^0.

Note that $\Phi_{JK}^0 \colon A_K^0 \to A_J^0$ is a smooth morphism of $\mathbb{C}$–algebras of pure relative dimension, since $\mathrm{id} \colon A_i^0 \to A_i^0$ and $\beta_i \colon B \to A_i^0$ are. Also Φ_{JK} maps independent generators $\coprod_{\emptyset \neq L \subseteq K} S_L^n$ of A_K^* over A_K^0 to independent generators of A_J^* over A_J^0. Hence $\Phi_{JK} \colon A_K^* \to A_J^*$ is quasifree.

Clearly $\beta_J = \Phi_{JK}^0 \circ \beta_K = \Phi_{JK} \circ \beta_K \colon B \to A_J^0$.

Also, if $\emptyset \neq L \subseteq K \subseteq J \subseteq I$ with $|J| \leqslant K$ then clearly $\Phi_{JL}^0 = \Phi_{JK}^0 \circ \Phi_{KL}^0 \colon A_L^0 \to A_J^0$, and $\Phi_{JL} = \Phi_{JK} \circ \Phi_{KL} \colon A_L^* \to A_J^*$.

(e) For all $\emptyset \neq J \subseteq I$ with $|J| \leqslant k$ and all $n = -1, -2, \dots$, we are given maps $\delta_J^n \colon S_J^n \to A_J^{n+1}$.

(f) Let $\emptyset \neq J \subseteq I$ with $|J| \leqslant k$. Define $\mathrm{d} \colon A_J^* \to A_J^{*+1}$ uniquely by the conditions that d satisfies the Leibnitz rule, and

$$\mathrm{d}\gamma = \Phi_{JK} \circ \delta_K^n(\gamma) \quad \text{for all } \emptyset \neq K \subseteq J,\ n \leqslant -1 \text{ and } \gamma \in S_K^n. \tag{48}$$

We require that $\mathrm{d} \circ \mathrm{d} = 0 \colon A_J^* \to A_J^{*+2}$, so that $A_J^\bullet = (A_J^*, \mathrm{d})$ is a cdga.

This defines $A_J^\bullet = (A_J^*, \mathrm{d})$ as a standard form cdga over $\mathbb{C}$. Observe if $\emptyset \neq K \subseteq J \subseteq I$ with $|J| \leqslant k$ then as $\Phi_{JK} \colon A_K^* \to A_J^*$ is a morphism of graded $\mathbb{C}$–algebras with $\Phi_{JK} \circ \mathrm{d}\gamma = \mathrm{d} \circ \Phi_{JK}(\gamma)$ for all γ in the generating sets $\coprod_{\emptyset \neq L \subseteq K} S_L^n$ for A_K^* over A_K^0, we have $\Phi_{JK} \circ \mathrm{d} = \mathrm{d} \circ \Phi_{JK}\colon A_K^* \to A_J^{*+1}$, and so $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ is a morphism of cdgas.

(g) For all $\varnothing \neq J \subseteq I$ with $|J| \leqslant k$, we are given a Zariski open inclusion $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$, with image $\operatorname{Im}\boldsymbol{\alpha}_J = \bigcap_{i\in J} \operatorname{Im}\boldsymbol{\alpha}_i$, such that (13) homotopy commutes.

If $\varnothing \neq K \subseteq J \subseteq I$ with $|J| \leqslant k$ then (14) homotopy commutes.

Remark 4.2 (i) In Hypothesis 4.1, the only actual data required are the finite sets S_J^n in (a), the maps $\delta_J^n\colon S_J^n \to A_J^{n+1}$ in (e), and the morphisms $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\, A_J^\bullet \hookrightarrow \boldsymbol{X}$ in (g).

Also, the only statements requiring proof are that $\mathrm{d} \circ \mathrm{d} = 0$ in (f), and that $\boldsymbol{\alpha}_J$ is a Zariski open inclusion with image $\bigcap_{i\in J} \operatorname{Im}\boldsymbol{\alpha}_i$, and that (13) and (14) homotopy commute in (g). All of (b), (c), (d) are definitions and deductions.

(ii) Most of the conclusions of Theorem 3.1 are immediate from the definitions in (a)–(g): that $A_J^\bullet$ is a standard form cdga, and $\beta_J\colon B \to A_J^0$ is smooth, and $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ is quasifree, and $\beta_J = \Phi_{JK} \circ \beta_K$, and $\Phi_{JL} = \Phi_{JK} \circ \Phi_{KL}$.

For the first step in the induction, we prove Hypothesis 4.1 when $k = 1$. Then the only subsets $\varnothing \neq J \subseteq I$ with $|J| \leqslant k$ are $J = \{i\}$ for $i \in I$, and the only subsets $\varnothing \neq K \subseteq J \subseteq I$ with $|J| \leqslant k$ are $J = K = \{i\}$ for $i \in I$.

As in Theorem 3.1 we are given data $\{(A_i^\bullet, \boldsymbol{\alpha}_i, \beta_i) \mid i \in I\}$, where $A_i^\bullet$ is a standard form cdga, so that A_i^* is freely generated over A_i^0 by finitely many generators in each degree $n = -1, -2, \dots$, as in Definition 2.1. For each $i \in I$ and each $n = -1, -2, \dots$ choose a subset $S_{\{i\}}^n \subseteq A_i^n$, as in part (a) for $J = \{i\}$, such that A_i^* is freely generated over A_i^0 by $\coprod_{n\leqslant -1} S_{\{i\}}^n$. Set $A_{\{i\}}^\bullet = A_i^\bullet$ and $\beta_{\{i\}} = \beta_i$, so that parts (b) and (c) hold for $J = \{i\}$.

Part (d) is a definition, and when $k = 1$ only says that when $J = K = \{i\}$ we have $\Phi_{\{i\}\{i\}} = \mathrm{id}\colon A_{\{i\}}^\bullet \to A_{\{i\}}^\bullet$. For (e), define

$$\delta_{\{i\}}^n\colon S_{\{i\}}^n \to A_{\{i\}}^{n+1} = A_i^{n+1} \quad \text{by} \quad \delta_{\{i\}}^n(\gamma) = \mathrm{d}\gamma,$$

using d in the cdga $A_i^\bullet = (A_i^*, \mathrm{d})$. Given (e), part (f) says that the differentials d in $A_{\{i\}}^\bullet = (A_{\{i\}}^*, \mathrm{d})$ and $A_i^\bullet = (A_i^*, \mathrm{d})$ agree, consistent with setting $A_{\{i\}}^\bullet = A_i^\bullet$, so that $\mathrm{d} \circ \mathrm{d} = 0$ in $A_{\{i\}}^\bullet$ as $A_i^\bullet$ is a cdga.

For (g), if $i \in I$ define $\boldsymbol{\alpha}_{\{i\}} = \boldsymbol{\alpha}_i\colon A_{\{i\}}^\bullet = A_i^\bullet \to \boldsymbol{X}$. Then the assumptions on $\{(A_i^\bullet, \boldsymbol{\alpha}_i, \beta_i) \mid i \in I\}$ in Theorem 3.1 imply that $\boldsymbol{\alpha}_{\{i\}}$ is a Zariski open inclusion, with image $\operatorname{Im}\boldsymbol{\alpha}_{\{i\}} = \operatorname{Im}\boldsymbol{\alpha}_i$, and (13) homotopy commutes for $J = \{i\}$ as (12) does. The only $\varnothing \neq K \subseteq J \subseteq I$ with $|J| \leqslant k = 1$ are $J = K = \{i\}$, and then (14) homotopy commutes as $\boldsymbol{\alpha}_J = \boldsymbol{\alpha}_K = \boldsymbol{\alpha}_{\{i\}}$ and $\Phi_{JK} = \mathrm{id}$. This completes Hypothesis 4.1 when $k=1$. Note that our definitions $A^\bullet_{\{i\}} = A^\bullet_i$, $\boldsymbol{\alpha}_{\{i\}} = \boldsymbol{\alpha}_i$, and $\beta_{\{i\}} = \beta_i$ for $i \in I$ are as required in Theorem 3.1(i).

Next we prove the inductive step. Let $l \geqslant 1$ be given, and suppose Hypothesis 4.1 holds with $k = l$. Keeping all the data in parts (a), (e), (g) for $|J| \leqslant l$ the same, we will prove Hypothesis 4.1 with $k = l+1$. To do this, for each $J \subseteq I$ with $|J| = l+1$, we have to construct the data of finite sets S^n_J for $n = -1, -2, \dots$ in (a), and maps $\delta^n_J \colon S^n_J \to A^{n+1}_J$ in (e), and the morphism $\boldsymbol{\alpha}_J \colon \mathbf{Spec}\, A^\bullet_J \hookrightarrow \boldsymbol{X}$ in (g), and then prove the claims in (f) that $\mathrm{d} \circ \mathrm{d} = 0$, and in (g) that $\boldsymbol{\alpha}_J$ is a Zariski open inclusion with image $\bigcap_{i \in J} \mathrm{Im}\, \boldsymbol{\alpha}_i$, and that (13) and (14) homotopy commute.

Note that as Hypothesis 4.1 involves no compatibility conditions between data for distinct $J, J' \subseteq I$ with $|J| = |J'| = k$, we can do this independently for each $J \subseteq I$ with $|J| = l+1$, that is, it is enough to give the proof for a single such J. So fix a subset $J \subseteq I$ with $|J| = l+1$.

We first define a standard form cdga $\widetilde{A}^\bullet_J$ which is an approximation to the cdga $A^\bullet_J$ that we want, and morphisms $\widetilde{\beta}_J \colon B \to \widetilde{A}^0_J$, $\widetilde{\Phi}_{JK} \colon A^\bullet_K \to \widetilde{A}^\bullet_J$ for all $\varnothing \neq K \subsetneq J$, so that $|K| \leqslant l$ and $A^\bullet_K$ is already defined:

- Define $\widetilde{A}^0_J = A^0_J$ and $\widetilde{\beta}_J = \beta_J \colon B \to \widetilde{A}^0_J = A^0_J$ as in Hypothesis 4.1(b).
- Define $\widetilde{A}^*_J$ to be the graded $\mathbb{C}$–algebra freely generated over A^0_J by generators $\coprod_{\varnothing \neq K \subsetneq J} S^n_K$ in degree n for $n = -1, -2, \dots$. This is the same as for A^*_J in Hypothesis 4.1(c), except that we do not include generators S^n_J, since S^n_J is not yet defined.
- If $\varnothing \neq K \subsetneq J$, so that $A^\bullet_K$ is defined, define $\Phi^0_{JK} \colon A^0_K \to A^0_J = \widetilde{A}^0_J$ as in Hypothesis 4.1(d), and define $\widetilde{\Phi}_{JK} \colon A^*_K \to \widetilde{A}^*_J$ to be the unique morphism of graded $\mathbb{C}$–algebras acting by Φ^0_{JK} in degree 0, and mapping $\Phi_{JK} \colon \gamma \mapsto \gamma$ for each $\gamma \in S^n_L$ for $\varnothing \neq L \subseteq K$ and $n = -1, -2, \dots$.
- The differential $\mathrm{d} \colon \widetilde{A}^*_J \to \widetilde{A}^{*+1}_J$ in the cdga $\widetilde{A}^\bullet_J = (\widetilde{A}^*_J, \mathrm{d})$ is determined uniquely as in (48) by

$$\mathrm{d}\gamma = \widetilde{\Phi}_{JK} \circ \delta^n_K(\gamma) \quad \text{for all } \varnothing \neq K \subsetneq J,\ n \leqslant -1 \text{ and } \gamma \in S^n_K.$$

Then $\widetilde{\Phi}_{JK} \colon A^\bullet_K \to \widetilde{A}^\bullet_J$ is a cdga morphism for all $\varnothing \neq K \subsetneq J$, as in Hypothesis 4.1(f) for Φ_{JK}.

That is, $\widetilde{A}^\bullet_J$ is the colimit in the ordinary category $\mathbf{cdga}_{\mathbb{C}}$ of the commutative diagram Γ with vertices the objects B and $A^\bullet_K$ for all K with $\varnothing \neq K \subsetneq J$, and edges the morphisms $\beta_K \colon B \to A^\bullet_K$ and $\Phi_{K_1K_2} \colon A^\bullet_{K_2} \to A^\bullet_{K_1}$ for $\varnothing \neq K_2 \subsetneq K_1 \subsetneq J$, and $\widetilde{\beta}_J \colon B \to \widetilde{A}^\bullet_J$,

$\widetilde{\Phi}_{JK}\colon A_K^\bullet \to \widetilde{A}_J^\bullet$ are the projections to the colimit. Since all the morphisms in Γ are almost-free in negative degrees and smooth in degree 0, these morphisms are sufficiently cofibrant to compute the homotopy colimits as well. Indeed, having such a morphism $A^\bullet \to C^\bullet$ we can factor it into $A^\bullet \to A^\bullet \otimes_{A^0} C^0 \to C^\bullet$. Each one of these morphisms is flat, and hence homotopy pullbacks can be computed without resolving. Finally we notice that the colimit of the entire diagram Γ can be calculated as a sequence of pullbacks. So $\widetilde{A}_J^\bullet$ is also the homotopy colimit of Γ in the ∞–category $\mathbf{cdga}_{\mathbb{C}}^\infty$. Hence $\mathbf{Spec}\,\widetilde{A}_J^\bullet$ is the homotopy limit of $\mathbf{Spec}\,\Gamma$ in the ∞–category $\mathbf{dSch}_{\mathbb{C}}$.

For $\varnothing \neq K \subsetneq J$, consider $\bigcap_{i\in K} \operatorname{Im}\boldsymbol{\alpha}_i$ as an open derived $\mathbb{C}$–subscheme of $\boldsymbol{X}$. Then by Hypothesis 4.1(g), $\boldsymbol{\alpha}_K\colon \mathbf{Spec}\,A_K^\bullet \to \bigcap_{i\in K}\operatorname{Im}\boldsymbol{\alpha}_i$ is an equivalence in $\mathbf{dSch}_{\mathbb{C}}$. We also have the open derived $\mathbb{C}$–subscheme $\bigcap_{i\in J}\operatorname{Im}\boldsymbol{\alpha}_i$ in $\boldsymbol{X}$, which is affine by Definition 2.6 as $\boldsymbol{X}$ has affine diagonal and $\operatorname{Im}\boldsymbol{\alpha}_i \simeq \mathbf{Spec}\,A_i^\bullet$ is affine for $i\in J$. Thus we may choose a standard form cdga $\widehat{A}_J^\bullet$ and an equivalence $\hat{\boldsymbol{\alpha}}_J\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \hookrightarrow \bigcap_{i\in J}\operatorname{Im}\boldsymbol{\alpha}_i$.

Define morphisms $\hat{\boldsymbol{\beta}}_J\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \to Z = \operatorname{Spec} B$ by $\hat{\boldsymbol{\beta}}_J = \boldsymbol{\pi}\circ\hat{\boldsymbol{\alpha}}_J$, and $\hat{\boldsymbol{\phi}}_{JK}\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \to \mathbf{Spec}\,A_K^\bullet$ for $\varnothing \neq K \subsetneq J$ as the composition

$$\mathbf{Spec}\,\widehat{A}_J^\bullet \xrightarrow{\hat{\boldsymbol{\alpha}}_J} \textstyle\bigcap_{i\in J}\operatorname{Im}\boldsymbol{\alpha}_i \hookrightarrow \bigcap_{i\in K}\operatorname{Im}\boldsymbol{\alpha}_i \xrightarrow{\boldsymbol{\alpha}_K^{-1}} \mathbf{Spec}\,A_K^\bullet,$$

where $\boldsymbol{\alpha}_K^{-1}$ is a quasi-inverse for the equivalence $\boldsymbol{\alpha}_K\colon \mathbf{Spec}\,A_K^\bullet \to \bigcap_{i\in K}\operatorname{Im}\boldsymbol{\alpha}_i$.

By the homotopy limit property of $\mathbf{Spec}\,\widetilde{A}_J^\bullet$, there exists a morphism $\boldsymbol{\psi}\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \to \mathbf{Spec}\,\widetilde{A}_J^\bullet$ in $\mathbf{dSch}_{\mathbb{C}}$ unique up to homotopy, with homotopies $\hat{\boldsymbol{\beta}}_J \simeq \mathbf{Spec}\,\tilde{\beta}_J \circ \boldsymbol{\psi}$ and $\hat{\boldsymbol{\phi}}_{JK} \simeq \mathbf{Spec}\,\widetilde{\Phi}_{JK}\circ\boldsymbol{\psi}$ for $\varnothing \neq K \subsetneq J$. We can then write $\boldsymbol{\psi} \simeq \mathbf{Spec}\,\Psi$ for $\Psi\colon \widetilde{A}_J^\bullet \to \widehat{A}_J^\bullet$ a morphism in $\mathbf{cdga}_{\mathbb{C}}^\infty$, unique up to homotopy. However, we do not yet know that Ψ descends to a morphism in $\mathbf{cdga}_{\mathbb{C}}$. The definitions of $\hat{\boldsymbol{\beta}}_J$, $\hat{\boldsymbol{\phi}}_{JK}$ and $\boldsymbol{\psi} \simeq \mathbf{Spec}\,\Psi$ give homotopies

$$\begin{aligned} &\boldsymbol{\pi}\circ\hat{\boldsymbol{\alpha}}_J \simeq \mathbf{Spec}\,\tilde{\beta}_J \circ \mathbf{Spec}\,\Psi\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \to Z,\\ &\hat{\boldsymbol{\alpha}}_J \simeq \boldsymbol{\alpha}_K \circ \mathbf{Spec}\,\widetilde{\Phi}_{JK}\circ \mathbf{Spec}\,\Psi\colon \mathbf{Spec}\,\widehat{A}_J^\bullet \to \boldsymbol{X} \quad \text{for } \varnothing \neq K \subsetneq J. \end{aligned} \tag{49}$$

Consider the composition of morphisms of classical $\mathbb{C}$–algebras

$$A_J^0 = \widetilde{A}_J^0 \longrightarrow H^0(\widetilde{A}_J^\bullet) \xrightarrow{H^0(\Psi)} H^0(\widehat{A}_J^\bullet). \tag{50}$$

Here $\operatorname{Spec} H^0(\Psi)$ is the natural morphism

$$\operatorname{Spec} H^0(\Psi)\colon X_J \to \prod\nolimits_{Z,\ \varnothing\neq K\subsetneq J} X_K, \tag{51}$$

writing X_K for the open $\mathbb{C}$–subscheme $\bigcap_{k\in K} t_0(\operatorname{Im}\boldsymbol{\alpha}_k)$ in X. This is the restriction of the multidiagonal $\Delta_X^{2^{|J|-2}}\colon X \to X\times_Z X \times_Z \cdots \times_Z X$, with $2^{|J|-2}$ copies of X on the right. Because X is separated, $\Delta_X^2\colon X \to X\times_Z X$ is a closed immersion, and thus $\Delta_X^{2^{|J|-2}}$ is a closed immersion. Also the domain X_J of (51) is the preimage under $\Delta_X^{2^{|J|-2}}$ of the target, since $X_J = \bigcap_{\emptyset \neq K \subsetneq J} X_K$ as $|J| \geqslant 2$.

Hence (51) is a closed immersion, so $H^0(\Psi)$ in (50) is surjective. Also $\widetilde{A}^0_J \to H^0(\widetilde{A}^\bullet_J)$ is surjective, so the composition (50) is surjective. Therefore we can replace $\widehat{A}^\bullet_J$ by an equivalent object in $\mathbf{cdga}^\infty_{\mathbb{C}}$, such that $\widehat{A}^0_J = \widetilde{A}^0_J$, and the following homotopy commutes in $\mathbf{cdga}^\infty_{\mathbb{C}}$:

$$\begin{array}{ccc} \widetilde{A}^0_J & \Longleftrightarrow & \widehat{A}^0_J \\ \downarrow & & \downarrow \\ \widetilde{A}^\bullet_J & \xrightarrow{\Psi} & \widehat{A}^\bullet_J \end{array} \tag{52}$$

Now $\Psi\colon \widetilde{A}^\bullet_J \to \widehat{A}^\bullet_J$ is a morphism in $\mathbf{cdga}^\infty_{\mathbb{C}}$. For this to descend to a morphism in $\mathbf{cdga}_{\mathbb{C}}$, the simplest condition is that $\widetilde{A}^\bullet_J$ should be cofibrant and $\widehat{A}^\bullet_J$ fibrant in the model category $\mathbf{cdga}_{\mathbb{C}}$. Here the object $\widehat{A}^\bullet_J$ is fibrant, as all objects are, but $\widetilde{A}^\bullet_J$ may not be cofibrant, ie a retract of an almost-free cdga. However, $\widetilde{A}^\bullet_J$ is cofibrant as an $\widetilde{A}^0_J$–algebra, as it is free in negative degrees, and (52) says that Ψ does descend to a morphism in $\mathbf{cdga}_{\mathbb{C}}$ in degree 0. Together these imply that Ψ descends to a morphism $\Psi\colon \widetilde{A}^\bullet_J \to \widehat{A}^\bullet_J$ in $\mathbf{cdga}_{\mathbb{C}}$.

Next, by induction on decreasing $n = -1, -2, \ldots$ we will choose the data S^n_J, δ^n_J in parts (a) and (e) of Hypothesis 4.1. Here is our inductive hypothesis:

Hypothesis 4.3 Let $N = 0, -1, -2, \ldots$ be given. Then:

(a) We are given finite subsets S^n_J for $n = -1, -2, \ldots, N$. Write

$$A^*_{J,N} = \widetilde{A}^*_J[S^1_J, \ldots, S^N_J]$$

for the graded $\mathbb{C}$–algebra freely generated over $\widetilde{A}^*_J$ by the sets of extra generators S^n_J in degree n for all $n = -1, -2, \ldots, N$.

(b) We are given maps $\delta^n_J\colon S^n_J \to A^{n+1}_{J,N}$ for $n = -1, -2, \ldots, N$. Define

$$\mathrm{d}\colon A^*_{J,N} \to A^{*+1}_{J,N}$$

uniquely by the conditions that d satisfies the Leibnitz rule, and d is as in $\widetilde{A}^\bullet_J = (\widetilde{A}^*_J, \mathrm{d})$ on $\widetilde{A}^*_J \subseteq A^*_{J,N}$, and on the extra generators $\gamma \in S^n_J$ for $n = -1, -2, \ldots, N$, we have $\mathrm{d}\gamma = \delta^n_J(\gamma) \in A^{n+1}_{J,N}$. We require that $\mathrm{d} \circ \mathrm{d} = 0\colon A^*_{J,N} \to A^{*+2}_{J,N}$, so that $A^\bullet_{J,N} = (A^*_{J,N}, \mathrm{d})$ is a cdga.

(c) We are given maps $\xi^n_J\colon S^n_J \to \widehat{A}^n_J$ for $n = -1, -2, \ldots, N$. Define

$$\Xi_N\colon A^*_{J,N} \to \widehat{A}^*_J$$

to be the morphism of graded $\mathbb{C}$–algebras such that $\Xi_N = \Psi$ on $\widetilde{A}^*_J \subseteq A^*_{J,N}$, and on the extra generators $\gamma \in S^n_J$ for $n = -1, -2, \ldots, N$, we have $\Xi_N(\gamma) = \xi^n_J(\gamma) \in \widehat{A}^n_{J,N}$.

We require that $\Xi_N\circ\mathrm{d}=\mathrm{d}\circ\Xi_N\colon A^*_{J,N}\to\widehat{A}^{*+1}_J$, so that $\Xi_N\colon A^\bullet_{J,N}\to\widehat{A}^\bullet_J$ is a cdga morphism.

We also require that $H^n(\Xi_N)\colon H^n(A^\bullet_{J,N})\to H^n(\widehat{A}^\bullet_J)$ should be an isomorphism for $n=0,-1,-2,\dots,N+1$, and surjective for $n=N$.

For the first step $N=0$, there is no data S^n_J, δ^n_J, ξ^n_J, and $A^\bullet_{J,0}=\widetilde{A}^\bullet_J$, and $\Xi_0=\Psi$, and the only thing to prove is that

$$H^0(\Psi)\colon H^0(\widetilde{A}^\bullet_J)\to H^0(\widehat{A}^\bullet_J)$$

is surjective, which holds as $\Psi^0=\mathrm{id}\colon \widetilde{A}^0_J\to\widetilde{A}^0_J=\widehat{A}^0_J$ from above. So Hypothesis 4.3 holds for $N=0$.

For the inductive step, let $m=0,-1,-2,\dots$ be given, and suppose Hypothesis 4.3 holds with $N=m$. Keeping all the data S^n_J, δ^n_J, ξ^n_J for $n=-1,\dots,m$ the same, we will prove Hypothesis 4.3 with $N=m-1$. Note that with $S^{-1}_J,\dots,S^m_J$ the same, the graded $\mathbb{C}$–algebras $A^*_{J,m}$, $A^*_{J,m-1}$ agree in degrees $0,-1,\dots,m$, so it makes sense to say that

$$\delta^n_J\colon S^n_J\to A^{n+1}_{J,m}\quad\text{and}\quad\delta^n_J\colon S^n_J\to A^{n+1}_{J,m-1}$$

are equal for $n=-1,-2,\dots,m$. We must choose data S^{m-1}_J, $\delta^{m-1}_J\colon S^{m-1}_J\to A^m_{J,m-1}$ and $\xi^{m-1}_J\colon S^{m-1}_J\to\widehat{A}^{m-1}_J$, and verify the last two conditions of Hypothesis 4.3(c).

Choose a finite subset $\dot{S}^{m-1}_J$ of $\mathrm{Ker}(H^m(\Xi_m)\colon H^m(A^\bullet_{J,m})\to H^m(\widehat{A}^\bullet_J))$ which generates $\mathrm{Ker}(\cdots)$ as an $H^0(A^\bullet_{J,m})$–module, and a finite subset $\ddot{S}^{m-1}_J$ of $H^{m-1}(\widehat{A}^\bullet_J)$ such that $\ddot{S}^{m-1}_J$ and $\mathrm{Im}(H^{m-1}(\Xi_m)\colon H^{m-1}(A^\bullet_{J,m})\to H^{m-1}(\widehat{A}^\bullet_J))$ generate $H^{m-1}(\widehat{A}^\bullet_J)$ as an $H^0(\widehat{A}^\bullet_J)$–module. Finite subsets suffice in each case since $A^\bullet_{J,m}$, $\widehat{A}^\bullet_J$ are of standard form, so that the modules $H^n(A^\bullet_{J,m})$, $H^n(\widehat{A}^\bullet_J)$ are finitely generated over $H^0(A^\bullet_{J,m})$, $H^0(\widehat{A}^\bullet_J)$ for all n. Set

$$S^{m-1}_J=\dot{S}^{m-1}_J\amalg\ddot{S}^{m-1}_J.$$

Then Hypothesis 4.3(a) defines $A^*_{J,m-1}$ as a graded $\mathbb{C}$–algebra, with $A^n_{J,m-1}=A^n_{J,m}$ in degrees $n\geqslant m$. For all $\gamma\in\dot{S}^{m-1}_J$, choose a representative $\delta^{m-1}_J(\gamma)$ in $A^m_{J,m-1}=A^m_{J,m}$ for the cohomology class γ in $H^m(A^\bullet_{J,m})$, so that

$$\mathrm{d}(\delta^{m-1}_J(\gamma))=0\quad\text{in }A^{m+1}_{J,m}.$$

Define $\delta^{m-1}_J(\gamma)=0$ in $A^m_{J,m-1}$ for all $\gamma\in\ddot{S}^{m-1}_J$. This defines $\delta^{m-1}_J\colon S^{m-1}_J\to A^m_{J,m-1}$ in Hypothesis 4.3(b), and hence $\mathrm{d}\colon A^*_{J,m-1}\to A^{*+1}_{J,m-1}$.

To see that $\mathrm{d}\circ\mathrm{d}=0\colon A^*_{J,m-1}\to A^{*+2}_{J,m-1}$, note that $A^*_{J,m-1}=A^*_{J,m}[S^{m-1}_J]$, so d on $A^*_{J,m-1}$ is determined by d on $A^*_{J,m}$, which already satisfies $\mathrm{d}\circ\mathrm{d}=0$ by induction, and d on the extra generators S_J^{m-1}, which satisfy $\mathrm{d}\circ\mathrm{d}=0$ as for $\gamma\in\dot{S}_J^{m-1}$ we have $\mathrm{d}\circ\mathrm{d}\gamma=\mathrm{d}(\delta_J^{m-1}(\gamma))=0$, and for $\gamma\in\ddot{S}_J^{m-1}$ we have $\mathrm{d}\gamma=0$ so $\mathrm{d}\circ\mathrm{d}\gamma=0$. Hence $A^\bullet_{J,m-1}=(A^*_{J,m-1},\mathrm{d})$ is a cdga, as we have to prove.

For all $\gamma\in\dot{S}_J^{m-1}$, because $\delta_J^{m-1}(\gamma)\in A^m_{J,m}$ represents a cohomology class in $\mathrm{Ker}(H^m(\Xi_m)\colon H^m(A^\bullet_{J,m})\to H^m(\hat{A}^\bullet_J))$, we see that $\Xi_m\circ\delta_J^{m-1}(\gamma)$ is exact in $\hat{A}^\bullet_J$, so we can choose an element $\xi_J^{m-1}(\gamma)\in\hat{A}_J^{m-1}$ with $\mathrm{d}\circ\xi_J^{m-1}(\gamma)=\Xi_m\circ\delta_J^{m-1}(\gamma)$. For all $\gamma\in\ddot{S}_J^{m-1}\subset H^{m-1}(\hat{A}^\bullet_J)$, choose an element $\xi_J^{m-1}(\gamma)\in\hat{A}_J^{m-1}$ representing γ, so that $\mathrm{d}\circ\xi_J^{m-1}(\gamma)=0$. This defines $\xi_J^{m-1}\colon S_J^{m-1}\to\hat{A}_J^{m-1}$.

Hypothesis 4.3(c) now defines $\Xi_{m-1}\colon A^*_{J,m-1}\to\hat{A}^*_J$. To prove $\Xi_{m-1}\circ\mathrm{d}=\mathrm{d}\circ\Xi_{m-1}$, note that $A^*_{J,m-1}=A^*_{J,m}[S_J^{m-1}]$, and on $A^*_{J,m}\subseteq A^*_{J,m-1}$ we have $\Xi_{m-1}=\Xi_m$, and $\Xi_m\circ\mathrm{d}=\mathrm{d}\circ\Xi_m$ by induction. So it is enough to prove that $\Xi_{m-1}\circ\mathrm{d}(\gamma)=\mathrm{d}\circ\Xi_{m-1}(\gamma)$ for all $\gamma\in S_J^{m-1}$. If $\gamma\in\dot{S}_J^{m-1}$ then

$$\Xi_{m-1}\circ\mathrm{d}(\gamma)=\Xi_{m-1}\circ\delta_J^{m-1}(\gamma)=\Xi_m\circ\delta_J^{m-1}(\gamma)=\mathrm{d}\circ\xi_J^{m-1}(\gamma)=\mathrm{d}\circ\Xi_{m-1}(\gamma),$$

as we want. Similarly, if $\gamma\in\ddot{S}_J^{m-1}$ then

$$\Xi_{m-1}\circ\mathrm{d}(\gamma)=\Xi_{m-1}\circ\delta_J^{m-1}(\gamma)=0=\mathrm{d}\circ\xi_J^{m-1}(\gamma)=\mathrm{d}\circ\Xi_{m-1}(\gamma).$$

Therefore $\Xi_{m-1}\circ\mathrm{d}=\mathrm{d}\circ\Xi_{m-1}$, and $\Xi_{m-1}\colon A^\bullet_{J,m-1}\to\hat{A}^\bullet_J$ is a cdga morphism.

Finally we have to show that $H^n(\Xi_{m-1})\colon H^n(A^\bullet_{J,m-1})\to H^n(\hat{A}^\bullet_J)$ is an isomorphism for $n=-1,-2,\dots,m$, and surjective for $n=m-1$. Since $\Xi_m\colon A^\bullet_{J,m}\to\hat{A}^\bullet_J$ and $\Xi_{m-1}\colon A^\bullet_{J,m-1}\to\hat{A}^\bullet_J$ coincide in degrees $0,-1,\dots,m$, in cohomology they coincide in degrees $0,-1,\dots,m+1$, so $H^n(\Xi_{m-1})$ is an isomorphism for $n=0,-1,\dots,m+1$ as $H^n(\Xi_m)$ is, by induction.

Because $H^m(\Xi_m)\colon H^m(A^\bullet_{J,m})\to H^m(\hat{A}^\bullet_J)$ is surjective, and the added generators $\dot{S}_J^{m-1}$ in $A^\bullet_{J,m-1}$ span $\mathrm{Ker}(H^m(\Xi_m))$, adding the generators $\dot{S}_J^{m-1}$ makes $H^m(\Xi_{m-1})\colon H^m(A^\bullet_{J,m-1})\to H^m(\hat{A}^\bullet_J)$ into an isomorphism. Also, since the added generators $\ddot{S}_J^{m-1}$ together with $\mathrm{Im}(H^{m-1}(\Xi_m))$ generate $H^{m-1}(\hat{A}^\bullet_J)$, adding $\ddot{S}_J^{m-1}$ makes $H^{m-1}(\Xi_{m-1})\colon H^{m-1}(A^\bullet_{J,m-1})\to H^{m-1}(\hat{A}^\bullet_J)$ surjective.

This proves Hypothesis 4.3 for $N=m-1$, so by induction Hypothesis 4.3 holds for all $N=0,-1,-2,\dots$. Taking the limit $\lim_{N\to-\infty}A^\bullet_{J,N}$ gives the cdga $A^\bullet_J$ defined in Hypothesis 4.1 using the data S_J^n, δ_J^n for all $n=-1,-2,\dots$ from parts (a) and (b) of Hypothesis 4.3 as $N\to-\infty$. The data ξ_J^n for $n=-1,-2,\dots$ from part (c) defines a morphism $\Xi=\lim_{N\to-\infty}\Xi_N\colon A^\bullet_J\to\hat{A}^\bullet_J$, where Ξ, $A^\bullet_J$ agree with Ξ_N, $A^\bullet_{J,N}$ in degrees $0,-1,\dots,N$ for all $N\leqslant 0$.

Hence $H^n(\Xi)\colon H^n(A^\bullet_J)\to H^n(\hat{A}^\bullet_J)$ agrees with $H^n(\Xi_N)\colon H^n(A^\bullet_{J,N})\to H^n(\hat{A}^\bullet_J)$ for all $n=0,-1,\dots,N+1$, so $H^n(\Xi)$ is an isomorphism for all $n\leqslant 0$ by part (c) of Hypothesis 4.3, and $\Xi\colon A_J^\bullet \to \hat{A}_J^\bullet$ is a quasi-isomorphism in $\mathbf{cdga}_{\mathbb{C}}$, and hence an equivalence in $\mathbf{cdga}_{\mathbb{C}}^\infty$. Thus $\mathbf{Spec}\,\Xi\colon \mathbf{Spec}\,\hat{A}_J^\bullet \to \mathbf{Spec}\,A_J^\bullet$ is an equivalence in $\mathbf{dSch}_{\mathbb{C}}$. So we can choose a quasi-inverse $\boldsymbol{\chi}\colon \mathbf{Spec}\,A_J^\bullet \to \mathbf{Spec}\,\hat{A}_J^\bullet$ in $\mathbf{dSch}_{\mathbb{C}}$.

Write $\iota\colon \tilde{A}_J^\bullet \hookrightarrow A_J^\bullet$ for the inclusion. Then $\Psi = \Xi \circ \iota\colon \tilde{A}_J^\bullet \to \hat{A}_J^\bullet$, since $\Xi_N|_{\tilde{A}_J^\bullet} = \Psi$, so taking the limit as $N \to -\infty$ gives $\Xi|_{\tilde{A}_J^\bullet} = \Psi$. Also the definitions of $\beta_J\colon B \to A_J^\bullet$ and $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ for $\varnothing \neq K \subsetneq J$ in parts (b) and (d) of Hypothesis 4.1 satisfy $\beta_J = \iota \circ \tilde{\beta}_J$ and $\Phi_{JK} = \iota \circ \tilde{\Phi}_{JK}$.

Define $\boldsymbol{\alpha}_J = \hat{\boldsymbol{\alpha}}_J \circ \boldsymbol{\chi}\colon \mathbf{Spec}\,A_J^\bullet \to \boldsymbol{X}$. Since $\hat{\boldsymbol{\alpha}}_J$ is a Zariski open inclusion with image $\bigcap_{i\in J} \operatorname{Im}\boldsymbol{\alpha}_i$, and $\boldsymbol{\chi}$ is an equivalence, $\boldsymbol{\alpha}_J\colon \mathbf{Spec}\,A_J^\bullet \to \boldsymbol{X}$ is a Zariski open inclusion with image $\bigcap_{i\in J} \operatorname{Im}\boldsymbol{\alpha}_i$, as in Hypothesis 4.1(g). Then we have

$$\begin{aligned}\boldsymbol{\pi}\circ\boldsymbol{\alpha}_J &= \boldsymbol{\pi}\circ\hat{\boldsymbol{\alpha}}_J\circ\boldsymbol{\chi}\\ &\simeq \mathbf{Spec}\,\tilde{\beta}_J\circ\mathbf{Spec}\,\Psi\circ\boldsymbol{\chi}\\ &\simeq \mathbf{Spec}\,\tilde{\beta}_J\circ\mathbf{Spec}\,\iota\circ\mathbf{Spec}\,\Xi\circ\boldsymbol{\chi} \simeq \mathbf{Spec}\,\tilde{\beta}_J\circ\mathbf{Spec}\,\iota = \mathbf{Spec}\,\beta_J,\end{aligned}$$

using (49) in the second step, $\Psi = \Xi\circ\iota$ in the third, $\mathbf{Spec}\,\Xi$, $\boldsymbol{\chi}$ quasi-inverse in the fourth, and $\beta_J = \iota\circ\tilde{\beta}_J$ in the fifth. Thus (13) homotopy commutes.

Similarly, if $\varnothing \neq K \subsetneq J$ then

$$\begin{aligned}\boldsymbol{\alpha}_J &= \hat{\boldsymbol{\alpha}}_J\circ\boldsymbol{\chi}\\ &\simeq \boldsymbol{\alpha}_K\circ\mathbf{Spec}\,\tilde{\Phi}_{JK}\circ\mathbf{Spec}\,\Psi\circ\boldsymbol{\chi}\\ &\simeq \boldsymbol{\alpha}_K\circ\mathbf{Spec}\,\tilde{\Phi}_{JK}\circ\mathbf{Spec}\,\iota\circ\mathbf{Spec}\,\Xi\circ\boldsymbol{\chi} \simeq \boldsymbol{\alpha}_K\circ\mathbf{Spec}\,\Phi_{JK},\end{aligned}$$

using (49) in the second step, $\Psi = \Xi\circ\iota$ in the third, and $\Phi_{JK} = \iota\circ\tilde{\Phi}_{JK}$ and $\mathbf{Spec}\,\Xi$, $\boldsymbol{\chi}$ quasi-inverse in the fourth. Hence (14) homotopy commutes.

This proves that Hypothesis 4.1 holds with $k = l+1$, and completes the inductive step begun shortly after Remark 4.2. Hence by induction, Hypothesis 4.1 holds for all $k = 1, 2, \ldots$ so Hypothesis 4.1 holds for $k = \infty$. Theorem 3.1 follows, since all the conclusions of Theorem 3.1(i)–(ii) are either part of Hypothesis 4.1, or for $A_{\{i\}}^\bullet = A_i^\bullet$, $\boldsymbol{\alpha}_{\{i\}} = \boldsymbol{\alpha}_i$, $\beta_{\{i\}} = \beta_i$ in part (i) were included in the first step of the induction. This completes the proof.

5 Proof of Theorem 3.7

5.1 Theorem 3.7(a): (∗) is an open condition

Suppose $\boldsymbol{X}$, $\omega_{\boldsymbol{X}}^*$, $A^\bullet$, $\boldsymbol{\alpha}$, V, E, F, s, t, ψ are as in Definition 3.6, and suppose that $U \subseteq V$ is open, E^- is a real vector subbundle of $E|_U$, and $v \in s^{-1}(0) \cap U$, such that the assumptions on $E^-|_v$ in condition $(*)$ hold at v. We must show that these assumptions also hold for all v' in an open neighbourhood of v in $s^{-1}(0)\cap U$. Suppose for a contradiction that this is false. Then we can choose a sequence $(v_i)_{i=1}^\infty$ in $s^{-1}(0)\cap U$ such that $v_i\to v$ as $i\to\infty$, and the assumptions on $E^-|_{v_i}$ in $(*)$ do not hold for any $i=1,2,\dots$.

By passing to a subsequence of $(v_i)_{i=1}^\infty$, we can assume $\dim\operatorname{Im}\mathrm{d}s|_{v_i}$ and $\dim\operatorname{Ker}t|_{v_i}$ are independent of $i=1,2,\dots$. By trivializing E near v, we can regard $(\operatorname{Im}\mathrm{d}s|_{v_i})_{i=1}^\infty$ and $(\operatorname{Ker}t|_{v_i})_{i=1}^\infty$ as sequences in complex Grassmannians, which are compact. Thus, passing to a subsequence of $(v_i)_{i=1}^\infty$, we can assume they converge, and there are complex vector subspaces $I_v,K_v\subseteq E|_v$ such that $\operatorname{Im}\mathrm{d}s|_{v_i}\to I_v$ and $\operatorname{Ker}t|_{v_i}\to K_v$ as $i\to\infty$.

Because $t\circ\mathrm{d}s=0$ on $s^{-1}(0)$ we have $\operatorname{Im}\mathrm{d}s|_{v_i}\subseteq\operatorname{Ker}t|_{v_i}$, and so $I_v\subseteq K_v$. Also $\operatorname{Im}\mathrm{d}s|_v\subseteq I_v$, since if $w\in T_vV$ we can find $w_i\in T_{v_i}V$ with $w_i\to w$ as $i\to\infty$, and then $\mathrm{d}s|_{v_i}(w_i)\to\mathrm{d}s|_v(w)$ as $i\to\infty$. Similarly $K_v\subseteq\operatorname{Ker}t|_v$.

We now have a quotient vector space $(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)$, which as in (32) carries a nondegenerate quadratic form $\tilde Q_v$. There are subspaces satisfying $I_v/(\operatorname{Im}\mathrm{d}s|_v)\subseteq K_v/(\operatorname{Im}\mathrm{d}s|_v)\subseteq(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)$. Also, for each $i\geqslant 1$ we have a quotient space $(\operatorname{Ker}t|_{v_i})/(\operatorname{Im}\mathrm{d}s|_{v_i})$ with quadratic forms $\tilde Q_{v_i}$. As $i\to\infty$ we have

$$(\operatorname{Ker}t|_{v_i})/(\operatorname{Im}\mathrm{d}s|_{v_i})\to K_v/I_v\cong[K_v/(\operatorname{Im}\mathrm{d}s|_v)]/[I_v/(\operatorname{Im}\mathrm{d}s|_v)]. \tag{53}$$

One can prove using a representative $\omega_{A^\bullet}$ for $\boldsymbol{\alpha}^*(\omega_{\boldsymbol{X}}^0)$ that

$$I_v/(\operatorname{Im}\mathrm{d}s|_v)=\{e\in(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)\mid\tilde Q_v(e,k)=0\text{ for all }k\in K_v/(\operatorname{Im}\mathrm{d}s|_v)\},$$

that is, $I_v/(\operatorname{Im}\mathrm{d}s|_v)$ and $K_v/(\operatorname{Im}\mathrm{d}s|_v)$ are orthogonal subspaces with respect to $\tilde Q_v$. Hence the restriction of $\tilde Q_v$ to $K_v/(\operatorname{Im}\mathrm{d}s|_v)$ is null along $I_v/(\operatorname{Im}\mathrm{d}s|_v)$, and descends to a nondegenerate quadratic form $\check Q_v$ on $[K_v/(\operatorname{Im}\mathrm{d}s|_v)]/[I_v/(\operatorname{Im}\mathrm{d}s|_v)]\cong K_v/I_v$. Then under the limit (53), we have $\tilde Q_{v_i}\to\check Q_v$ as $i\to\infty$.

By $(*)$ for (U,E^-) at v, we have $\operatorname{Im}(\mathrm{d}s|_v)\cap E^-|_v=\{0\}$, and the map Π_v in (35), $\Pi_v\colon E^-|_v\cap\operatorname{Ker}(t|_v)\to(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)$, has image $\operatorname{Im}\Pi_v$ of half the total dimension, with $\operatorname{Re}\tilde Q_v$ negative definite on $\operatorname{Im}\Pi_v$. Since $\tilde Q_v$ is zero on $I_v/(\operatorname{Im}\mathrm{d}s|_v)$, it follows that $\operatorname{Im}\Pi_v\cap(I_v/(\operatorname{Im}\mathrm{d}s|_v))=\{0\}$, and thus

$$E^-|_v\cap I_v=\{0\}. \tag{54}$$

Condition (34), that $t|_v(E^-|_v)=t|_v(E|_v)$, is equivalent to $E^-|_v+\operatorname{Ker}(t|_v)=E|_v$, in subspaces of $E|_v$. As $\operatorname{Im}\Pi_v$ is a maximal negative definite subspace for $\operatorname{Re}\tilde Q_v$ in $(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)$, and $K_v/(\operatorname{Im}\mathrm{d}s|_v)$ is the orthogonal to a null subspace $I_v/(\operatorname{Im}\mathrm{d}s|_v)$ with respect to $\operatorname{Re}\tilde Q_v$, it follows that $\operatorname{Im}\Pi_v+K_v/(\operatorname{Im}\mathrm{d}s|_v)=(\operatorname{Ker}t|_v)/(\operatorname{Im}\mathrm{d}s|_v)$.

Lifting to $\operatorname{Ker} t|_v$ gives $[E^-|_v \cap (\operatorname{Ker} t|_v)] + K_v = \operatorname{Ker} t|_v$. Thus the subspace $E^-|_v + K_v$ in $E|_v$ contains $E^-|_v$ and $\operatorname{Ker} t|_v$, so, as $E^-|_v + \operatorname{Ker}(t|_v) = E|_v$, we see that

$$E^-|_v + K_v = E|_v. \tag{55}$$

Write $\check{\Pi}_v\colon E^-|_v \cap K_v \to K_v/I_v$ for the natural projection. It is injective by (54). Using (54)–(55) and the facts that $\operatorname{Im}\Pi_v$ has half the dimension of $(\operatorname{Ker} t|_v)/(\operatorname{Im} \mathrm{d}s|_v)$, and

$$\dim[I_v/(\operatorname{Im} \mathrm{d}s|_v)] + \dim[K_v/(\operatorname{Im} \mathrm{d}s|_v)] = \dim[(\operatorname{Ker} t|_v)/(\operatorname{Im} \mathrm{d}s|_v)]$$

as $I_v/(\operatorname{Im} \mathrm{d}s|_v)$, $K_v/(\operatorname{Im} \mathrm{d}s|_v)$ are orthogonal subspaces, by a dimension count we find that $\operatorname{Im}\check{\Pi}_v$ has half the total dimension of K_v/I_v. Also, since the quadratic form $\check{Q}_v$ on $K_v/I_v \cong [K_v/(\operatorname{Im} \mathrm{d}s|_v)]/[I_v/(\operatorname{Im} \mathrm{d}s|_v)]$ descends from the restriction of $\widetilde{Q}_v$ to $K_v/(\operatorname{Im} \mathrm{d}s|_v)$, and $\operatorname{Im}\check{\Pi}_v$ descends from $\operatorname{Im}\Pi_v \cap [K_v/(\operatorname{Im} \mathrm{d}s|_v)]$, and $\operatorname{Re}\widetilde{Q}_v$ is negative definite on $\operatorname{Im}\Pi_v$, we see that $\operatorname{Re}\check{Q}_v$ is negative definite on $\operatorname{Im}\check{\Pi}_v$.

Because $E^-|_{v_i} \to E^-|_v$ and $\operatorname{Im} \mathrm{d}s|_{v_i} \to I_v$ as $i \to \infty$, we see from (54) that

$$E^-|_{v_i} \cap (\operatorname{Im} \mathrm{d}s|_{v_i}) = \{0\} \quad \text{for } i \gg 0. \tag{56}$$

Since $E^-|_{v_i} \to E^-|_v$ and $\operatorname{Ker} t|_{v_i} \to K_v$ as $i \to \infty$, we see from (55) that we have $E^-|_{v_i} + \operatorname{Ker} t|_{v_i} = E|_{v_i}$ for $i \gg 0$. But this is equivalent to

$$t|_{v_i}(E^-|_{v_i}) = t|_{v_i}(E|_{v_i}) \quad \text{in } F|_{v_i} \text{ for } i \gg 0. \tag{57}$$

Using (56)–(57), the same dimension count as above implies that $\operatorname{Im}\widetilde{\Pi}_{v_i}$ has half the dimension of $(\operatorname{Ker} t|_{v_i})/(\operatorname{Im} \mathrm{d}s|_{v_i})$ for $i \gg 0$. Under the limit (53), we have $\widetilde{Q}_{v_i} \to \check{Q}_v$ and $\operatorname{Im}\widetilde{\Pi}_{v_i} \to \operatorname{Im}\check{\Pi}_v$. Thus, as $\operatorname{Re}\check{Q}_v$ is negative definite on $\operatorname{Im}\check{\Pi}_v$, we see that $\operatorname{Re}\widetilde{Q}_{v_i}$ is negative definite on $\operatorname{Im}\widetilde{\Pi}_{v_i}$ for $i \gg 0$. Together with (56)–(57), this shows that the assumptions on $E^-|_{v_i}$ in $(*)$ hold for $i \gg 0$, which contradicts the choice of sequence $(v_i)_{i=1}^\infty$. This proves Theorem 3.7(a).

5.2 Theorem 3.7(b): extending pairs (U, E^-) satisfying $(*)$

Suppose X, ω_X^*, $A^\bullet$, $\boldsymbol{\alpha}$, V, E, F, s, t, ψ are as in Definition 3.6, and (U, E^-) satisfying $(*)$ is as in Definition 3.6, and $C \subseteq V$ is closed with $C \subseteq U$. Our goal is to construct $(\widetilde{U}, \widetilde{E}^-)$ satisfying $(*)$ for $V, E, \dots$ with $C \cup s^{-1}(0) \subseteq \widetilde{U} \subseteq V$, such that $E^-|_{U'} = \widetilde{E}^-|_{U'}$ for U' an open neighbourhood of C in $U \cap \widetilde{U}$.

Using the notation of Section 3.2, $s^{-1}(0)^{\mathrm{alg}}$ is a finite type closed $\mathbb{C}$–subscheme of V^{alg}, and the maps $v \mapsto \dim \operatorname{Ker} \mathrm{d}s|_v$ and $v \mapsto \dim \operatorname{Ker} t|_v$ are upper semicontinuous, algebraically constructible functions $s^{-1}(0)^{\mathrm{alg}} \to \mathbb{N}$, noting that $t|_v$ is independent of choices for $v \in s^{-1}(0)^{\mathrm{alg}}$. Therefore by some standard facts about constructible sets in algebraic geometry, we can choose a stratification of Zariski topological spaces $s^{-1}(0)^{\mathrm{alg}} = \coprod_{a\in A} W_a^{\mathrm{alg}}$, where A is a finite indexing set, and W_a^{alg} is a smooth, connected, locally closed $\mathbb{C}$–subscheme of $s^{-1}(0)^{\mathrm{alg}} \subseteq V^{\mathrm{alg}}$ for each $a \in A$, with closure $\overline{W}_a^{\mathrm{alg}}$ in $s^{-1}(0)^{\mathrm{alg}}$ a finite union of strata W_b, such that $v \mapsto \dim \operatorname{Ker} \mathrm{d}s|_v$ and $v \mapsto \dim \operatorname{Ker} t|_v$ are both constant functions on W_a^{alg}.

Writing $W_a \subseteq s^{-1}(0) \subseteq V$ for the set of $\mathbb{C}$–points of W_a^{alg}, each W_a is a connected, locally closed complex submanifold of V lying in $s^{-1}(0)$, with closure $\overline{W}_a$ a finite union of submanifolds W_b, such that $s^{-1}(0) = \coprod_{a\in A} W_a$. On each W_a, the maps $v \mapsto \dim \operatorname{Ker} \mathrm{d}s|_v$ and $v \mapsto \dim \operatorname{Ker} t|_v$ are constant. This implies that $\operatorname{Ker} \mathrm{d}s|_{W_a}$ is a holomorphic vector subbundle of $TV|_{W_a}$, and $\operatorname{Im} \mathrm{d}s|_{W_a}$ a holomorphic vector subbundle of $E|_{W_a}$, and $\operatorname{Ker} t|_{W_a}$ a holomorphic vector subbundle of $E|_{W_a}$, and $\operatorname{Im} t|_{W_a}$ a holomorphic vector subbundle of $F|_{W_a}$. Since $t \circ \mathrm{d}s = 0$ on $s^{-1}(0)$, we have $\operatorname{Im} \mathrm{d}s|_{W_a} \subseteq \operatorname{Ker} t|_{W_a} \subseteq E|_{W_a}$.

Thus we have a holomorphic vector bundle $(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})$ over W_a, whose fibre at $v \in W_a$ is identified with $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ for $x = \psi(v)$ by (20). As in (6) we have a quadratic form Q_x on $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$, and as in (32) $\widetilde{Q}_v$ is the quadratic form on $(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})|_v$ identified with Q_x by (20). One can prove using a representative $\omega_{A^\bullet}$ for $\boldsymbol{\alpha}^*(\omega_{\boldsymbol{X}}^0)$ that $\widetilde{Q}_v$ depends holomorphically on $v \in W_a$. Hence $\widetilde{Q}_v = \widetilde{Q}_a|_v$ for $\widetilde{Q}_a \in H^0(S^2[(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})]^*)$, a nondegenerate holomorphic quadratic form on the fibres of $(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})$.

The idea of the proof is to choose $\widetilde{E}^-$ near W_a by induction on increasing $\dim W_a$, starting with $a \in A$ with $\dim W_a = 0$, then $a \in A$ with $\dim W_a = 1$, and so on. Since $\dim(\overline{W}_a \setminus W_a) < \dim W_a$, we see that $\overline{W}_a \setminus W_a$ is a finite union of W_b with $\dim W_b < \dim W_a$, so when we choose $\widetilde{E}^-$ near W_a we will already have chosen $\widetilde{E}^-$ near $\overline{W}_a \setminus W_a$, and the extension over W_a should be compatible with this.

Our inductive hypothesis $(\ddagger)_m$ for $m = 0, 1, 2, \dots$ is:

$(\ddagger)_m$ For all $a \in A$ with $\dim W_a \leqslant m$ we have chosen a pair $(\check{U}_a, \check{E}_a^-)$ satisfying $(*)$ for V, E, F, s, $t, \dots$ with $W_a \subseteq \check{U}_a \subseteq V$, such that there is an open neighbourhood $\hat{U}_a$ of $C \cap \check{U}_a$ in $U \cap \check{U}_a$ with $E^-|_{\hat{U}_a} = \check{E}_a^-|_{\hat{U}_a}$, and if $b \in A$ with $W_b \subseteq \overline{W}_a \setminus W_a$ (which implies that $\dim W_b < \dim W_a \leqslant m$, so $(\check{U}_b, \check{E}_b^-)$ is defined), then there is an open neighbourhood $\hat{U}_{ab}$ of W_b in $\check{U}_b$ such that $\check{E}_a^-|_{\check{U}_a \cap \hat{U}_{ab}} = \check{E}_b^-|_{\check{U}_a \cap \hat{U}_{ab}}$.

First consider how to choose $(\check{U}_a, \check{E}_a^-)$ satisfying $(*)$ with $W_a \subseteq \check{U}_a \subseteq V$ for $a \in A$ with no compatibility conditions, either with (U, E^-) near C, or with $(\check{U}_b, \check{E}_b^-)$ for $W_b \subseteq \overline{W}_a \setminus W_a$. We can do this as follows:

(i) Choose a real vector subbundle $\dot{E}_a$ of $(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})$, whose real rank is half the real rank of $(\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})$, such that $\operatorname{Re} \widetilde{Q}_a$ is negative definite on $\dot{E}_a$.

(ii) Lift $\dot{E}_a$ to a real vector subbundle $\ddot{E}_a$ of $\operatorname{Ker} t|_{W_a}$. That is, the projection $\operatorname{Ker} t|_{W_a} \to (\operatorname{Ker} t|_{W_a})/(\operatorname{Im} \mathrm{d}s|_{W_a})$ induces an isomorphism $\ddot{E}_a \to \dot{E}_a$.

(iii) Choose a real vector subbundle $\dddot{E}_a$ of $E|_{W_a}$ with $E|_{W_a} = \dddot{E}_a \oplus \operatorname{Ker} t|_{W_a}$.

(iv) Set $\check{E}_a^-|_{W_a} = \ddot{E}_a \oplus \dddot{E}_a$. Then $\check{E}_a^-|_{W_a}$ is a real vector subbundle of $E|_{W_a}$, and the assumptions on $\check{E}_a^-|_v$ in condition $(*)$ in Section 3.3 hold for all $v \in W_a$.

(v) Choose any real vector subbundle $\check{E}_a^-$ of $E|_{\check{U}_a}$ on a small open neighbourhood $\check{U}_a$ of W_a in V, extending the given $\check{E}_a^-|_{W_a} = \ddot{E}_a \oplus \dddot{E}_a$ on W_a.

Observe that by Theorem 3.7(a), proved in Section 5.1, condition $(*)$ holds for $\check{E}_a^-$ on an open neighbourhood of W_a. So by making $\check{U}_a$ smaller, we can suppose $(\check{U}_a, \check{E}_a^-)$ satisfies $(*)$.

All of these steps are possible. Any $(\check{U}_a, \check{E}_a^-)$ satisfying $(*)$ with $W_a \subseteq \check{U}_a \subseteq V$ arises from steps (i)–(v) (though $\dddot{E}_a$ in (iii) is not uniquely determined by $\check{E}_a^-$). Furthermore (taking germs in (v)), the space of choices in each step is contractible.

Now suppose $m = 0, 1, \dots$ and $(\ddagger)_{m-1}$ holds if $m > 0$, and $a \in A$ with $\dim W_a = m$. To choose $(\check{U}_a, \check{E}_a^-)$ with the compatibility conditions required in $(\ddagger)_m$, we follow (i)–(v), but modified as follows. In step (i), we choose $\dot{E}_a$ with

$$\dot{E}_a|_{W_a \cap \hat{U}_a} = [((E^- \cap \operatorname{Ker} t)|_{W_a \cap \hat{U}_a}) + (\operatorname{Im} \mathrm{d}s|_{W_a \cap \hat{U}_a})]/(\operatorname{Im} \mathrm{d}s|_{W_a \cap \hat{U}_a}), \tag{58}$$

for some small open neighbourhood $\hat{U}_a$ of $C \cap W_a$ in U, and if $b \in A$ with $W_b \subseteq \overline{W}_a \setminus W_a$ then

$$\dot{E}_a|_{W_a \cap \check{U}_{ab}} = [((\check{E}_b^- \cap \operatorname{Ker} t|_{W_a \cap \hat{U}_{ab}})) + (\operatorname{Im} \mathrm{d}s|_{W_a \cap \hat{U}_{ab}})]/(\operatorname{Im} \mathrm{d}s|_{W_a \cap \hat{U}_{ab}}), \tag{59}$$

for some small open neighbourhood $\hat{U}_{ab}$ of W_b in $\check{U}_b$.

To see this is possible, first note that the first part of $(\ddagger)_{m-1}$ with b in place of a implies that (58) and (59) are compatible, that is they prescribe the same value for $\dot{E}_a$ on $W_a \cap \hat{U}_a \cap \hat{U}_{ab}$, provided the open neighbourhoods $\hat{U}_a$, $\hat{U}_{ab}$ are small enough. Also given distinct $b, b' \in A$ with $W_b, W_{b'} \subseteq \overline{W}_a \setminus W_a$, either (a) $W_{b'} \subseteq \overline{W}_b \setminus W_b$, or (b) $W_b \subseteq \overline{W}_{b'} \setminus W_{b'}$, or (c) $W_b \cap \overline{W}_{b'} = \overline{W}_b \cap W_{b'} = \varnothing$. In cases (a) and (b) we can use the second part of $(\ddagger)_{m-1}$ to show that (59) for b, b' are compatible provided $\hat{U}_{ab}$, $\hat{U}_{ab'}$ are small enough, and in case (c) we can choose $\hat{U}_{ab}$, $\hat{U}_{ab'}$ with $\hat{U}_{ab} \cap \hat{U}_{ab'} = \varnothing$, so compatibility is trivial.

Thus, if $\hat{U}_a$ and $\hat{U}_{ab}$ for all b are small enough then (58) and (59) for all b are compatible, and can be combined into a single equation prescribing $\dot{E}_a$ on $\check{W}_a := W_a \cap (\hat{U}_a \cup \bigcup_b \hat{U}_{ab})$. We then have to extend $\dot{E}_a$ from $\check{W}_a$ to W_a, satisfying the required conditions. This may not be possible: if we have chosen E^- or $\check{E}_b^-$ badly near the "edge" of $\check{W}_a$ in W_a, then the prescribed values of $\dot{E}_a$ may not extend continuously to the closure $\overline{\check{W}}_a$ of $\check{W}_a$ in W_a. However, we can deal with this problem by shrinking all the $\hat{U}_a$, $\hat{U}_{ab}$, such that the closure $\overline{\check{W}}_a$ of the new $\check{W}_a$ lies inside the old $\check{W}_a$. Then it is guaranteed that the prescribed value of $\dot{E}_a$ on $\check{W}_a$ extends smoothly to an open neighbourhood of $\overline{\check{W}}_a$ in W_a, so we can choose $\dot{E}_a$ on W_a satisfying all the required conditions (58)–(59).

In a similar way, for each of steps (ii)–(v) we can show that making the open neighbourhoods $\hat{U}_a$, $\hat{U}_{ab}$ smaller if necessary, we can make choices consistent with the compatibility conditions on $(\check{U}_a, \check{E}_a^-)$ in $(\ddagger)_m$. So by induction, $(\ddagger)_m$ holds for all $m = 0, 1, \ldots$. Fix data $(\check{U}_a, \check{E}_a^-)$, $\hat{U}_a$, $\hat{U}_{ab}$ satisfying $(\ddagger)_m$ for $m = \dim V$.

Next, choose open neighbourhoods U' of C in $U \subseteq V$ and $\tilde{U}_a$ of W_a in $\check{U}_a$ for each $a \in A$, such that $U' \cap \tilde{U}_a \subseteq \hat{U}_a$ for $a \in A$, and $\tilde{U}_a \cap \tilde{U}_b \subseteq \hat{U}_{ab}$ if $a, b \in A$ with $W_b \subseteq \overline{W}_a \setminus W_a$, and $\tilde{U}_a \cap \tilde{U}_b = \varnothing$ if $a, b \in A$ with $\overline{W}_a \cap W_b = W_a \cap \overline{W}_b = \varnothing$. This is possible provided U' and $\tilde{U}_a$ for $a \in A$ are all small enough.

Define $\tilde{U} = U' \cup \bigcup_{a \in A} \tilde{U}_a$, which is an open neighbourhood of $C \cup \bigcup_{a \in A} W_a = C \cup s^{-1}(0)$ in V. Define a vector subbundle $\tilde{E}^-$ of $E|_{\tilde{U}}$ by $\tilde{E}^-|_{U'} = E^-|_{U'}$ and $\tilde{E}^-|_{\tilde{U}_a} = \check{E}_a^-|_{\tilde{U}_a}$ for $a \in A$. These values agree on the overlaps $U' \cap \tilde{U}_a$ and $\tilde{U}_a \cap \tilde{U}_b$ by construction, so $\tilde{E}^-$ is well defined. Also $(\tilde{U}, \tilde{E}^-)$ satisfies $(*)$, since (U, E^-) and the $(\check{U}_a, \check{E}_a^-)$ do, and U' is an open neighbourhood of C in $U \cap \tilde{U}$ with $E^-|_{U'} = \tilde{E}^-|_{U'}$ by definition. This proves Theorem 3.7(b).

5.3 Theorem 3.7(c): $s^{-1}(0) = (s^+)^{-1}(0)$ locally in U

In Section 3.4 we explained how to pull back pairs (U_K, E_K^-) satisfying $(*)$ along a quasifree $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$. We can also *push forward* (U_J, E_J^-) along Φ_{JK}.

Definition 5.1 Let X, ω_X^*, n, $\Phi_{JK}\colon A_K^\bullet \to A_J^\bullet$ and V_J, $E_J, \ldots, \chi_{JK}$, ξ_{JK} be as in Definition 3.10, and suppose (U_J, E_J^-) satisfies $(*)$ for $A_J^\bullet$. Our goal is to construct (U_K, E_K^-) satisfying $(*)$ for $A_K^\bullet$, with $\psi_J(s_J^{-1}(0) \cap U_J) = \psi_K(s_K^{-1}(0) \cap U_K) \subseteq X_{\mathrm{an}}$, and if (U_J, E_J^-), (U_K, E_K^-) also satisfy $(\dagger)$, a coordinate change of Kuranishi neighbourhoods, as in Section 2.5:

$$(U_K, \theta_{KJ}, \eta_{KJ})\colon (U_K, E_K^+, s_K^+, \psi_K^+) \to (U_J, E_J^+, s_J^+, \psi_J^+). \tag{60}$$

Let $v_J \in s_J^{-1}(0) \cap U_J$ with $\phi_{JK}(v_J) = v_K \in s_K^{-1}(0) \subseteq V_K$ and $\psi_J(v_J) = \psi_K(v_K) = x \in X_{\mathrm{an}}$. We claim that we can choose splittings of real vector spaces

$$\begin{aligned} T_{v_J}V_J &= \widetilde{T}_{v_J}V_J \oplus T'_{v_J}V_J, & E_J|_{v_J} &= \widetilde{E}_J|_{v_J} \oplus E'_J|_{v_J} \oplus E''_J|_{v_J}, \\ E_J^-|_{v_J} &= \widetilde{E}_J^-|_{v_J} \oplus \widetilde{E}''_J|_{v_J}, & F_J|_{v_J} &= \widetilde{F}_J|_{v_J} \oplus F''_J|_{v_J} \oplus F'''_J|_{v_J}, \end{aligned} \tag{61}$$

fitting into a commutative diagram of the form

$$\begin{tikzcd} & & E_J^-|_{v_J} = \widetilde{E}_J^-|_{v_J} \oplus \widetilde{E}''_J|_{v_J} \arrow[d, "\mathrm{inc}"'] \arrow[dr, "t_J|_{E_J^-|_{v_J}}"] & & \\ 0 \arrow[r] & \begin{matrix}\widetilde{T}_{v_J}V_J \oplus \\ T'_{v_J}V_J\end{matrix} \arrow[r, "\mathrm{d}s_J|_{v_J}"] \arrow[d, "\mathrm{d}\phi_{JK}|_{v_J}"] & \begin{matrix}\widetilde{E}_J|_{v_J} \oplus \\ E'_J|_{v_J} \oplus \\ E''_J|_{v_J}\end{matrix} \arrow[r, "t_J|_{v_J}"] \arrow[d, "\chi_{JK}|_{v_J}"] & \begin{matrix}\widetilde{F}_J|_{v_J} \oplus \\ F''_J|_{v_J} \oplus \\ F'''_J|_{v_J}\end{matrix} \arrow[r] \arrow[d, "\xi_{JK}|_{v_J}"] & \cdots \\ 0 \arrow[r] & T_{v_K}V_K \arrow[r, "\mathrm{d}s_K|_{v_K}"] & E_K|_{v_K} \arrow[r, "t_K|_{v_K}"] & F_K|_{v_K} \arrow[r] & \cdots \end{tikzcd} \tag{62}$$

where

$$\mathrm{inc} = \begin{pmatrix} * & * \\ 0 & * \\ 0 & * \end{pmatrix}, \quad t_J|_{E_J^-|_{v_J}} = \begin{pmatrix} * & 0 \\ 0 & \cong \\ 0 & 0 \end{pmatrix}, \quad \mathrm{d}s_J|_{v_J} = \begin{pmatrix} \widetilde{\mathrm{d}s_K|_{v_K}} & 0 \\ * & \cong \\ 0 & 0 \end{pmatrix}, \quad t_J|_{v_J} = \begin{pmatrix} \widetilde{t_K|_{v_K}} & 0 & 0 \\ 0 & 0 & \cong \\ 0 & 0 & 0 \end{pmatrix},$$

$$\mathrm{d}\phi_{JK}|_{v_J} = \begin{pmatrix} \cong & 0 \end{pmatrix}, \quad \chi_{JK}|_{v_J} = \begin{pmatrix} \cong & 0 & 0 \end{pmatrix}, \quad \xi_{JK}|_{v_J} = \begin{pmatrix} \cong & 0 & 0 \end{pmatrix}.$$

To prove this, note that the rows of (62) are $\mathbb{T}_{\mathbf{Spec}\,A_J^\bullet}|_{v_J}$, $\mathbb{T}_{\mathbf{Spec}\,A_K^\bullet}|_{v_K}$, and are complexes, and the lower columns are induced by Φ_{JK}, are surjective as Φ_{JK} is quasifree, and induce isomorphisms on cohomology as in Section 3.2. Then:

(i) Define $T'_{v_J}V_J = \operatorname{Ker} \mathrm{d}\phi_{JK}|_{v_J}$.

(ii) Choose arbitrary $\widetilde{T}_{v_J}V_J$ with $T_{v_J}V_J \cong \widetilde{T}_{v_J}V_J \oplus T'_{v_J}V_J$. Then $\widetilde{T}_{v_J}V_J \cong T_{v_K}V_K$ as $\mathrm{d}\phi_{JK}$ is surjective.

(iii) Define $E'_J|_{v_J} = \mathrm{d}s_J|_{v_J}[T'_{v_J}V_J]$. Then $E'_J|_{v_J} \cong T'_{v_J}V_J$ as the columns of (62) are isomorphisms in cohomology, and $E'_J|_{v_J} \subseteq \operatorname{Ker}(\chi_{JK}|_{v_J})$ as the left-hand square of (62) commutes.

(iv) Choose $E''_J|_{v_J}$ with $\operatorname{Ker}(\chi_{JK}|_{v_J}) = E'_J|_{v_J} \oplus E''_J|_{v_J}$.

(v) Since the columns of (62) are isomorphisms on cohomology, $t_J|_{v_J}$ is injective on $E''_J|_{v_J}$. Define $F''_J|_{v_J} = t_J|_{v_J}[E''_J|_{v_J}]$. Then $F''_J|_{v_J} \cong E''_J|_{v_J}$. Also $F''_J|_{v_J} \subseteq \operatorname{Ker} \xi_{JK}|_{v_J}$, as the right-hand square of (62) commutes.

(vi) Choose $F_J'''|_{v_J}$ with $\operatorname{Ker}\xi_{JK}|_{v_J} = F_J''|_{v_J} \oplus F_J'''|_{v_J}$.

(vii) Since the columns of (62) are isomorphisms on cohomology, we have

$$\begin{aligned} F_J''|_{v_J} &= t_J|_{v_J}[E_J'|_{v_J} \oplus E_J''|_{v_J}] = t_J|_{v_J}[\operatorname{Ker}\chi_{JK}|_{v_J}] \\ &= \operatorname{Ker}\xi_{JK}|_{v_J} \cap \operatorname{Im} t_J|_{v_J} = (F_J''|_{v_J} \oplus F_J'''|_{v_J}) \cap \operatorname{Im} t_J|_{v_J}. \end{aligned}$$

Thus we may choose $\tilde F_J|_{v_J}$ with $F_J|_{v_J} = \tilde F_J|_{v_J} \oplus F_J''|_{v_J} \oplus F_J'''|_{v_J}$ and $\operatorname{Im} t_J|_{v_J} \subseteq \tilde F_J|_{v_J} \oplus F_J''|_{v_J}$. So the third row of $t_J|_{v_J}$ in (62) is zero. Also $\tilde F_J|_{v_J} \cong F_K|_{v_K}$ by (vi) as ξ_{JK} is surjective.

(viii) Set $\tilde E_J^-|_{v_J} = E_J^-|_{v_J} \cap t_J|_{v_J}^{-1}(\tilde F_J|_{v_J})$. We claim $\chi_{JK}|_{v_J}$ is injective on $\tilde E_J^-|_{v_J}$. To see this, note that we have an exact sequence

$$0 \longrightarrow E_J^-|_{v_J} \cap \operatorname{Ker} t_J|_{v_J} \longrightarrow \tilde E_J^-|_{v_J} \longrightarrow t_J|_{v_J}[E_J^-|_{v_J}] \cap \tilde F_J|_{v_J} \longrightarrow 0,$$

since $\operatorname{Ker} t_J|_{v_J} \subseteq t_J|_{v_J}^{-1}(\tilde F_J|_{v_J})$. The last part of $(*)$ implies that $\chi_{JK}|_{v_J}$ maps $E_J^-|_{v_J} \cap \operatorname{Ker} t_J|_{v_J}$ injectively into $\operatorname{Ker} t_K|_{v_K}$. Also $\xi_{JK}|_{v_J}$ is injective on $\tilde F_J|_{v_J}$, and the right square of (62) commutes, so the claim follows.

(ix) Choose $\tilde E_J|_{v_J} \subseteq E_J|_{v_J}$ such that

$$\tilde E_J^-|_{v_J} \subseteq \tilde E_J|_{v_J} \quad \text{and} \quad E_J|_{v_J} = \tilde E_J|_{v_J} \oplus \operatorname{Ker}(\chi_{JK}|_{v_J}) \stackrel{\text{(iv)}}{=} \tilde E_J|_{v_J} \oplus E_J'|_{v_J} \oplus E_J''|_{v_J}$$

and $t_J|_{v_J}[\tilde E_J|_{v_J}] \subseteq \tilde F_J|_{v_J}$. This is possible as $\chi_{JK}|_{v_J}$ is injective on $\tilde E_J^-|_{v_J}$, and using (v), (vii) and (viii). Then $\tilde E_J|_{v_J} \cong E_K|_{v_K}$ as χ_{JK} is surjective.

(x) Choose $\tilde E_J''|_{v_J}$ such that $E_J^-|_{v_J} = \tilde E_J^-|_{v_J} \oplus \tilde E_J''|_{v_J}$ and $t_J|_{v_J}[\tilde E_J''|_{v_J}] \subseteq F_J''|_{v_J}$. This is possible by (viii) and because $\operatorname{Im} t_J|_{v_J} \subseteq \tilde F_J|_{v_J} \oplus F_J''|_{v_J}$.

Since $t_J|_{v_J}(E_J^-|_{v_J}) = t_J|_{v_J}(E_J|_{v_J})$ by (34) and $F_J''|_{v_J} = t_J|_{v_J}[E_J''|_{v_J}]$, we see that $t_J|_{v_J}[\tilde E_J''|_{v_J}] = F_J''|_{v_J}$. Also $t_J|_{v_J}\colon \tilde E_J''|_{v_J} \to F_J''|_{v_J}$ is injective, as, by (viii), $\operatorname{Ker} t_J|_{E_J^-|_{v_J}} \subseteq \tilde E_J^-|_{v_J}$. Hence $\tilde E_J''|_{v_J} \cong F_J''|_{v_J}$.

We can do all this, not just at one $v_J \in s_J^{-1}(0) \cap U_J$, but in an open neighbourhood U_J' of $s_J^{-1}(0) \cap U_J$ in U_J. That is, we can choose U_J', and splittings

$$\begin{aligned} TV_J|_{U_J'} &= \tilde T V_J \oplus T'V_J, & E_J|_{U_J'} &= \tilde E_J \oplus E_J' \oplus E_J''|_{v_J}, \\ E_J^-|_{U_J'} &= \tilde E_J^- \oplus \tilde E_J'', & F_J|_{U_J'} &= \tilde F_J \oplus F_J'' \oplus F_J''', \end{aligned} \tag{63}$$

with $\tilde E_J^- \subseteq \tilde E_J$, such that (62) holds at each $v_J \in s_J^{-1}(0) \cap U_J$. To see this, note that the argument above can be carried out on $s_J^{-1}(0) \cap U_J$ regarded as a C^∞–subscheme of U_J, in the sense of C^∞–algebraic geometry in [17], and the splittings (63) with $\tilde E_J^- \subseteq \tilde E_J$ can then be extended from $s_J^{-1}(0) \cap U_J$ to an open neighbourhood U_J'. Making U_J' smaller, we can suppose that the component of χ_{JK} mapping $\tilde E_J \to \phi_{JK}|_{U_J'}^*(E_K)$

is an isomorphism. We can also choose the splittings so that away from $s_J^{-1}(0)\cap U_J$, the map $t_J|_{U'_J}$ has the form

$$t_J|_{U'_J}=\begin{pmatrix} * & * & 0\\ * & * & \cong\\ * & * & 0\end{pmatrix}\colon \tilde E_J|_{v_J}\oplus E'_J\oplus E''_J\to \tilde F_J\oplus F''_J\oplus F'''_J. \tag{64}$$

Write $s_J|_{U'_J}=\tilde s_J\oplus s'_J\oplus s''_J$, for $\tilde s_J\in C^\infty(\tilde E_J)$, $s'_J\in C^\infty(E'_J)$ and $s''_J\in C^\infty(E''_J)$. Then (64) and $t_J\circ s_J=0$ together imply that $s''_J=0$. From (62) we see that $\mathrm{d}s'_J|_{v_J}\colon T_{v_J}V_J\to E'_J|_{v_J}$ is surjective and $\mathrm{d}\phi_{JK}|_{v_J}\colon \mathrm{Ker}(\mathrm{d}s'_J|_{v_J})\to T_{v_K}V_K$ is an isomorphism, at each $v_J\in s_J^{-1}(0)\cap U_J$. Hence s'_J is transverse near v_J, so that $(s'_J)^{-1}(0)$ is an embedded submanifold of V_J near v_J with tangent space $\mathrm{Ker}(\mathrm{d}s'_J|_{v_J})$ at v_J, and $\phi_{JK}|_{(s'_J)^{-1}(0)}\colon (s'_J)^{-1}(0)\to V_K$ is a local diffeomorphism near v_J. Thus, making U'_J smaller, we can suppose that s'_J is transverse on U'_J, so that $(s'_J)^{-1}(0)$ is an embedded submanifold of U'_J, and $\phi_{JK}|_{(s''_J)^{-1}(0)}\colon (s'_J)^{-1}(0)\to V_K$ is a local diffeomorphism. But ϕ_{JK} is injective on $s_J^{-1}(0)\cap U_J$, so making U'_J smaller, we can also suppose $\phi_{JK}|_{(s'_J)^{-1}(0)}$ is a diffeomorphism with an open set U_K in V_K, with inverse $\theta_{KJ}\colon U_K\xrightarrow{\cong}(s'_J)^{-1}(0)\subseteq U'_J\subseteq U_J$.

We now have a vector bundle $\theta_{KJ}^*(E_J)$ over U_K, and we have vector subbundles $\theta_{KJ}^*(\tilde E_J,E'_J,E''_J,E_J^-,\tilde E_J^-,\tilde E''_J)$ with $\theta_{KJ}^*(E_J)=\theta_{KJ}^*(\tilde E_J)\oplus\theta_{KJ}^*(E'_J)\oplus\theta_{KJ}^*(E''_J)$, $\theta_{KJ}^*(E_J^-)=\theta_{KJ}^*(\tilde E_J^-)\oplus\theta_{KJ}^*(\tilde E''_J)$ and $\theta_{KJ}^*(\tilde E_J^-)\subseteq\theta_{KJ}^*(\tilde E_J)$. Since $\phi_{JK}\circ\theta_{KJ}=\mathrm{id}_{U_K}$, pulling back $\chi_{JK}\colon E_J\to\phi_{JK}^*(E_K)$ by θ_{KJ} gives a surjective vector bundle morphism $\theta_{KJ}^*(\chi_{JK})\colon\theta_{KJ}^*(E_J)\to E_K|_{U_K}$, where $\theta_{KJ}^*(\chi_{JK})$ restricts to an isomorphism $\theta_{KJ}^*(\tilde E_J)\to E_K$. We also have a section $\theta_{KJ}^*(s_J)$ of $\theta_{KJ}^*(E_J)$, whose components in $\theta_{KJ}^*(\tilde E_J)$, $\theta_{KJ}^*(E'_J)$, $\theta_{KJ}^*(E''_J)$ are $\theta_{KJ}^*(\tilde s_J)$, 0, 0. Applying θ_{KJ}^* to (25) and using $E''_J\subseteq\mathrm{Ker}\,\chi_{JK}$ shows that

$$\theta_{KJ}^*(\chi_{JK})[\theta_{KJ}^*(s_J)]=\theta_{KJ}^*(\chi_{JK})[\theta_{KJ}^*(\tilde s_J)]=s_K|_{U_K}. \tag{65}$$

Define a vector subbundle $E_K^-\subseteq E_K|_{U_K}$ by $E_K^-=\theta_{KJ}^*(\chi_{JK})[\theta_{KJ}^*(\tilde E_J^-)]$. This is valid as $\theta_{KJ}^*(\tilde E_J^-)\subseteq\theta_{KJ}^*(\tilde E_J)$, and $\theta_{KJ}^*(\chi_{JK})$ is an isomorphism on $\theta_{KJ}^*(\tilde E_J)$. We claim that (U_K,E_K^-) satisfies condition $(*)$. To see this, let $v_K\in s_K^{-1}(0)\cap U_K$, and set $v_J=\theta_{KJ}(v_K)$. Then $v_J\in s_J^{-1}(0)\cap U'_J$ with $\phi_{JK}(v_J)=v_K$, so (61)–(62) hold, with the columns of (62) isomorphisms on cohomology. From this and $(*)$ for (U_J,E_J^-) at v_J, we can deduce $(*)$ for (U_K,E_K^-) at v_K.

Writing $E_J^+=E_J|_{U_J}/E_J^-$, $s_J^+=s_J+E_J^-\in C^\infty(E_J^+)$, and similarly for E_K^+, s_K^+, define a vector bundle morphism

$$\eta_{KJ}\colon E_K^+\to\theta_{KJ}^*(E_J^+),\quad \eta_{KJ}\colon e_K+E_K^-\mapsto\theta_{KJ}^*(\chi_{JK})|^{-1}_{\theta_{KJ}^*(\tilde E_J)}[e_K]+\theta_{KJ}^*(E_J^-).$$

This is well defined as $\theta_{KJ}^*(\chi_{JK})|_{\theta_{KJ}^*(\tilde{E}_J)}\colon \theta_{KJ}^*(\tilde{E}_J) \to E_K$ is an isomorphism, with inverse

$$\theta_{KJ}^*(\chi_{JK})|_{\theta_{KJ}^*(\tilde{E}_J)}^{-1}\colon E_K \to \theta_{KJ}^*(\tilde{E}_J),$$

which, by definition of E_K^-, maps $E_K^- \to \theta_{KJ}^*(\tilde{E}_J^-) \subseteq \theta_{KJ}^*(E_J^-)$. Also (65) implies that $\eta_{KJ}(s_K^+) = \theta_{KJ}^*(s_J^+)$. Using (62) we can also show that the analogue of (8) for θ_{KJ}, η_{KJ} at v_K is exact. Therefore, if (U_J, E_J^-), (U_K, E_K^-) also satisfy (†), then $(U_K, \theta_{KJ}, \eta_{KJ})$ in (60) is a coordinate change. This completes Definition 5.1.

We now prove Theorem 3.7(c). Suppose $\boldsymbol{X}, \omega_{\boldsymbol{X}}^*, A^\bullet, \boldsymbol{\alpha}, V, E, F, s, t, \psi$ and (U, E^-) satisfying $(*)$ are as in Definition 3.6. Then $\boldsymbol{X}' := \boldsymbol{\alpha}(\mathbf{Spec}\, A^\bullet) \subseteq \boldsymbol{X}$ is an affine derived $\mathbb{C}$–subscheme of $\boldsymbol{X}$. Let $v \in s^{-1}(0) \cap U$, and set $x = \psi(v) \in X_{\mathrm{an}}$. Write $(A_1^\bullet, \boldsymbol{\alpha}_1) = (A^\bullet, \boldsymbol{\alpha})$, $V_1 = V$, $E_1 = E$, $v_1 = v$ and so on. Applying Theorem 2.10 to $(\boldsymbol{X}', \omega_{\boldsymbol{X}}^*|_{\boldsymbol{X}'})$ at x gives a pair $(A_2^\bullet, \omega_{A_2^\bullet})$ in -2–Darboux form and a Zariski open inclusion $\boldsymbol{\alpha}_2\colon \mathbf{Spec}\, A_2^\bullet \hookrightarrow \boldsymbol{X}' \subseteq \boldsymbol{X}$ which is minimal at $x \in \mathrm{Im}\, \boldsymbol{\alpha}_2$ with $\boldsymbol{\alpha}_2^*(\omega_{\boldsymbol{X}}^*) \simeq \omega_{A_2^\bullet}$. Section 3.2 applied to $A_2^\bullet, \boldsymbol{\alpha}_2$ gives $V_2, E_2, s_2, \dots$. Set $v_2 = \psi_2^{-1}(x) \in s_2^{-1}(0) \subseteq V_2$.

Applying Theorem 3.1 to the derived $\mathbb{C}$–scheme $\boldsymbol{X}'$ with $I = \{1,2\}$ and initial data $\{(A_1^\bullet, \boldsymbol{\alpha}_1), (A_2^\bullet, \boldsymbol{\alpha}_2)\}$ gives $(A_{12}^\bullet, \boldsymbol{\alpha}_{12})$ with image $\mathrm{Im}\, \boldsymbol{\alpha}_{12} = \mathrm{Im}\, \boldsymbol{\alpha}_1 \cap \mathrm{Im}\, \boldsymbol{\alpha}_2$ and quasifree morphisms $\Phi_{12,1}\colon A_1^\bullet \to A_{12}^\bullet$, $\Phi_{12,2}\colon A_2^\bullet \to A_{12}^\bullet$ such that (14) homotopy commutes in $\mathbf{dSch}_{\mathbb{C}}$. Section 3.2 applied to $A_{12}^\bullet$ gives $V_{12}, E_{12}, s_{12}, \dots$ and to $\Phi_{12,1}$ and $\Phi_{12,2}$ gives $\phi_{12,1}\colon V_{12} \to V_1 = V$, $\chi_{12,1}$, $\xi_{12,1}$ and $\phi_{12,2}\colon V_{12} \to V_2$, $\chi_{12,2}$, $\xi_{12,2}$, simplifying notation a little. Set $v_{12} = \psi_{12}^{-1}(x) \in s_{12}^{-1}(0) \subseteq V_{12}$, so that $\phi_{12,1}(v_{12}) = v_1$ and $\phi_{12,2}(v_{12}) = v_2$.

We have (U, E^-) satisfying $(*)$ for $A_1^\bullet, \boldsymbol{\alpha}_1, V_1, E_1, s_1, \dots$. Thus by Lemma 3.12, we can choose (U_{12}, E_{12}^-) satisfying $(*)$ for $V_{12}, E_{12}, s_{12}, \dots$ and compatible with (U, E^-) under $\phi_{12,1}$ and $\chi_{12,1}$ in the sense of Section 3.4, such that $v_{12} \in s_{12}^{-1}(0) \cap \phi_{12,1}^{-1}(U) \subseteq U_{12} \subseteq V_{12}$. Also Section 3.4 defines $\chi_{12,1}^+$ such that if (U, E^-) and (U_{12}, E_{12}^-) satisfy (†) (we do *not* assume this), then

$$(U_{12}, \phi_{12,1}|_{U_{12}}, \chi_{12,1}^+)\colon (U_{12}, E_{12}^+, s_{12}^+, \psi_{12}^+) \to (U, E^+, s^+, \psi^+)$$

is a coordinate change of Kuranishi neighbourhoods, as in Corollary 3.11.

Now apply Definition 5.1 to push forward (U_{12}, E_{12}^-) in $V_{12}, E_{12}, s_{12}, \dots$ along $\phi_{12,2}, \chi_{12,2}, \xi_{12,2}$. This yields (U_2, E_2^-) satisfying $(*)$ for $V_2, E_2, s_2, \dots$ with $\phi_{12,2}(s_{12}^{-1}(0) \cap U_{12}) \subseteq U_2 \subseteq V_2$, so in particular $v_2 \in U_2$, and data $\theta_{2,12}, \eta_{2,12}$ such that if (U_2, E_2^-) and (U_{12}, E_{12}^-) satisfy (†) (we do *not* assume this), then

$$(U_2, \theta_{2,12}, \eta_{2,12})\colon (U_2, E_2^+, s_2^+, \psi_2^+) \to (U_{12}, E_{12}^+, s_{12}^+, \psi_{12}^+) \tag{66}$$

is a coordinate change of Kuranishi neighbourhoods, as in (60).

Since $(A_2^\bullet, \omega_{A_2^\bullet})$ is in -2–Darboux form and minimal at x, Example 3.8 proves that there exists an open neighbourhood U_2' of v_2 in U_2 such that $s_2^{-1}(0) \cap U_2' = (s_2^+)^{-1}(0) \cap U_2'$. Then $(U_2', E_2^-|_{U_2'})$ satisfies (†). The construction in Definition 5.1 implies that $\theta_{2,12}$ identifies $s_2^{-1}(0)$ near v_2 with $s_{12}^{-1}(0)$ near v_{12}, and identifies $(s_2^+)^{-1}(0)$ near v_2 with $(s_{12}^+)^{-1}(0)$ near v_{12} (the second follows from the fact that the analogue of (8) for $\theta_{2,12}$, $\eta_{2,12}$ at v_2, v_{12} is exact, so (66) is a coordinate change of Kuranishi neighbourhoods near v_2, v_{12}). Since $s_2^{-1}(0) = (s_2^+)^{-1}(0)$ near v_2, it follows that $s_{12}^{-1}(0) = (s_{12}^+)^{-1}(0)$ near v_{12}. That is, there exists an open neighbourhood U_{12}' of v_{12} in U_{12} such that $s_{12}^{-1}(0) \cap U_{12}' = (s_{12}^+)^{-1}(0) \cap U_{12}'$.

Similarly, we have that $\phi_{12,1}$ identifies $s_{12}^{-1}(0)$ near v_{12} with $s^{-1}(0)$ near v, and identifies $(s_{12}^+)^{-1}(0)$ near v_{12} with $(s^+)^{-1}(0)$ near v, so there exists an open neighbourhood U_v' of v in U such that $s^{-1}(0) \cap U_v' = (s^+)^{-1}(0) \cap U_v'$. This holds for all $v \in s^{-1}(0) \cap U$. Define $U' = \bigcup_{v \in s^{-1}(0)} U_v'$. Then U' is an open neighbourhood of $s^{-1}(0) \cap U$ in U, and $s^{-1}(0) \cap U' = (s^+)^{-1}(0) \cap U'$. Theorem 3.7(c) follows.

6 Proofs of some auxiliary results

Next we prove Propositions 3.13, 3.14 and 3.17.

6.1 Proof of Proposition 3.13

Let Z be a paracompact, Hausdorff topological space and $\{R_i \mid i \in I\}$ an open cover of Z. By paracompactness we can choose a locally finite refinement $\{S_i \mid i \in I\}$. That is, $S_i \subseteq R_i \subseteq Z$ is open with $\bigcup_{i \in I} S_i = Z$, and each $z \in Z$ has an open $z \in U_z \subseteq Z$ with $U_z \cap S_i \neq \varnothing$ for only finitely many $i \in I$.

By a standard result in topology known as the shrinking lemma, we can choose open sets $T_i^1 \subseteq Z$ with closures $\overline{T}_i^1 \subseteq Z$ for $i \in I$ such that $T_i^1 \subseteq \overline{T}_i^1 \subseteq S_i$ for $i \in I$ and $\bigcup_{i \in I} T_i^1 = Z$. The next part of the proof broadly follows that of McDuff and Wehrheim [29, Lemma 7.1.7], who prove a similar result with Z compact and I finite. By induction on $k = 2, 3, \dots$ choose open $Ti^k \subseteq Z$ with

$$T_i \subseteq \overline{T}_i^1 \subseteq T_i^2 \subseteq \overline{T}_i^2 \subseteq T_i^3 \subseteq \overline{T}_i^3 \subseteq \cdots \subseteq S_i \subseteq Z \tag{67}$$

for $i \in I$. Here to choose T_i^k we note that Z is normal as it is paracompact and Hausdorff, so we can choose open $T_i^k, U \subseteq Z$ with $\overline{T}_i^{k-1} \subseteq T_i^k$, $Z \setminus S_i \subseteq U$ and $T_i^k \cap U = \varnothing$. Then $T_i^k \subseteq Z \setminus U \subseteq S_i$, and $Z \setminus U$ is closed, so we have $\overline{T}_i^k \subseteq S_i$.

Now for each finite $\varnothing \neq J \subseteq I$, define a closed subset $C_J \subseteq Z$ by

$$C_J = \bigcap_{j \in J} \overline{T}_j^{|J|} \setminus \bigcap_{i \in I \setminus J} T_i^{|J|+1}. \tag{68}$$

Then part (i) of the proposition follows from $\overline{T}{}_j^{|J|} \subseteq S_j \subseteq R_j$ for $j \in J$ by (67), and (ii) from $\{S_i \mid i \in I\}$ locally finite with $C_J \subseteq \bigcap_{i \in I} S_i$. For (iii), suppose $\varnothing \neq J, K \subseteq I$ are finite with $J \not\subseteq K$ and $K \not\subseteq J$. Without loss of generality, suppose $|J| \leqslant |K|$. Then there exists $j \in J \setminus K$, and (68) gives $C_J \subseteq \overline{T}{}_j^{|J|}$ and $C_K \subseteq Z \setminus T_j^{|K|+1}$, which forces $C_J \cap C_K = \varnothing$ as $\overline{T}{}_j^{|J|} \subseteq T_j^{|K|+1}$ by (67).

For part (iv), if $z \in Z$, define

$$J_z = \bigcup_{\substack{J \subseteq I \text{ finite} \\ z \in \bigcap_{j \in J} \overline{T}{}_j^{|J|}}} J. \tag{69}$$

Then J_z is finite since $\{S_i \mid i \in I\}$ is locally finite, so $z \in S_j$ for only finitely many $j \in I$, and J_z is nonempty as $\{T_i^1 \mid i \in I\}$ covers Z, so $z \in T_i^1 \subseteq \overline{T}{}_i^2$ for some $i \in I$, and $J = \{i\}$ is a possible set in the union (69). If $j \in J_z$ then $j \in J$ for some J in the union (69), so that $z \in \overline{T}{}_j^{|J|} \subseteq \overline{T}{}_j^{|J_z|}$ as $|J| \leqslant |J_z|$. If $i \in I \setminus J_z$ then we have that $z \notin \bigcap_{j \in J_z \cup \{i\}} \overline{T}{}_j^{|J_z|+1}$, as $J_z \cup \{i\}$ is not one of the sets J in (69), but $z \in \bigcap_{j \in J_z} \overline{T}{}_j^{|J_z|+1}$, so we conclude that $z \notin \overline{T}{}_i^{|J_z|+1}$. Hence $z \in C_{J_z}$ by (68), and part (iv) follows. This completes the proof of Proposition 3.13.

6.2 Proof of Proposition 3.14

We work in the situation of Section 3.5 just after Remark 3.28, so that we have data X_{an}, I, V_J, E_J, s_J, ψ_J and $C_J \subseteq R_J = \bigcap_{i \in J} R_i \subseteq X_{\mathrm{an}}$ for all $J \in A$, and ϕ_{JK}, χ_{JK} for all $J, K \in A$ with $K \subsetneq J$. We will first prove the following inductive hypothesis $(+)_m$, by induction on $m = 1, 2, \ldots$:

$(+)_m$ For all $J \in A$ with $|J| \leqslant m$, we can choose $(\widetilde{U}_J, \widetilde{E}_J^-)$ satisfying condition $(*)$ for $A_J^\bullet$, V_J, E_J, F_J, s_J, t_J, $\psi_J, \ldots$ such that $\psi_J^{-1}(C_J) \subseteq \widetilde{U}_J \subseteq V_J$, and if $J, K \in A$ with $K \subsetneq J$ and $0 < |K| < |J| \leqslant m$ then there exists open $\widetilde{U}_{JK} \subseteq \widetilde{U}_J$ with $\psi_J^{-1}(C_J \cap C_K) \subseteq \widetilde{U}_{JK}$ such that, in the sense of Section 3.4, $(\widetilde{U}_{JK}, \widetilde{E}_J^-|_{\widetilde{U}_{JK}})$ is compatible with $(\widetilde{U}_K, \widetilde{E}_K^-)$. That is, $\phi_{JK}(\widetilde{U}_{JK}) \subseteq \widetilde{U}_K \subseteq V_K$ and $\chi_{JK}|_{\widetilde{U}_{JK}}(\widetilde{E}_J^-|_{\widetilde{U}_{JK}}) \subseteq \phi_{JK}|_{\widetilde{U}_{JK}}^*(\widetilde{E}_K^-) \subseteq \phi_{JK}|_{\widetilde{U}_{JK}}^*(E_K)$.

For the first step, to prove $(+)_1$ for all $J = \{i\}$ with $i \in I$, we choose $(\widetilde{U}_J, \widetilde{E}_J^-)$ for $A_J^\bullet$, V_J, $E_J, \ldots$ satisfying $(*)$ with $s_J^{-1}(0) \subseteq \widetilde{U}_J$, so that $\psi_J^{-1}(C_J) \subseteq \widetilde{U}_J$, by applying Theorem 3.7(b) with $C = U = \varnothing$. The second part of $(+)_1$ is trivial, as there are no $J, K \in A$ with $0 < |K| < |J| \leqslant 1$.

For the inductive step, suppose $(+)_{m-1}$ holds for some $m > 1$. We will prove that $(+)_m$ holds. Using the existing choices of $(\widetilde{U}_J, \widetilde{E}_J^-)$ and $\widetilde{U}_{JK}$ for $J, K \in A$ with $|J|, |K| < m$ from $(+)_{m-1}$, it remains to choose $(\widetilde{U}_J, \widetilde{E}_J^-)$ when $|J| = m$, and $\widetilde{U}_{JK}$ when $0 < |K| < |J| = m$. So fix $J \subseteq I$ with $|J| = m$.

Then $(+)_{m-1}$ gives $(\widetilde{U}_K, \widetilde{E}_K^-)$ satisfying $(*)$ for all $\varnothing \neq K \subsetneq J$. Using the notation of Lemma 3.12, set $\widetilde{U}'_{JK} = \phi_{JK}^{-1}(\widetilde{U}_K) \subseteq V_J$, and define

$$\widetilde{E}'_{JK} = \chi_{JK}|_{\widetilde{U}'_{JK}}^{-1}(\widetilde{E}_K^-),$$

a vector subbundle of $E_J|_{\widetilde{U}'_{JK}}$. Then $\widetilde{U}'_{JK}$ is an open neighbourhood of $\psi_J^{-1}(C_K)$ in V_J, by (27).

If $\varnothing \neq L \subsetneq K \subsetneq J$ then by $(+)_{m-1}$ we have that there exists open $\widetilde{U}_{KL} \subseteq \widetilde{U}_K$ with $\psi_K^{-1}(C_K \cap C_L) \subseteq \widetilde{U}_{KL}$ such that

$$\phi_{KL}(\widetilde{U}_{KL}) \subseteq \widetilde{U}_L \quad \text{and} \quad \chi_{KL}|_{\widetilde{U}_{KL}}(\widetilde{E}_K^-) \subseteq \phi_{KL}|_{\widetilde{U}_{KL}}^*(\widetilde{E}_L^-) \subseteq \phi_{KL}|_{\widetilde{U}_{KL}}^*(\widetilde{E}_L).$$

Pulling back by ϕ_{JK}, applying χ_{JK}, and using the last part of Corollary 3.5(ii) then shows that we have an open neighbourhood $\widetilde{U}'_{JKL} = \phi_{JK}^{-1}(\widetilde{U}_{KL})$ of $\psi_J^{-1}(C_K \cap C_L)$ in $\widetilde{U}'_{JK} \cap \widetilde{U}'_{JL} \subseteq V_J$, such that

$$\widetilde{E}'_{JK}|_{\widetilde{U}'_{JKL}} \subseteq \widetilde{E}'_{JL}|_{\widetilde{U}'_{JKL}} \subseteq E_J|_{\widetilde{U}'_{JKL}}.$$

As in Lemma 3.12, choose vector subbundles $\widetilde{E}''_{JK} \subseteq E_J|_{\widetilde{U}'_{JK}}$ with

$$E_J|_{\widetilde{U}'_{JK}} = \widetilde{E}'_{JK} \oplus \widetilde{E}''_{JK} \quad \text{on } \widetilde{U}'_{JK} \text{ for all } \varnothing \neq K \subsetneq J.$$

Choose a connection ∇ on E_J. As in Lemma 3.12, $\widetilde{E}'''_{JK} := (\nabla s_J)[\operatorname{Ker} \mathrm{d}\phi_{JK}]$ is a vector subbundle of E_J near $s_J^{-1}(0)$ in V_J, for all $\varnothing \neq K \subsetneq J$. Making the open neighbourhoods $\widetilde{U}'_{JK}$, $\widetilde{U}'_{JKL}$ smaller, we can suppose $\widetilde{E}'''_{JK}$ is a vector subbundle of $E_J|_{\widetilde{U}'_{JK}}$. If $\varnothing \neq L \subsetneq K \subsetneq J \subseteq I$ then $\operatorname{Ker} \mathrm{d}\phi_{JK} \subseteq \operatorname{Ker} \mathrm{d}\phi_{JL}$, as $\phi_{JL} = \phi_{KL} \circ \phi_{JK}$, and so

$$\widetilde{E}'''_{JK}|_{\widetilde{U}'_{JKL}} \subseteq \widetilde{E}'''_{JL}|_{\widetilde{U}'_{JKL}} \subseteq E_J|_{\widetilde{U}'_{JKL}}.$$

Next, by reverse induction on $l = m-1, m-2, \dots, 1$, we will prove the following inductive hypothesis $(\times)_{J,l}$:

$(\times)_{J,l}$ For all $\varnothing \neq L \subsetneq J$ with $l \leqslant |L|$ we can choose an open neighbourhood $\widehat{U}_{JL}$ of $\psi_J^{-1}(C_J \cap C_L)$ in $\widetilde{U}_{JL}$ and a vector subbundle $\widehat{E}_{JL}^-$ of $E'_{JL}|_{\widehat{U}_{JL}}$ such that

$$E_J|_{\widehat{U}_{JL}} = \widehat{E}_{JL}^- \oplus E''_{JL}|_{\widehat{U}_{JL}} \oplus E'''_{JL}|_{\widehat{U}_{JL}}, \tag{70}$$

or equivalently, identifying E'_{JL} with E_J/E''_{JL} on $\widehat{U}_{JL}$,

$$E'_{JL}|_{\widehat{U}_{JL}} = \widehat{E}_{JL}^- \oplus [(E''_{JL} \oplus E'''_{JL})/E''_{JL}]|_{\widehat{U}_{JL}}, \tag{71}$$

and such that if $\varnothing \neq L \subsetneq K \subsetneq J$ with $l \leqslant |L| < |K|$ then there exists an open neighbourhood $\widehat{U}_{JKL}$ of $\psi_J^{-1}(C_J \cap C_K \cap C_L)$ in $\widehat{U}_{JK} \cap \widehat{U}_{JL}$ with $\widehat{E}_{JL}^-|_{\widehat{U}_{JKL}} = \widehat{E}_{JK}^-|_{\widehat{U}_{JKL}}$.

For the first step $l = m-1$, for each $L \subsetneq J$ with $|L| = m-1$ we take $\hat{U}_{JL} = \tilde{U}_{JL}$ and take $\hat{E}^-_{JL}$ to be an arbitrary complement to $[(E''_{JL} \oplus E'''_{JL})/E''_{JL}]$ in $E'_{JL}|_{\tilde{U}_{JL}}$, as in (71), which implies (70). The second part of $(\times)_{J,m-1}$ is trivial as there are no K, L with $m-1 \leqslant |L| < |K| < |J| = m$.

For the inductive step, suppose $(\times)_{J,l+1}$ holds for some $1 \leqslant l < m-1$, and fix $L \subsetneq J$ with $|L| = l$. Choose open neighbourhoods $\hat{U}_{JKL}$ of $\psi_J^{-1}(C_J \cap C_K \cap C_L)$ in V_J for all $L \subsetneq K \subsetneq J$ with the properties that:

(a) $\hat{U}_{JKL} \subseteq \hat{U}_{JK} \cap \tilde{U}_{JL}$, where $\hat{U}_{JK}$ is already chosen by $(\times)_{J,l+1}$.

(b) If $L \subsetneq K_1, K_2 \subsetneq J$ with $K_1 \subsetneq K_2$ and $K_2 \subsetneq K_1$ then $\hat{U}_{JK_1L} \cap \hat{U}_{JK_2L} = \varnothing$.

(c) If $L \subsetneq K_2 \subsetneq K_1 \subsetneq J$ then $\hat{U}_{JK_1L} \cap \hat{U}_{JK_2L} \subseteq \hat{U}_{JK_1K_2}$, where $\hat{U}_{JK_1K_2}$ is already chosen by $(\times)_{J,l+1}$.

This is possible, using Proposition 3.13(iii) to ensure (b).

Next, we have to choose an open neighbourhood $\hat{U}_{JL}$ of $\psi_J^{-1}(C_J \cap C_L)$ in $\tilde{U}_{JL}$ and choose a vector subbundle $\hat{E}^-_{JL}$ of $E'_{JL}|_{\hat{U}_{JL}}$ satisfying (70)–(71), such that for all K with $L \subsetneq K \subsetneq J$ we have that $\hat{U}_{JKL} \subseteq \hat{U}_{JL}$ and $\hat{E}^-_{JL}|_{\hat{U}_{JKL}} = \hat{E}^-_{JK}|_{\hat{U}_{JKL}}$.

First note from Lemma 3.12 that (70)–(71) near $\psi_J^{-1}(C_J \cap C_L)$ are equivalent to $(\hat{U}_{JL}, \hat{E}^-_{JL})$ near $\psi_J^{-1}(C_J \cap C_L)$ satisfying $(*)$ and being compatible with $(\tilde{U}_L, \tilde{E}^-_L)$. By $(\times)_{J,l+1}$ we already know that $\hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ near $\psi_J^{-1}(C_J \cap C_L)$ satisfies $(*)$ and is compatible with $(\tilde{U}_K, \tilde{E}^-_K)$, and thus $\hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ is compatible with $(\tilde{U}_L, \tilde{E}^-_L)$ near $\psi_J^{-1}(C_J \cap C_L)$ since $(\tilde{U}_K, \tilde{E}^-_K)$ is compatible with $(\tilde{U}_L, \tilde{E}^-_L)$ by $(+)_{m-1}$. Thus the prescribed value $\hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ for $\hat{E}^-_{JL}$ on $\hat{U}_{JKL}$ satisfies (70)–(71) near $\psi_J^{-1}(C_J \cap C_L)$, and making $\hat{U}_{JKL}$ smaller, we can suppose $\hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ satisfies (70)–(71) on $\hat{U}_{JKL}$. This proves that (70)–(71) are compatible with the conditions $\hat{E}^-_{JL}|_{\hat{U}_{JKL}} = \hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ for all $\varnothing \neq L \subsetneq K \subsetneq J$.

Next, observe that the prescribed values $\hat{E}^-_{JK}|_{\hat{U}_{JKL}}$ for $\hat{E}^-_{JL}$ on $\hat{U}_{JKL}$ for different K_1, K_2 with $L \subsetneq K_1$, $K_2 \subsetneq J$ agree on the overlaps $\hat{U}_{JK_1L} \cap \hat{U}_{JK_2L}$. This follows from (b) and (c) above and $\hat{E}^-_{JK_1}|_{\hat{U}_{JK_1K_2}} = \hat{E}^-_{JK_2}|_{\hat{U}_{JK_1K_2}}$, which holds by $(\times)_{J,l+1}$. Therefore the last part of $(\times)_{J,l}$ can be rewritten to say that we have one prescribed value for $\hat{E}^-_{JL}$ on the subset $\dot{U}_{JL} := \bigcup_{\{K \mid L \subsetneq K \subsetneq J\}} \hat{U}_{JKL}$, which satisfies (70)–(71) on $\dot{U}_{JL}$.

So, we are given a prescribed value of $\hat{E}^-_{JL}$ on an open set $\dot{U}_{JL} \subseteq V_J$ satisfying (71), and we have to extend it to a larger open set $\hat{U}_{JL} \subseteq V_J$ containing both $\dot{U}_{JL}$ and $\psi_J^{-1}(C_J \cap C_K \cap C_L)$. This may not be possible: if we have chosen previous values of $\hat{E}^-_{JK}$ badly near the "edge" of $\dot{U}_{JL}$ in V_J, then the prescribed values of $\hat{E}^-_{JL}$ may not extend continuously to the closure $\overline{\dot{U}}_{JL}$ of $\dot{U}_{JL}$ in V_J, and in particular, may not extend continuously over points in $[\psi_J^{-1}(C_J \cap C_K \cap C_L)] \cap [\overline{\dot{U}}_{JL} \setminus \dot{U}_{JL}]$. However, we can deal with this problem by shrinking all the open sets $\hat{U}_{JKL}$, such that the closure $\overline{\dot{U}}_{JL}$ of the new $\dot{U}_{JL}$ lies inside the old $\dot{U}_{JL}$. Then it is guaranteed that the prescribed value of $\hat{E}^-_{JL}$ on $\dot{U}_{JL}$ extends smoothly to an open neighbourhood of $\overline{\dot{U}}_{JL}$ in V_J, so we can choose $(\hat{U}_{JL}, \hat{E}^-_{JL})$ satisfying all the required conditions. As this holds for all $L \subsetneq J$ with $|L| = l$, this completes the inductive step, and $(\times)_{J,l}$ holds for all $l = m-1, m-2, \dots, 1$.

Fix data $\hat{U}_{JL}$, $\hat{E}^-_{JL}$, $\hat{U}_{JKL}$ as in $(\times)_{J,1}$. For all $\varnothing \neq K \subsetneq J$, choose open neighbourhoods $\check{U}_{JK}$ of $\psi_J^{-1}(C_J \cap C_K)$ in $\hat{U}_{JK}$ such that if $K_1 \subsetneq K_2$ and $K_2 \subsetneq K_1$ then $\check{U}_{JK_1} \cap \check{U}_{JK_2} = \varnothing$, and if $\varnothing \neq L \subsetneq K \subsetneq J$ then $\check{U}_{JK} \cap \check{U}_{JL} \subseteq \hat{U}_{JKL}$. This is possible provided the $\check{U}_{JK}$ are small enough, using Proposition 3.13(iii) to ensure $\check{U}_{JK_1} \cap \check{U}_{JK_2} = \varnothing$.

Define

$$\check{U}_J = \bigcup_{\{K \mid \varnothing \neq K \subsetneq J\}} \check{U}_{JK}.$$

The set $\check{U}_J$ is an open neighbourhood of the closed set $\check{C}_J$ in V_J, where $\check{C}_J = \bigcup_{\{K \mid \varnothing \neq K \subsetneq J\}} \psi_J^{-1}(C_J \cap C_K)$ in V_J. Define a vector subbundle $\check{E}^-_J$ of $E_J|_{\check{U}_J}$ by

$$\check{E}^-_J|_{\check{U}_{JK}} = \hat{E}^-_{JL}|_{\check{U}_{JK}} \quad \text{for all } \varnothing \neq K \subsetneq J.$$

These prescribed values for different K_1, K_2 are compatible, by construction, on the overlap $\check{U}_{JK_1} \cap \check{U}_{JK_2}$, so $\check{E}^-_J$ is well defined.

Now apply Theorem 3.7(b) to $A^\bullet_J$, V_J, E_J, $s_J, \dots$, with closed set $\check{C}_J \subseteq V_J$ and pair $(\check{U}_J, \check{E}^-_J)$ satisfying $(*)$ with $\check{C}_J \subseteq \check{U}_J$. This shows that there exists a pair $(\tilde{U}_J, \tilde{E}^-_J)$ satisfying $(*)$ for $A^\bullet_J$, V_J, E_J, $s_J, \dots$, and an open neighbourhood $\check{U}'_J$ of $\check{C}_J$ in $\check{U}_J \cap \tilde{U}_J$ such that $\check{E}^-_J|_{\check{U}'_J} = \tilde{E}^-_J|_{\check{U}'_J}$. Set

$$\tilde{U}_{JK} = \check{U}'_J \cap \check{U}_{JK} \quad \text{for all } \varnothing \neq K \subsetneq J.$$

Then $\tilde{U}_{JK}$ is an open neighbourhood of $\psi_J^{-1}(C_J \cap C_K)$ in V_J, and $\tilde{E}^-_J|_{\tilde{U}_{JK}} = \check{E}^-_J|_{\tilde{U}_{JK}} = \hat{E}^-_{JK}|_{\tilde{U}_{JK}}$, which is compatible with $(\tilde{U}_K, \tilde{E}^-_K)$ by definition. This completes the proof of the inductive step of $(+)_m$. So by induction, $(+)_m$ holds for all $m = 1, 2, \dots$.

Fix data $(\tilde{U}_J, \tilde{E}^-_J)$ for all $J \in A$ and $\tilde{U}_{JK}$ for all $J, K \in A$ with $K \subsetneq J$ as in $(+)_m$ as $m \to \infty$ (or $m = |I|$ if I is finite). For all $J \in A$, choose open neighbourhoods U_J of $\psi_J^{-1}(C_J)$ in $\tilde{U}_J$, such that setting $E^-_J = \tilde{E}^-_J|_{U_J}$ and $S_J = \psi_J(s_J^{-1}(0) \cap U_J)$, so that S_J is an open neighbourhood of C_J in X_{an}, then (U_J, E^-_J) satisfies condition (†), and for all $J, K \in A$, if $J \not\subseteq K$ and $K \not\subseteq J$ then $S_J \cap S_K = \varnothing$, and if $K \subsetneq J$ then $\psi_J^{-1}(S_J \cap S_K) \subseteq \tilde{U}_{JK}$. If $K \subsetneq J$, we define $U_{JK} = \tilde{U}_{JK} \cap U_J \cap \phi_{JK}^{-1}(U_K)$. Then $s_J^{-1}(0) \cap U_{JK} = \psi_J^{-1}(S_J \cap S_K)$, and $(U_{JK}, E^-_J|_{U_{JK}})$ is compatible with (U_K, E^-_K).

To see that we can choose U_J for all $J \in A$ satisfying all these conditions, note that by Theorem 3.7(c), if U_J is small enough then (U_J, E_J^-) satisfies (†), as $(\tilde{U}_J, \tilde{E}_J^-)$ satisfies $(*)$. If $J \not\subseteq K$ and $K \not\subseteq J$ then Proposition 3.13(iii) implies that $S_J \cap S_K = \varnothing$ provided both U_J, U_K are sufficiently small. Similarly, if $K \subsetneq J$ then we have $\psi_J^{-1}(S_J \cap S_K) \subseteq \tilde{U}_{JK}$ provided both U_J, U_K are sufficiently small. Now if I is infinite, it is possible that an individual set U_J may have to satisfy infinitely many smallness conditions, for compatibility with infinitely many sets $\varnothing \neq K \subseteq I$. However, the local finiteness condition Proposition 3.13(ii) means that in an open neighbourhood of any $v_J \in \psi_J^{-1}(C_J)$, only finitely many smallness conditions on U_J are relevant, so we can solve them. This completes the proof of Proposition 3.14.

6.3 Proof of Proposition 3.17

Let $(\boldsymbol{X}, \omega_{\boldsymbol{X}^*})$, X_{an}, $\mathcal{K}$ and $\boldsymbol{X}_{\mathrm{dm}}$ be as in Theorems 3.15 and 3.16, and use the notation of Section 3.5. First we relate orientations on $(\boldsymbol{X}, \omega_{\boldsymbol{X}^*})$ and $\boldsymbol{X}_{\mathrm{dm}}$ at one point $x \in X_{\mathrm{an}}$. Pick $J \in A$ with $x \in S_J = \operatorname{Im} \psi_J^+$. From (7) and (9) we have

$$\{\text{orientations on } (\boldsymbol{X}, \omega_{\boldsymbol{X}}^*) \text{ at } x\} \cong \{\mathbb{C}\text{–orientations on } (H^1(\mathbb{T}_{\boldsymbol{X}}|_x), Q_x)\}, \tag{72}$$

$$\{\text{orientations on } \boldsymbol{X}_{\mathrm{dm}} \text{ at } x\} \cong \{\text{orientations on } T_x^* \boldsymbol{X}_{\mathrm{dm}} \oplus O_x \boldsymbol{X}_{\mathrm{dm}}\}, \tag{73}$$

where $Q_x = \omega_{\boldsymbol{X}}^0 \cdot$ is the nondegenerate complex quadratic form on $H^1(\mathbb{T}_{\boldsymbol{X}}|_x)$ in (6). There is a unique v_J in $s_J^{-1}(0) \cap U_J = (s_J^+)^{-1}(0) \subseteq U_J \subseteq V_J$ with $\psi_J(v_J) = x$. Equation (20) gives an isomorphism of complex vector spaces

$$H^1(\mathbb{T}_{\boldsymbol{\alpha}_J}|_{v_J})\colon \frac{\operatorname{Ker}(t_J|_{v_J}\colon E_J|_{v_J} \to F_J|_{v_J})}{\operatorname{Im}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}V_J \to E_J|_{v_J})} \to H^1(\mathbb{T}_{\boldsymbol{X}}|_x). \tag{74}$$

Write $\tilde{Q}_{v_J}$ for the complex quadratic form on $\operatorname{Ker}(t_J|_{v_J})/\operatorname{Im}(\mathrm{d}s_J|_{v_J})$ identified with Q_x by (74), as in Definition 3.6. Then by (72) we have

$$\begin{aligned}\{\text{orientations on } (\boldsymbol{X}, \omega_{\boldsymbol{X}}^*) \text{ at } x\}& \\ &\cong \big\{\mathbb{C}\text{–orientations on } \big(\operatorname{Ker}(t_J|_{v_J})/\operatorname{Im}(\mathrm{d}s_J|_{v_J}), \tilde{Q}_{v_J}\big)\big\}.\end{aligned} \tag{75}$$

Condition $(*)$ for (U_J, E_J^-) at v_J requires that

$$\Pi_{v_J}\colon E_J^-|_{v_J} \cap \operatorname{Ker}(t_J|_{v_J}\colon E_J|_{v_J} \to F_J|_{v_J}) \to \frac{\operatorname{Ker}(t_J|_{v_J}\colon E_J|_{v_J} \to F_J|_{v_J})}{\operatorname{Im}(\mathrm{d}s_J|_{v_J}\colon T_{v_J}V_J \to E_J|_{v_J})}$$

should be injective, with image $\operatorname{Im}\Pi_{v_J}$ a real vector subspace of half the real dimension of $\operatorname{Ker}(t_J|_{v_J})/\operatorname{Im}(\mathrm{d}s_J|_{v_J})$, on which the real quadratic form $\operatorname{Re}\tilde{Q}_{v_J}$ is negative definite. As $(U_J, E_J^+, s_J^+, \psi_J|_{s_J^{-1}(0)\cap U_J})$ is a Kuranishi neighbourhood on $\boldsymbol{X}_{\mathrm{dm}}$ by the proof of Theorem 3.16, equation (10) gives an exact sequence

$$0 \longrightarrow T_x\boldsymbol{X}_{\mathrm{dm}} \longrightarrow T_{v_J}V_J \xrightarrow{\mathrm{d}s_J^+|_{v_J}} E_J^+|_{v_J} \longrightarrow O_x\boldsymbol{X}_{\mathrm{dm}} \longrightarrow 0.$$

Condition $(*)$ implies that $\mathrm{Ker}(\mathrm{d}s_J|_{v_J}) = \mathrm{Ker}(\mathrm{d}s_J^+|_{v_J})$, so we have

$$T_x \boldsymbol{X}_{\mathrm{dm}} \cong \mathrm{Ker}(\mathrm{d}s_J|_{v_J} \colon T_{v_J} V_J \to E_J|_{v_J}). \tag{76}$$

Also from $(*)$ we see there is a canonical isomorphism

$$O_x \boldsymbol{X}_{\mathrm{dm}} \cong \frac{\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})}{\mathrm{Im}\, \Pi_{v_J}}. \tag{77}$$

By (76), $T_x \boldsymbol{X}_{\mathrm{dm}}$ is a complex vector space, so $T_x \boldsymbol{X}_{\mathrm{dm}}$ and $T_x^* \boldsymbol{X}_{\mathrm{dm}}$ have natural orientations as real vector spaces. Thus by (77) we have a bijection

$$\begin{aligned} &\{\text{orientations on } T_x^* \boldsymbol{X}_{\mathrm{dm}} \oplus O_x \boldsymbol{X}_{\mathrm{dm}}\} \\ &\qquad \cong \{\text{orientations on } [\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})] / \mathrm{Im}\, \Pi_{v_J}\}. \end{aligned} \tag{78}$$

Suppose we are given a complex basis $e_1, \dots, e_k$ of $\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J}) \cong \mathbb{C}^k$ that is orthonormal with respect to $\widetilde{Q}_{v_J}$. As $e_1, \dots, e_k$ are orthonormal with respect to $\widetilde{Q}_{v_J}$, the real quadratic form $\mathrm{Re}\, \widetilde{Q}_{v_J}$ is positive definite on the real span $\langle e_1, \dots, e_k \rangle_{\mathbb{R}}$, and $\mathrm{Re}\, \widetilde{Q}_{v_J}$ is negative definite on $\mathrm{Im}\, \Pi_{v_J}$, and thus $\langle e_1, \dots, e_k \rangle_{\mathbb{R}} \cap \mathrm{Im}\, \Pi_{v_J} = \{0\}$. Therefore $e_1 + \mathrm{Im}\, \Pi_{v_J}, \dots, e_k + \mathrm{Im}\, \Pi_{v_J}$ are linearly independent in the real vector space $[\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})] / \mathrm{Im}\, \Pi_{v_J} \cong \mathbb{R}^k$, so they are a basis as $\mathrm{Im}\, \Pi_{v_J}$ has half the real dimension of $\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})$. Define an identification

$$\begin{aligned} &\{\mathbb{C}\text{–orientations on } \big(\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J}), \widetilde{Q}_{v_J}\big)\} \\ &\qquad \cong \{\text{orientations on } [\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})] / \mathrm{Im}\, \Pi_{v_J}\}, \end{aligned} \tag{79}$$

such that orientations on both sides are identified if, whenever $e_1, \dots, e_k$ is an oriented orthonormal complex basis for $\big(\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J}), \widetilde{Q}_{v_J}\big)$, then we have that $e_1 + \mathrm{Im}\, \Pi_{v_J}, \dots, e_k + \mathrm{Im}\, \Pi_{v_J}$ is an oriented basis for $[\mathrm{Ker}(t_J|_{v_J}) / \mathrm{Im}(\mathrm{d}s_J|_{v_J})] / \mathrm{Im}\, \Pi_{v_J}$. Combining equations (73), (75), (78) and (79) gives an identification

$$\{\text{orientations on } (\boldsymbol{X}, \omega_{\boldsymbol{X}}^*) \text{ at } x\} \cong \{\text{orientations on } \boldsymbol{X}_{\mathrm{dm}} \text{ at } x\}. \tag{80}$$

It is not difficult to show that the isomorphism (80) is independent of the choice of $J \in A$ with $x \in S_J$, and depends continuously on $x \in X_{\mathrm{an}}$. Thus we get a canonical one-to-one correspondence between the sets in Proposition 3.17(a),(c). The last part of Theorem 2.18 gives a one-to-one correspondence between the sets in Proposition 3.17(b),(c). This completes the proof.

6.4 Proof of Proposition 3.18

Suppose $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$ is a separated, -2–shifted symplectic derived $\mathbb{C}$–scheme with virtual dimension $\mathrm{vdim}_{\mathbb{C}} \boldsymbol{X} = n$, whose complex analytic topological space X_{an} is second countable. Let $\mathcal{K}$, $\mathcal{K}'$ be different possible Kuranishi atlases constructed in Theorem 3.15, and $\boldsymbol{X}_{\mathrm{dm}}$, $\boldsymbol{X}'_{\mathrm{dm}}$ the corresponding derived manifolds in Theorem 3.16.

As in Section 3.5, let $\mathcal{K}$ be constructed using the family $\{(A_i^\bullet, \boldsymbol{\alpha}_i) \mid i \in I\}$, and data $A_J^\bullet$, $\boldsymbol{\alpha}_J$ for $J \in A$, Φ_{JK} for $K \subseteq J$ in A from Theorem 3.1, where $A = \{J \mid \varnothing \neq J \subseteq I \text{ and } J \text{ is finite}\}$, and as in Section 3.2, use notation V_J, E_J, F_J, s_J, t_J, ψ_J and $R_J = \bigcap_{i \in J} R_i \subseteq X_{\mathrm{an}}$ from $A_J^\bullet$, $\boldsymbol{\alpha}_J$ and ϕ_{JK}, χ_{JK}, ξ_{JK} from Φ_{JK}. Let $\mathcal{K}$ be defined using closed subsets $C_J \subseteq X_{\mathrm{an}}$ for $J \in A$ in Proposition 3.13 and pairs (U_J, E_J^-) and open subsets $U_{JK} \subseteq U_J$ in Proposition 3.14. Similarly, let $\mathcal{K}'$ be constructed using $\{(A_{i'}'^\bullet, \boldsymbol{\alpha}'_{i'}) \mid i' \in I'\}$, $A_{J'}'^\bullet$, $\boldsymbol{\alpha}'_{J'}$, $V'_{J'}$, $E'_{J'}, \dots, U'_{J'K'} \subseteq U'_{J'}$.

We must build a derived manifold with boundary $\boldsymbol{W}_{\mathrm{dm}}$ with topological space $X_{\mathrm{an}} \times [0,1]$ and $\operatorname{vdim} \boldsymbol{W}_{\mathrm{dm}} = n+1$, and an equivalence $\partial \boldsymbol{W}_{\mathrm{dm}} \simeq \boldsymbol{X}_{\mathrm{dm}} \amalg \boldsymbol{X}'_{\mathrm{dm}}$ topologically identifying $\boldsymbol{X}_{\mathrm{dm}}$ with $X_{\mathrm{an}} \times \{0\}$ and $\boldsymbol{X}'_{\mathrm{dm}}$ with $X_{\mathrm{an}} \times \{1\}$.

Write $\tilde{\boldsymbol{\pi}}\colon \tilde{\boldsymbol{X}} \to Z$ to be the projection $\boldsymbol{\pi}_{\mathbb{A}^1}\colon \boldsymbol{X} \times \mathbb{A}^1 \to \mathbb{A}^1$, so that $Z = \mathbb{A}^1 = \operatorname{Spec} B$ with $B = \mathbb{C}[z]$, and $Z_{\mathrm{an}} = \mathbb{C}$. Define $\omega_{\tilde{\boldsymbol{X}}/Z} = \boldsymbol{\pi}_{\boldsymbol{X}}^*(\omega_{\boldsymbol{X}}^0)$. Then $\omega_{\tilde{\boldsymbol{X}}/Z}$ is a family of -2–shifted symplectic structures on $\boldsymbol{X}/Z$ in the sense of Section 3.7, the constant family over $Z = \mathbb{A}^1$ with fibre $(\boldsymbol{X}, \omega_{\boldsymbol{X}}^*)$. We now carry out the programme of Section 3.7 for $\tilde{\boldsymbol{\pi}}\colon \tilde{\boldsymbol{X}} \to Z, \omega_{\tilde{\boldsymbol{X}}/Z}$, choosing data as follows:

(a) Set $\tilde{I} = I \amalg I'$, the disjoint union of I and I'.

(b) Define $(\tilde{A}_i^\bullet, \tilde{\boldsymbol{\alpha}}_i, \tilde{\beta}_i)$ for $i \in I$ by

$$\tilde{A}_i^\bullet = A_i^\bullet \otimes_{\mathbb{C}} \mathbb{C}[z, (z-1)^{-1}],$$

so that $\operatorname{\mathbf{Spec}} \tilde{A}_i^\bullet = (\operatorname{\mathbf{Spec}} A_i^\bullet) \times (\mathbb{A}^1 \setminus \{1\})$, and

$$\tilde{\boldsymbol{\alpha}}_i = \boldsymbol{\alpha}_i \times \mathrm{inc}\colon (\operatorname{\mathbf{Spec}} A_i^\bullet) \times (\mathbb{A}^1 \setminus \{1\}) \to \boldsymbol{X} \times \mathbb{A}^1,$$

and

$$\tilde{\beta}_i\colon \mathbb{C}[z] \to A_i^0 \otimes_{\mathbb{C}} \mathbb{C}[z, (z-1)^{-1}] \text{ by } \tilde{\beta}_i\colon z \mapsto 1 \otimes z.$$

Similarly, define $(\tilde{A}_{i'}'^\bullet, \tilde{\boldsymbol{\alpha}}_{i'}, \tilde{\beta}_{i'})$ for $i' \in I'$ by $\tilde{A}_{i'}^\bullet = A_{i'}'^\bullet \otimes_{\mathbb{C}} \mathbb{C}[z, z^{-1}]$, so $\operatorname{\mathbf{Spec}} \tilde{A}_{i'}'^\bullet = (\operatorname{\mathbf{Spec}} A_{i'}'^\bullet) \times (\mathbb{A}^1 \setminus \{0\})$, and $\tilde{\boldsymbol{\alpha}}_{i'} = \boldsymbol{\alpha}'_{i'} \times \mathrm{inc}\colon (\operatorname{\mathbf{Spec}} A_{i'}'^\bullet) \times (\mathbb{A}^1 \setminus \{0\}) \to \boldsymbol{X} \times \mathbb{A}^1$, and $\tilde{\beta}'_{i'}\colon \mathbb{C}[z] \to A_{i'}'^0 \otimes_{\mathbb{C}} \mathbb{C}[z, z^{-1}]$ by $\tilde{\beta}'_{i'}\colon z \mapsto 1 \otimes z$.

(c) Write $\tilde{A} = \{\tilde{J} \mid \varnothing \neq \tilde{J} \subseteq \tilde{I} \text{ and } \tilde{J} \text{ is finite}\}$. Then $A \subseteq \tilde{A}$ and $A' \subseteq \tilde{A}$.

(d) When we apply Theorem 3.1 to choose $\tilde{A}_{\tilde{J}}^\bullet$, $\tilde{\boldsymbol{\alpha}}_{\tilde{J}}$, $\tilde{\beta}_{\tilde{J}}$ for $\tilde{J} \in \tilde{A}$ and $\tilde{\Phi}_{\tilde{J}\tilde{K}}$ for $\tilde{K} \subseteq \tilde{J}$, we make these choices so that

$$\tilde{A}_J^\bullet = A_J^\bullet \otimes_{\mathbb{C}} \mathbb{C}[z, (z-1)^{-1}] \quad \text{and} \quad \tilde{A}_{J'}^\bullet = A_{J'}'^\bullet \otimes_{\mathbb{C}} \mathbb{C}[z, z^{-1}],$$

$$\tilde{\boldsymbol{\alpha}}_J = \boldsymbol{\alpha}_J \times \mathrm{inc}\colon (\operatorname{\mathbf{Spec}} A_J^\bullet) \times (\mathbb{A}^1 \setminus \{1\}) \to \boldsymbol{X} \times \mathbb{A}^1,$$

$$\tilde{\boldsymbol{\alpha}}_{J'} = \boldsymbol{\alpha}'_{J'} \times \mathrm{inc}\colon (\mathbf{Spec}\, A'^{\bullet}_{J'}) \times (\mathbb{A}^1 \setminus \{0\}) \to \boldsymbol{X} \times \mathbb{A}^1,$$
$$\tilde{\beta}_J\colon z \mapsto 1 \otimes z \quad \text{and} \quad \tilde{\beta}_{J'}\colon z \mapsto 1 \otimes z,$$
$$\tilde{\Phi}_{JK} = \Phi_{JK} \otimes \mathrm{id}\colon A^{\bullet}_K \otimes_{\mathbb{C}} \mathbb{C}[z,(z-1)^{-1}] \to A^{\bullet}_J \otimes_{\mathbb{C}} \mathbb{C}[z,(z-1)^{-1}],$$
$$\tilde{\Phi}_{J'K'} = \Phi'_{J'K'} \otimes \mathrm{id}\colon A'^{\bullet}_{K'} \otimes_{\mathbb{C}} \mathbb{C}[z,z^{-1}] \to A'^{\bullet}_{J'} \otimes_{\mathbb{C}} \mathbb{C}[z,z^{-1}],$$

for all $K \subseteq J$ in A and $K' \subseteq J'$ in A'. This is clearly possible. Note that this does not determine $\tilde{A}^{\bullet}_{\tilde{J}}$, $\tilde{\boldsymbol{\alpha}}_{\tilde{J}}$, $\tilde{\beta}_{\tilde{J}}$ or $\Phi_{\tilde{J}\tilde{K}}$ if $\tilde{J} \in \tilde{A} \setminus (A \amalg A')$.

(e) When we translate to complex geometry using Section 3.2, part (d) implies that $\tilde{V}_J = V_J \times (\mathbb{C} \setminus \{1\})$ for $J \in A \subseteq \tilde{A}$. Also $\tilde{E}_J$, $\tilde{F}_J$, $\tilde{s}_J$, $\tilde{t}_J$, $\tilde{\phi}_{JK}$, $\tilde{\chi}_{JK}$ for $J, K \in A$ are obtained from $E_J, \dots, \chi_{JK}$ by taking products with $\mathbb{C} \setminus \{1\}$. Similarly, $\tilde{V}_{J'}$, $\tilde{E}_{J'}$, $\tilde{F}_{J'}$, $\tilde{s}_{J'}$, $\tilde{t}_{J'}$, $\tilde{\phi}_{J'K'}$, $\tilde{\chi}_{J'K'}$ for $J', K' \in A' \subseteq \tilde{A}$ are obtained from $V_{J'}, \dots, \chi_{J'K'}$ by taking products with $\mathbb{C} \setminus \{0\}$.

(f) When we choose data $\tilde{C}_{\tilde{J}}$, $(\tilde{U}_{\tilde{J}}, \tilde{E}^-_{\tilde{J}})$ for $\tilde{J} \in \tilde{A}$, we do this so that

$$\tilde{C}_J \cap (X_{\mathrm{an}} \times \{0\}) = C_J \times \{0\}, \qquad \tilde{U}_J \cap V_J \times \{0\} = U_J \times \{0\},$$
$$\tilde{E}^-_J|_{U_J \times \{0\}} = E^-_J \times 0, \qquad \tilde{C}_{J'} \cap (X_{\mathrm{an}} \times \{1\}) = C'_{J'} \times \{1\},$$
$$\tilde{U}_{J'} \cap V'_{J'} \times \{1\} = U'_{J'} \times \{1\}, \qquad \tilde{E}^-_{J'}|_{U'_{J'} \times \{1\}} = E'^-_{J'} \times 1,$$

whenever $J \in A$ and $J' \in A'$. This is clearly possible.

Theorem 3.23 constructs a relative Kuranishi atlas $\tilde{\mathcal{K}}$ for $\pi_{\mathbb{C}}\colon X_{\mathrm{an}} \times \mathbb{C} \to \mathbb{C}$, of dimension $n+2$. By construction, over $X_{\mathrm{an}} \times \{0\}$ this restricts to the Kuranishi atlas $\mathcal{K}$, and over $X_{\mathrm{an}} \times \{1\}$ it restricts to $\mathcal{K}'$.

Theorem 3.24 gives a derived manifold $\tilde{\boldsymbol{X}}_{\mathrm{dm}}$ with $\mathrm{vdim}\, \tilde{\boldsymbol{X}}_{\mathrm{dm}} = n+2$ and topological space $X_{\mathrm{an}} \times \mathbb{C}$, with a morphism $\tilde{\boldsymbol{\pi}}_{\mathrm{dm}}\colon \tilde{\boldsymbol{X}}_{\mathrm{dm}} \to \mathbb{C}$. From Theorem 3.24(iii) we see that $\tilde{\boldsymbol{X}}^{\boldsymbol{0}}_{\mathrm{dm}} = \tilde{\boldsymbol{\pi}}^{-1}_{\mathrm{dm}}(0) \simeq \boldsymbol{X}_{\mathrm{dm}}$ and $\tilde{\boldsymbol{X}}^1_{\mathrm{dm}} = \tilde{\boldsymbol{\pi}}^{-1}_{\mathrm{dm}}(1) \simeq \boldsymbol{X}'_{\mathrm{dm}}$.

Now define $\boldsymbol{W}_{\mathrm{dm}} = \tilde{\boldsymbol{X}}_{\mathrm{dm}} \times_{\tilde{\boldsymbol{\pi}}_{\mathrm{dm}},\mathbb{C},\mathrm{inc}} [0,1]$, as a fibre product in the 2–category $\mathbf{dMan^c}$ of d-manifolds with corners from [18; 19; 20], where $\mathrm{inc}\colon [0,1] \hookrightarrow \mathbb{C}$ is the inclusion. By properties of fibre products in $\mathbf{dMan^c}$ from [18; 19; 20], this has topological space $X_{\mathrm{an}} \times [0,1]$ and $\mathrm{vdim}\, \boldsymbol{W}_{\mathrm{dm}} = n+1$, and boundary

$$\partial \boldsymbol{W}_{\mathrm{dm}} \simeq \tilde{\boldsymbol{X}}_{\mathrm{dm}} \times_{\tilde{\boldsymbol{\pi}}_{\mathrm{dm}},\mathbb{C},\mathrm{inc}} \partial[0,1] \simeq \tilde{\boldsymbol{X}}_{\mathrm{dm}} \times_{\tilde{\boldsymbol{\pi}}_{\mathrm{dm}},\mathbb{C},\mathrm{inc}} \{0,1\} \simeq \boldsymbol{X}_{\mathrm{dm}} \amalg \boldsymbol{X}'_{\mathrm{dm}}. \tag{81}$$

This proves the first part of Proposition 3.18.

For the last part, orientations on $(\boldsymbol{X}, \omega^*_{\boldsymbol{X}})$ correspond naturally to orientations for $\tilde{\boldsymbol{\pi}}\colon \tilde{\boldsymbol{X}} \to Z, \omega_{\tilde{\boldsymbol{X}}/Z}$, by pullback along $\tilde{\boldsymbol{X}} \to \boldsymbol{X}$, and these correspond to orientations on $\tilde{\boldsymbol{X}}_{\mathrm{dm}}$ by Proposition 3.26, and thus (using oriented fibre products) to orientations on $\boldsymbol{W}_{\mathrm{dm}}$. Since $\partial[0,1] = -\{0\} \amalg \{1\}$ in oriented manifolds, we see that as in (81) that $\partial \boldsymbol{W}_{\mathrm{dm}} \simeq -\boldsymbol{X}_{\mathrm{dm}} \amalg \boldsymbol{X}'_{\mathrm{dm}}$ in oriented derived manifolds. This completes the proof.

References

[1] **K Behrend**, **B Fantechi**, *The intrinsic normal cone*, Invent. Math. 128 (1997) 45–88 MR

[2] **O Ben-Bassat**, **C Brav**, **V Bussi**, **D Joyce**, *A "Darboux theorem" for shifted symplectic structures on derived Artin stacks, with applications*, Geom. Topol. 19 (2015) 1287–1359 MR

[3] **D Borisov**, *Derived manifolds and Kuranishi models*, preprint (2012) arXiv

[4] **D Borisov**, **J Noel**, *Simplicial approach to derived differential manifolds*, preprint (2011) arXiv

[5] **C Brav**, **V Bussi**, **D Dupont**, **D Joyce**, **B Szendrői**, *Symmetries and stabilization for sheaves of vanishing cycles*, J. Singul. 11 (2015) 85–151 MR

[6] **C Brav**, **V Bussi**, **D Joyce**, *A "Darboux theorem" for derived schemes with shifted symplectic structure* (2013) arXiv

[7] **V Bussi**, **D Joyce**, **S Meinhardt**, *On motivic vanishing cycles of critical loci*, preprint (2013) arXiv

[8] **Y Cao**, *Donaldson–Thomas theory for Calabi–Yau four-folds*, preprint (2013) arXiv

[9] **Y Cao**, **N C Leung**, *Donaldson–Thomas theory for Calabi–Yau 4–folds*, preprint (2014) arXiv

[10] **Y Cao**, **N C Leung**, *Orientability for gauge theories on Calabi–Yau manifolds*, Adv. Math. 314 (2017) 48–70 MR

[11] **P E Conner**, *Differentiable periodic maps*, 2nd edition, Lecture Notes in Math. 738, Springer (1979) MR

[12] **S K Donaldson**, **P B Kronheimer**, *The geometry of four-manifolds*, Oxford Univ. Press (1990) MR

[13] **S K Donaldson**, **R P Thomas**, *Gauge theory in higher dimensions*, from "The geometric universe" (S A Huggett, L J Mason, K P Tod, S T Tsou, N M J Woodhouse, editors), Oxford Univ. Press (1998) 31–47 MR

[14] **K Fukaya**, **Y-G Oh**, **H Ohta**, **K Ono**, *Lagrangian intersection Floer theory: anomaly and obstruction, I–II*, AMS/IP Studies in Advanced Mathematics 46, Amer. Math. Soc., Providence, RI (2009) MR

[15] **K Fukaya**, **K Ono**, *Arnold conjecture and Gromov–Witten invariant*, Topology 38 (1999) 933–1048 MR

[16] **R Hartshorne**, *Algebraic geometry*, Graduate Texts in Math. 52, Springer (1977) MR

[17] **D Joyce**, *Algebraic geometry over C^∞–rings*, preprint (2009) arXiv To appear in Mem. Amer. Math. Soc.

[18] **D Joyce**, *D-manifolds and d-orbifolds: a theory of derived differential geometry*, book project (2012) Available at http://people.maths.ox.ac.uk/~joyce/dmanifolds.html

[19] **D Joyce**, *D-manifolds, d-orbifolds and derived differential geometry: a detailed summary*, preprint (2012) arXiv

[20] **D Joyce**, *An introduction to d-manifolds and derived differential geometry*, from "Moduli spaces" (L Brambila-Paz, O García-Prada, P Newstead, R P Thomas, editors), London Math. Soc. Lecture Note Ser. 411, Cambridge Univ. Press (2014) 230–281 MR

[21] **D Joyce**, *A new definition of Kuranishi space*, preprint (2014) arXiv

[22] **D Joyce**, *A classical model for derived critical loci*, J. Differential Geom. 101 (2015) 289–367 MR

[23] **D Joyce**, *Kuranishi spaces as a 2–category*, preprint (2015) arXiv

[24] **D Joyce**, *Kuranishi spaces and symplectic geometry*, multiple-volume book project (2017) Available at http://people.maths.ox.ac.uk/~joyce/Kuranishi.html

[25] **D Joyce**, **Y Song**, *A theory of generalized Donaldson–Thomas invariants*, Mem. Amer. Math. Soc. 1020, Amer. Math. Soc., Providence, RI (2012) MR

[26] **M Kontsevich**, **Y Soibelman**, *Stability structures, motivic Donaldson–Thomas invariants and cluster transformations* (2008) arXiv

[27] **J Lurie**, *Derived algebraic geometry, V: Structured spaces*, preprint (2009) arXiv

[28] **D McDuff**, *Notes on Kuranishi atlases*, preprint (2014) arXiv

[29] **D McDuff**, **K Wehrheim**, *Kuranishi atlases with trivial isotropy: the* 2013 *state of affairs*, preprint (2012) arXiv

[30] **J Milnor**, *On the Steenrod homology theory*, from "Collected papers of John Milnor, IV: Homotopy, homology and manifolds" (J McCleary, editor), Amer. Math. Soc., Providence, RI (2009) 83–100 MR Reprint of the 1961 original

[31] **T Pantev**, **B Toën**, **M Vaquié**, **G Vezzosi**, *Shifted symplectic structures*, Publ. Math. Inst. Hautes Études Sci. 117 (2013) 271–328 MR

[32] **D I Spivak**, *Derived smooth manifolds*, Duke Math. J. 153 (2010) 55–128 MR

[33] **R P Thomas**, *A holomorphic Casson invariant for Calabi–Yau* 3*–folds, and bundles on* $K3$ *fibrations*, J. Differential Geom. 54 (2000) 367–438 MR

[34] **B Toën**, *Higher and derived stacks: a global overview*, from "Algebraic geometry, 1" (D Abramovich, A Bertram, L Katzarkov, R Pandharipande, M Thaddeus, editors), Proc. Sympos. Pure Math. 80, Amer. Math. Soc., Providence, RI (2009) 435–487 MR

[35] **B Toën**, *Derived algebraic geometry*, EMS Surv. Math. Sci. 1 (2014) 153–240 MR

[36] **B Toën**, **G Vezzosi**, *Homotopical algebraic geometry, II: Geometric stacks and applications*, 902, Amer. Math. Soc., Providence, RI (2008) MR

Mathematisches Institut, Georg-August-Universität Göttingen
Göttingen, Germany

The Mathematical Institute, University of Oxford
Oxford, United Kingdom

dennis.borisov@gmail.com, joyce@maths.ox.ac.uk

https://sites.google.com/site/dennisborisov/,
http://people.maths.ox.ac.uk/~joyce/

Proposed: Simon Donaldson
Seconded: Richard Thomas, Jim Bryan

Received: 7 April 2015
Revised: 3 October 2016

Geometry & Topology 21 (2017) 3313–3389

Koszul duality patterns in Floer theory

TOLGA ETGÜ
YANKI LEKILI

We study symplectic invariants of the open symplectic manifolds X_Γ obtained by plumbing cotangent bundles of 2–spheres according to a plumbing tree Γ. For any tree Γ, we calculate (DG) algebra models of the Fukaya category $\mathcal{F}(X_\Gamma)$ of closed exact Lagrangians in X_Γ and the wrapped Fukaya category $\mathcal{W}(X_\Gamma)$. When Γ is a Dynkin tree of type A_n or D_n (and conjecturally also for E_6, E_7, E_8), we prove that these models for the Fukaya category $\mathcal{F}(X_\Gamma)$ and $\mathcal{W}(X_\Gamma)$ are related by (derived) Koszul duality. As an application, we give explicit computations of symplectic cohomology of X_Γ for $\Gamma = A_n, D_n$, based on the Legendrian surgery formula of Bourgeois, Ekholm and Eliashberg.

57R58; 16E45

1 Introduction

Let us begin by recalling a simple example that we learned from Blumberg, Cohen and Teleman [14]. Consider a *simply connected* smooth compact manifold S and its cotangent bundle $M = T^*S$ with its canonical symplectic structure. The zero section S is a Lagrangian submanifold. We choose a basepoint $x \in S$ and consider the corresponding cotangent fiber $L = T^*_x S$. This is another Lagrangian submanifold, a noncompact one. Throughout, our Lagrangian submanifolds will be equipped with a *brane structure*. This means that they will be given an orientation, a spin structure (in particular, we assume here that S is spinnable) and they will be equipped with a grading in the sense of Seidel [59].

Fix a coefficient field $\mathbb{K}$. Lagrangian Floer theory gives us $\mathbb{Z}$–graded A_∞–algebras over $\mathbb{K}$

$$\mathscr{A} = \mathrm{CF}^*(S,S), \quad \mathscr{B} = \mathrm{CW}^*(L,L).$$

Indeed, S is an object of $\mathcal{F}(M)$, the Fukaya category of closed exact Lagrangian branes in the Liouville manifold M (see Seidel [61]). The endomorphisms of the object S in this category are given by the Fukaya–Floer A_∞–algebra $\mathrm{CF}^*(S,S)$. On the other hand, L is an object of $\mathcal{W}(M)$, the wrapped Fukaya category of M

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3313

(see Abouzaid and Seidel [6]). The endomorphisms of the object L in this category are given by the wrapped Floer cochain complex $\mathrm{CW}^*(L,L)$, which again has an associated A_∞–structure (well-defined up to quasi-isomorphism).

Now, in this setting, by construction, there exists a full and faithful embedding

$$\mathcal{F}(M) \to \mathcal{W}(M)$$

since by definition $\mathcal{W}(M)$ allows certain noncompact Lagrangians in M with controlled behavior at infinity, in addition to the exact compact Lagrangians in M. Furthermore, it is a general fact (see Abouzaid [2]) that a cotangent fiber generates the wrapped Fukaya category in the derived sense. Hence, in particular, one has a Yoneda functor to the DG-category of A_∞–modules over $\mathscr{B}$,

$$\mathcal{Y}\colon \mathcal{F}(M) \to \mathscr{B}^{\mathrm{mod}},$$

which is a cohomologically full and faithful embedding. This sends an exact compact Lagrangian T to the (right) A_∞–module $\mathcal{Y}_T = \mathrm{CW}^*(L,T)$ over $\mathscr{B}$. As a consequence, one can compute $\mathscr{A}$ via its quasi-isomorphic image under $\mathcal{Y}$:

$$\mathscr{A} \simeq \hom_{\mathscr{B}}(\mathbb{K},\mathbb{K}), \tag{1}$$

where we write $\mathbb{K}$ for the right A_∞–module over $\mathscr{B}$ with underlying vector space $\mathbb{K}\cdot x = \mathrm{CW}^*(L,S)$. Equipping S and L with suitable brane structures, one can arrange that the degree $|x|$ is 0. The only nontrivial module map is the multiplication by the idempotent element in $\mathrm{CW}^0(L,L) = \mathbb{K}\cdot e$, which acts as the identity. The other products (including the higher products) are necessarily trivial. This can be seen from the fact that $\mathrm{CW}^*(L,L)$ is supported in nonpositive degrees (as we shall see below). Note that we are following the conventions of [61], where, for example, the A_∞–module maps are given by Floer products

$$\mathrm{CW}^*(L,S)\otimes \mathrm{CW}^*(L,L)^{\otimes k} \to \mathrm{CW}^*(L,S)[1-k], \quad k=0,\dots.$$

Throughout, upwards shift of grading by n is written as $[-n]$.

On the other hand, $\mathrm{CW}^*(L,S)$ is also a (left) A_∞–module over $\mathrm{CF}^*(S,S)$, where A_∞–module maps are given by Floer products

$$\mathrm{CF}^*(S,S)^{\otimes k}\otimes \mathrm{CW}^*(L,S) \to \mathrm{CW}^*(L,S)[1-k], \quad k=0,\dots.$$

To be in line with the conventions of [61], we prefer to view this as a right $\mathscr{A}^{\mathrm{op}}$–module (which entails slightly different sign conventions). In fact, in our setting, it turns out that $\mathscr{A}$ is quasi-isomorphic to $\mathscr{A}^{\mathrm{op}}$.

Somewhat more surprisingly, one can also compute $\mathscr{B}$ as

$$\mathscr{B}^{\mathrm{op}} \simeq \hom_{\mathscr{A}^{\mathrm{op}}}(\mathbb{K}, \mathbb{K}). \tag{2}$$

This is an instance of *Koszul duality*.

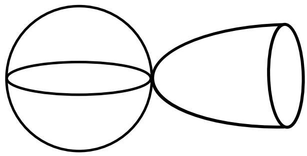


Figure 1: A picture of Koszul duality

To see this, we observe that both $\mathscr{A}$ and $\mathscr{B}$ have topological models due to Abouzaid [3; 4]. Indeed, there are A_∞–equivalences

$$\mathscr{A} \simeq C^*(S;\mathbb{K}) \quad \text{and} \quad \mathscr{B} \simeq C_{-*}(\Omega_x S;\mathbb{K}),$$

where $\Omega_x S$ is the based loop space of S at x. Notice the cohomological grading in place. In particular, $\mathscr{A}$ is supported in nonnegative degrees and $\mathscr{B}$ is supported in nonpositive degrees.

Now, (1) becomes an Eilenberg–Moore equivalence (of DGA's)

$$\mathrm{RHom}_{C_{-*}(\Omega_x S)}(\mathbb{K}, \mathbb{K}) \simeq C^*(S;\mathbb{K}),$$

and (2) is Adams' cobar equivalence (see Adams [8] and Jones and McCleary [44])

$$\mathrm{RHom}_{C^*(S)^{\mathrm{op}}}(\mathbb{K}, \mathbb{K}) \simeq C_{-*}(\Omega_x S)^{\mathrm{op}}$$

($^{\mathrm{op}}$ operations get removed from both sides if one considers $\mathbb{K}$ as a left $C^*(S)$–module).

This duality is relevant to us because it induces an isomorphism at the level of Hochschild cohomology. Namely, by a general result of Keller [47] (see also Félix, Menichi and Thomas [36]) one obtains an isomorphism of Gerstenhaber algebras (in fact, of B_∞–algebras at the chain level)

$$\mathrm{HH}^*(C^*(S), C^*(S)) \cong \mathrm{HH}^*(C_{-*}(\Omega_x S), C_{-*}(\Omega_x S)).$$

In view of Abouzaid's generation result [4], the right-hand side is in turn isomorphic to $\mathrm{HH}^*(\mathcal{W}(M))$ as a Gerstenhaber algebra. On the other hand, the work of Bourgeois, Ekholm and Eliashberg [17] can be interpreted, over a field $\mathbb{K}$ of characteristic 0, to give an isomorphism of Gerstenhaber algebras

$$\mathrm{HH}^*(\mathcal{W}(M)) \cong \mathrm{SH}^*(M).$$

The group on the right-hand side is called *symplectic cohomology*. Strictly speaking, the results of [17] relate symplectic and Hochschild *homologies*. However, in our computations, we will use an explicit DG-algebra as a model for $\mathcal{W}(M)$, which has an (open) Calabi–Yau property (in the sense of Ginzburg [39]) implying an isomorphism between Hochschild homology and cohomology. This allows us to use the cohomological statement above that we find more convenient. Symplectic (co)homology of a Liouville manifold is a symplectic invariant based on an extension of Hamiltonian Floer (co)homology to noncompact symplectic manifolds. It was introduced by Viterbo [70] in its current form. We recommend Seidel [60] for an excellent introduction to symplectic cohomology and the recent manuscript Abouzaid [5] for more. Briefly, this is a very interesting invariant of a Stein manifold that is sensitive to the underlying symplectic structure (cf Eliashberg and Gromov [31]). Symplectic cohomology is in general difficult to calculate explicitly. However, Bourgeois, Ekholm and Eliashberg [16; 17] recently outlined a proof of a surgery formula for symplectic (co)homology. Combining this with the very recent work of Ekholm and Ng [28], one obtains a purely combinatorial description of symplectic cohomology of any 4–dimensional Weinstein manifold. (In the absence of 1–handles and when the coefficient field is $\mathbb{Z}_2$, one had Chekanov [19] as a precursor to [28].) This combinatorial description is in general still highly complicated. It involves noncommutative and infinite-dimensional chain complexes.

In the above setting, assuming that $\mathscr{A} = C^*(S)$ is a formal DG-algebra, that is, it is quasi-isomorphic to its homology $A = H^*(S)$, we get a promising way of computing symplectic cohomology. Namely, one has

$$\mathrm{HH}^*(H^*(S), H^*(S)) = \mathrm{SH}^*(M).$$

By a famous result of Deligne, Griffiths, Morgan and Sullivan [25], the formality assumption holds if S is a Kähler manifold and $\mathbb{K}$ has characteristic 0. Note that as a consequence of formality of $C^*(S)$, one has a bigrading on $\mathrm{HH}^*(C^*(S), C^*(S))$; there is a cohomological grading r associated with the Hochschild cochain complex and there is an internal grading s coming from the grading on $H^*(S)$. To get an isomorphism to $\mathrm{SH}^*(M)$, where the grading is given by a Conley–Zehnder type index, one has to consider the grading of the total complex corresponding to $r+s$.

Let us note that one could arrive at the same conclusion by combining theorems of Viterbo [70] and Cohen and Jones [20].

In this paper, we give a generalization of the above (in dimension 4) to Liouville manifolds $M = X_\Gamma$ obtained via plumbings of T^*S^2 according to a plumbing tree Γ. We will work over semisimple rings k of the form

$$\mathrm{k} = \bigoplus_v \mathbb{K} e_v,$$

where $e_v^2 = e_v$ and $e_v e_w = 0$ for $v \neq w$ and the index set of the sum is the vertex set Γ_0 of Γ.

To wit, using Floer complexes over $\mathbb{K}$, we set

$$\mathscr{A}_\Gamma = \bigoplus_{v,w} \mathrm{CF}^*(S_v, S_w),$$

where the S_v are the Lagrangian spheres corresponding to the zero sections of the cotangent bundles T^*S^2 that we are plumbing, and similarly we have

$$\mathscr{B}_\Gamma = \bigoplus_{v,w} \mathrm{CW}^*(L_v, L_w),$$

where L_v is a cotangent fiber at a chosen basepoint on S_v (away from the plumbing region).

In fact, assuming $\operatorname{char} \mathbb{K} = 0$, it turns out that $\mathscr{A}_\Gamma$ tends to be a formal DG-algebra (we can prove this when Γ is of type A_n or D_n, and conjecture it for E_6, E_7, E_8), hence, in such cases, we may replace it with $A_\Gamma = H^*(\mathscr{A}_\Gamma)$. Furthermore, very early on, we will replace $\mathscr{B}_\Gamma$ with a quasi-isomorphic DG-algebra (see [17, Proposition 4.11 and Theorem 5.8]) which has a combinatorial description. Namely, we will use Chekanov's DG-algebra [19], with the cohomological grading, associated to a Legendrian link $\Lambda_\Gamma = \bigcup \Lambda_v$ giving a Legendrian surgery diagram for X_Γ where the components are indexed by vertices v of Γ and each component Λ_v is a Legendrian unknot in $\mathbb{R}^3$ (see Figure 3). In the context of [17], the homologically graded version of this is also called the Legendrian contact homology algebra.

At this point, one obtains an explicit description of the DG-algebra $\mathscr{B}_\Gamma$. A careful choice of the surgery diagram (with suitable decorations) enables us to observe that the DG-algebra $\mathscr{B}_\Gamma$ is a *deformation* of Ginzburg's (cohomologically graded) DG-algebra $\mathscr{G}_\Gamma$ associated with the tree Γ (see Theorem 8).[1] Note that Ginzburg [39] associates a CY3 DG-algebra to any graph Γ and a potential function. In this paper, Γ is a tree and the potential function vanishes. On the other hand, since we are plumbing copies of T^*S^2, our DG-algebras are CY2. This generalization of the construction of Ginzburg's DG-algebra in order to obtain a CY2 DG-algebra appears in Van den Bergh [12]. (See Definition 5 for the definition of $\mathscr{G}_\Gamma$.)

The observation that $\mathscr{B}_\Gamma$ is a deformation of the corresponding Ginzburg DG-algebra $\mathscr{G}_\Gamma$ enables us to use the vast literature on the study of Ginzburg's DG-algebras to study symplectic invariants of X_Γ. Now, our discussion branches into two according to whether the underlying tree Γ is of Dynkin type or not.

[1] An earlier version of this manuscript claimed an isomorphism between $\mathscr{B}_\Gamma$ and $\mathscr{G}_\Gamma$, due to our blindness to some higher energy curves. We are indebted to the referee for opening our eyes.

Dynkin case For Γ of type A_n or D_n, we prove the following theorem:

Theorem 1 *For $\Gamma = A_n$ and $\mathbb{K}$ arbitrary field, or $\Gamma = D_n$ and $\mathbb{K}$ a field with* $\operatorname{char}\mathbb{K} \neq 2$*, there is a quasi-isomorphism of DG-algebras*

$$\mathscr{B}_\Gamma \simeq \mathscr{G}_\Gamma.$$

For $\Gamma = A_n$, this result follows from a direct computation of $\mathscr{B}_\Gamma$. However, for $\Gamma = D_n$, direct computation only shows that $\mathscr{B}_\Gamma$ is a certain deformation of $\mathscr{G}_\Gamma$. We then appeal to standard deformation theory arguments to show that this deformation is trivial when $\operatorname{char}\mathbb{K} \neq 2$. In fact, we also prove that $\mathscr{B}_\Gamma$ and $\mathscr{G}_\Gamma$ are not quasi-isomorphic when $\Gamma = D_n$ and $\operatorname{char}\mathbb{K} = 2$ by showing that the relevant obstruction class in $\mathrm{HH}^2(\mathscr{G}_\Gamma)$ is nontrivial.

We conjecture that $\mathscr{B}_\Gamma \simeq \mathscr{G}_\Gamma$ for $\Gamma = E_6, E_7$ if $\operatorname{char}\mathbb{K} \neq 2, 3$ and for $\Gamma = E_8$ if $\operatorname{char}\mathbb{K} \neq 2, 3, 5$, but we leave the study of these exceptional cases to a future work.

Assuming for brevity $\operatorname{char}\mathbb{K} \neq 2$, and $\Gamma = A_n$ or D_n, we can now write $\mathscr{B}_\Gamma \simeq \mathscr{G}_\Gamma$. For Γ of type ADE, $\mathscr{G}_\Gamma$ turns out to be nonformal; see Hermes [41]. Its cohomology has locally finite grading. Indeed, for an (algebraically closed) field with characteristic 0, it was computed in [41] that

$$H^*(\mathscr{G}_\Gamma) \cong \Pi_\Gamma \rtimes_\nu \mathrm{k}[u]$$

as a k–algebra, where Π_Γ is the preprojective algebra associated with the tree Γ, $|u| = -1$, and the multiplication is twisted by the Nakayama automorphism ν of Π_Γ. This is an involution, which is induced by an involution of the underlying Dynkin graph (see Section 3).

Because $\mathscr{G}_\Gamma$ is not formal, it is not immediately clear how to compute Hochschild cohomology of $\mathscr{G}_\Gamma$. To help with this, we prove in Section 5 the following:

Theorem 2 *Let $\mathbb{K}$ be any field. For any tree Γ, the associative algebra A_Γ is Koszul dual to the DG-algebra $\mathscr{G}_\Gamma$, in the sense that there are DG-algebra isomorphisms*

$$\mathrm{RHom}_{\mathscr{G}_\Gamma}(\mathrm{k},\mathrm{k}) \simeq A_\Gamma \quad \textit{and} \quad \mathrm{RHom}_{A_\Gamma^{\mathrm{op}}}(\mathrm{k},\mathrm{k}) \simeq \mathscr{G}_\Gamma^{\mathrm{op}}.$$

Therefore, by Keller's result [47], we can use this to compute $\mathrm{SH}^*(X_\Gamma)$ as

$$\mathrm{SH}^*(X_\Gamma) \cong \mathrm{HH}^*(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(A_\Gamma).$$

Since A_Γ is a rather small finite-dimensional algebra over k, one can find a minimal projective resolution to compute the latter group. Indeed, Brenner, Butler and King [18] give a minimal periodic (graded) resolution for A_Γ. However, we will find a shortcut to compute $\mathrm{HH}^*(A_\Gamma)$ as a bigraded algebra for $\Gamma = A_n, D_n$ over a field $\mathbb{K}$ of arbitrary characteristic. An explicit presentation of $\mathrm{HH}^*(A_\Gamma)$ as a (graded) commutative $\mathbb{K}$–algebra is provided in Theorem 40 for A_n and in Theorem 44 for D_n.

As we noted above in the case $\Gamma = D_n$ and when $\operatorname{char}\mathbb{K} = 2$, $\mathscr{B}_\Gamma$ is indeed a nontrivial deformation of $\mathscr{G}_\Gamma$. In this case $\mathscr{A}_\Gamma$ is not formal and indeed $\mathscr{B}_\Gamma$ and $\mathscr{A}_\Gamma$ are Koszul dual in the above sense. Therefore, it appears that a natural statement (that applies in all characteristics) may be that $\mathscr{A}_\Gamma$ and $\mathscr{B}_\Gamma$ are Koszul dual when Γ is a Dynkin tree.

Non-Dynkin case In this case, we only know that $\mathscr{B}_\Gamma$ is a deformation of $\mathscr{G}_\Gamma$ and even at the formal level there are many nontrivial deformations of $\mathscr{G}_\Gamma$ since $\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma)$ is big (see Theorem 30) and $\mathrm{HH}^3(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma) = 0$. Hence, it is not clear whether the deformation corresponding to $\mathscr{B}_\Gamma$ is trivial or not. On the other hand, as $\mathscr{B}_\Gamma$ (being a model for the wrapped Fukaya category of X_Γ) is also a Calabi–Yau (CY) algebra, one can see the deformation of $\mathscr{G}_\Gamma$ to $\mathscr{B}_\Gamma$ as a deformation of CY2-algebras. In characteristic 0, this allows one to conclude that the corresponding formal deformation is trivial as follows.

$\mathscr{G}_\Gamma$ is in a sense simpler for Γ non-Dynkin. Namely, in this case, the homology $H^*(\mathscr{G}_\Gamma)$ turns out to be concentrated in degree 0 and

$$H^0(\mathscr{G}_\Gamma) \cong \Pi_\Gamma$$

is the preprojective algebra associated with the tree Γ. For a non-Dynkin tree Γ, working over $\mathbb{K}$ of characteristic 0, Hermes [41] proved that $\mathscr{G}_\Gamma$ is formal, that is, $\mathscr{G}_\Gamma$ is quasi-isomorphic to its homology Π_Γ (see also Corollary 26 for another proof that works over any field). Furthermore, it is well-known that Π_Γ is Koszul in the classical sense (cf [54; 10]) over $\Bbbk$. The quadratic dual to Π_Γ is given by the associative algebra $A_\Gamma = H^*(\mathscr{A}_\Gamma)$ — the zigzag algebra associated with the tree Γ [43].

The Gerstenhaber algebra structure of the Hochschild cohomology of the preprojective algebra Π_Γ in the non-Dynkin case has already been computed by Crawley-Boevey, Etingof and Ginzburg in [23] when $\mathbb{K}$ has characteristic zero, and by Schedler [57] in general. $\mathrm{HH}^*(\Pi_\Gamma) \neq 0$ only for $* = 0, 1, 2$. We give a brief review of these computations of $\mathrm{HH}^*(\Pi_\Gamma)$ for completeness; see Section 6.1 for a full description. Now, $\mathscr{B}_\Gamma$ can be seen as a deformation of the CY2 algebra Π_Γ. If we consider the corresponding formal deformation, then the associated Kodaira–Spencer class lives in $\operatorname{Ker}(\Delta\colon \mathrm{HH}^2(\Pi_\Gamma) \to \mathrm{HH}^1(\Pi_\Gamma))$, where Δ is the BV-operator (see for example [35]). Now, it can be observed from the description given in Section 6.1 that this kernel is trivial if $\operatorname{char}\mathbb{K} = 0$. This result does not hold in arbitrary characteristic; see Remark 15 (cf Remark 33) for a proof that this deformation is nontrivial over a field $\mathbb{K}$ of characteristic 2.

Finally, let us remark that the above argument only shows that the associated formal deformation is trivial. This does not mean that $\mathscr{B}_\Gamma$ is quasi-isomorphic to $\mathscr{G}_\Gamma$ — such a quasi-isomorphism holds only after a certain completion. As was shown in our subsequent work [33], $H^0(\mathscr{B}_\Gamma)$ is isomorphic to the multiplicative preprojective algebra associated with Γ, introduced by Crawley-Boevey and Shaw [24]. On the other hand $H^0(\mathscr{G}_\Gamma)$ is isomorphic to the additive preprojective algebra Π_Γ. It is known that additive and multiplicative preprojective algebras are isomorphic only when $\operatorname{char}\mathbb{K}=0$ and Γ is Dynkin, and in general, they are isomorphic when $\operatorname{char}\mathbb{K}=0$ only after completion, as follows from the above deformation theory argument.

In Section 2, we provide geometric preliminaries on plumbings of cotangent bundles. In Section 3, we give a computation of Legendrian contact homology of the Legendrian link Λ_Γ associated to a tree Γ, show that it is isomorphic to a deformation of the corresponding CY2 Ginzburg DG-algebra $\mathscr{G}_\Gamma$ (Theorem 8) and that this deformation is trivial for $\Gamma = A_n$ or D_n, when $\operatorname{char}\mathbb{K}\neq 2$ in the latter case (Theorem 13). Section 4 computes the Floer cohomology algebra $\mathscr{A}_\Gamma$ of the spheres in X_Γ. The main result appears in Section 5, where we show that $\mathscr{G}_\Gamma$ and $A_\Gamma = H^*(\mathscr{A}_\Gamma)$ are Koszul duals for any tree Γ. Finally, as an application of our main result, in Section 6, we explicitly compute Hochschild cohomology of $\mathscr{G}_\Gamma$, hence also of $\mathscr{B}_\Gamma$ for $\Gamma = A_n$ and D_n, assuming $\operatorname{char}\mathbb{K}\neq 2$ if $\Gamma = D_n$.

Convention Throughout, we adhere to the following conventions. $\mathbb{K}$ is a field, of arbitrary characteristic unless otherwise specified, and k is a semisimple ring, generated over $\mathbb{K}$ by finitely many mutually orthogonal idempotents. Letters $A, B, \dots$ denote associative algebras over k. All our modules are *right* modules and all our multiplications are read from *right to left*. Letters $\mathscr{A}, \mathscr{B}, \dots$ denote A_∞– or DG-algebras over k. We follow the sign conventions as given in Seidel [61, Chapter 1] and its sequel Seidel [63]. In particular, an A_∞–algebra $\mathscr{A}$ over k is a $\mathbb{Z}$–graded k–module with a collection of k–linear maps

$$\mu^d\colon \mathscr{A}^{\otimes d} \to \mathscr{A}[2-d] \quad \text{for } d \geq 1,$$

where $[2-d]$ means μ^d lowers the degree by $d-2$. These maps are required to satisfy the A_∞–relations

$$\sum_{m,n} (-1)^{|a_1|+\cdots+|a_n|-n} \mu^{d-m+1}(a_d, \dots, a_{n+m+1}, \mu^m(a_{n+m}, \dots, a_{n+1}), a_n, \dots a_1) = 0.$$

A DG-algebra over k is an A_∞–algebra over k such that $\mu^d = 0$ for $d \geq 3$. In this case, we put

$$da = (-1)^{|a|}\mu^1(a), \quad a_2 a_1 = (-1)^{|a_1|}\mu^2(a_2, a_1). \tag{3}$$

With this convention the A_∞–equation for $d=2$ gives us the usual graded Leibniz rule

$$d(a_2a_1) = (da_2)a_1 + (-1)^{|a_2|}a_2(da_1).$$

$\mathscr{A}^{\mathrm{op}}$ denotes the opposite of an A_∞–algebra $\mathscr{A}$ and its operations are given by

$$\mu^d_{\mathscr{A}^{\mathrm{op}}}(a_d,\dots,a_1) = (-1)^{|a_1|+\dots+|a_d|-d}\mu^d_{\mathscr{A}}(a_1,\dots,a_d).$$

With the above conventions, a DG-algebra and its opposite are related via

$$d^{\mathrm{op}}(a) = (-1)^{|a|-1}da, \quad a_2a_1 = a_1a_2.$$

All our complexes are cohomological, ie the differential increases the grading by 1. It often happens that our complexes are bigraded. In this case, we denote these gradings by the pair (r,s), where r refers to a cohomological (or length) grading and s refers to an internal grading (the notation $|a|$ is used to denote the internal grading of a specific element). The grading $r+s$ is referred to as the total degree. If a second grading is not specified in the notation, for example as in $\mathrm{HH}^*(A_\Gamma)$, it is understood that the grading $*$ refers to the total degree.

The notation $\mathrm{HH}^*(A)$ is used to denote Hochschild cohomology of a graded $\mathbb{K}$–algebra A with coefficients in A. It is a bigraded algebra over $\mathbb{K}$. We write $\deg(x)$ for the total degree $r+s$ of a specific element. There are two binary $\mathbb{K}$–linear operations: an associative graded commutative product of bidegree $(0,0)$ and a Lie bracket of bidegree $(-1,0)$. These are called the cup product and Gerstenhaber bracket, respectively. The product is graded commutative:

$$xy = (-1)^{\deg(x)\deg(y)}yx.$$

The Gerstenhaber bracket is graded antisymmetric on $\mathrm{HH}^*(A)[1]$, that is,

$$[x,y] = -(-1)^{(\deg(x)-1)(\deg(y)-1)}[y,x].$$

Finally, Hochschild cohomology of a (formal) Calabi–Yau algebra can be equipped with a Batalin–Vilkovisky operator Δ of bidegree $(-1,0)$, and we have the following compatibility equation between these structures:

$$[x,y] = (-1)^{|x|}\Delta(xy) - (-1)^{|x|}\Delta(x)y - x\Delta(y).$$

Acknowledgements Lekili is partially supported by a Royal Society Fellowship and the NSF grant DMS-1509141. We thank Mohammed Abouzaid, Ben Antieau, Georgios Dimitroglou Rizell, Tobias Ekholm, Sheel Ganatra, Travis Schedler, Paul Seidel and Ivan Smith. We are especially grateful to the referee for a careful reading of the manuscript: in an earlier version of this paper, we used a more complicated Lagrangian

projection than the one given in Figure 4, which resulted in higher energy curves being immersed and elusive. We are indebted to the referee for communicating to us the existence of these higher-order contributions to the differential of $\mathscr{B}_\Gamma$.

2 Plumbing of cotangent bundles of 2–spheres

Let Γ be a finite tree. In the body of the paper, we will study Weinstein manifolds that are given by a plumbing of cotangent bundles of the 2–sphere according to the tree Γ. These are exact symplectic manifolds with a convexity condition at infinity. We briefly recall the construction of these manifolds (cf [3]).

Associated to each vertex of Γ, we prepare a copy of D^*S^2, the unit cotangent bundle of the 2–sphere with its canonical symplectic structure. Now, say we have an edge that connects the vertices v and w, and let us write D^*S_v and D^*S_w for the corresponding copies of T^*S^2 and choose basepoints $s_v \in S_v$ and $s_w \in S_w$. Near s_v and s_w one can find real coordinates p_1, p_2, q_1, q_2 where the coordinates q_1, q_2 correspond to variations on the base and the coordinates p_1, p_2 correspond to variations in the fiber direction. Furthermore, on these neighborhoods symplectic form can be identified with $dp \wedge dq$. We then glue D^*S_v and D^*S_w together near s_v and s_w via a symplectomorphism that sends (q, p) to $(p, -q)$.

This leads to a symplectic manifold which has a boundary with corners. One then smoothens the boundary and completes it to obtain a Weinstein manifold. The precise details of this construction are somewhat technical; we refer to [3, Section 2.3] (see also [37, Section 7.6]).

An alternative description of X_Γ can be given via *Legendrian surgery* à la Eliashberg [29] and Gompf [40], which we will take as primary.[2] In this description, we exhibit X_Γ as a surgery along a Legendrian link Λ on $(S^3, \xi_{\mathrm{std}})$ such that the vertices v of Γ correspond to the components Λ_v of this link, which are Legendrian unknots. Two such Legendrian unknots are "plumbed together" if there is an edge in Γ between the corresponding vertices. To be precise, by choosing a vertex as the *root* of our tree, we put our tree Γ in a standard form as in Figure 2, and the corresponding Legendrian unknots in a standard form in $(\mathbb{R}^3, dz - y\,dx)$, which when projected to (x, z) (front projection) gives the surgery diagram as in Figure 3.

The surgery construction equips X_Γ with a Weinstein structure (in fact, a Stein structure) by extending the standard Weinstein structure on D^4 via attaching 2–handles [73]

[2] Both the plumbing and surgery constructions lead to homotopic Weinstein manifolds but we do not check this here. Throughout, we use the surgery construction and appeal to the plumbing picture only for differential topological aspects.

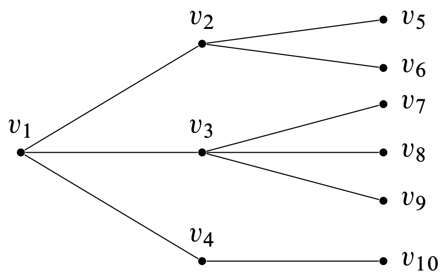


Figure 2: Standard form of Γ

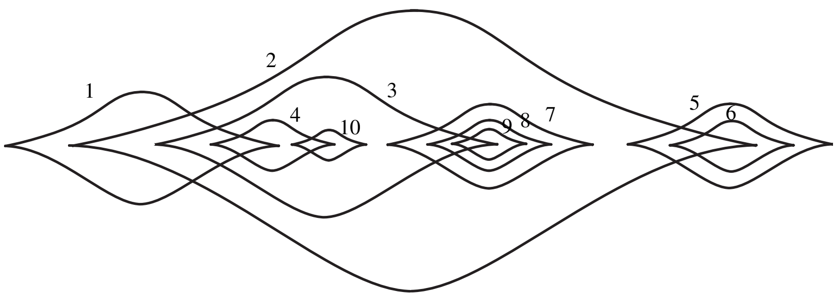


Figure 3: X_Γ as given by Legendrian surgery along Λ

along Legendrian unknots Λ_i. Each such Legendrian unknot bounds an embedded Lagrangian disk in D^4 and another capping disk given by the attaching disk of the corresponding Weinstein 2–handle. These fit together, as can be seen from the case of T^*S^2, to give the Lagrangian spheres S_v in X_Γ corresponding to the vertices of Γ, whereas the edges of Γ encode the intersection pattern of these spheres. The symplectic form ω on X_Γ is exact and it can be written as a primitive of a one-form θ for which the embedding of each sphere S_v is an exact Lagrangian submanifold of X_Γ. Both of these are easy facts since $H_2(X_\Gamma;\mathbb{Z})$ is generated by the Lagrangian spheres S_v and $H^1(S_v;\mathbb{Z})=0$.

Furthermore, the cocores of the 2–handles give noncompact (exact) Lagrangians L_v which are asymptotically Legendrian. The Lagrangian L_v intersects S_w only if $v=w$, in which case the intersection is transverse at a unique point x_v. In the plumbing description, the L_v correspond to the cotangent fibers $T^*_{x_v}S_v \subset T^*S_v$, where the x_v are basepoints on each S_v away from the plumbing regions.

In the next section, we will be concerned with Reeb chords between the components of Λ in $(\mathbb{R}^3, dz - y\,dx)$. The Reeb flow is in the direction of the vector field $\partial/\partial z$, hence it is more convenient for computations to consider the Lagrangian projection, ie the projection to (x,y) as in Figure 4. Then the crossings of the projection Λ are in one-to-one correspondence with Reeb chords from Λ to itself. There is some freedom in drawing the Lagrangian projection; we prefer the one given in Figure 4 as it makes enumeration of holomorphic curves manifest. (Notice that the diagram has the property that each component links at most one other component on its left. Clearly this is an artifact of the way we put our tree in a standard form and is not necessary.)

In Figure 4, besides a basepoint on each component, we also indicated an orientation on our Legendrian link Λ by putting an arrow on each component. This, in turn, induces orientations on the Lagrangian spheres S_v. Notice that

$$S_v \cdot S_w = +1$$

if v and w are adjacent vertices. This ensures that the Floer complex $\mathrm{CF}^*(S_v, S_w)$ is supported at an odd degree (see [59, Section 2d]).

We orient the noncompact Lagrangians L_v so that the algebraic intersection number $L_v \cdot S_v$ is given by

$$L_v \cdot S_v = -1.$$

As above, this ensures that the Floer complex $\mathrm{CF}^*(L_v, S_v)$ is supported at an even degree (which we will fix below to be 0 by picking suitable grading structures).

The classical topology of X_Γ is easy to study via the plumbing description, which shows that X_Γ deformation retracts onto a wedge of spheres formed by the union of

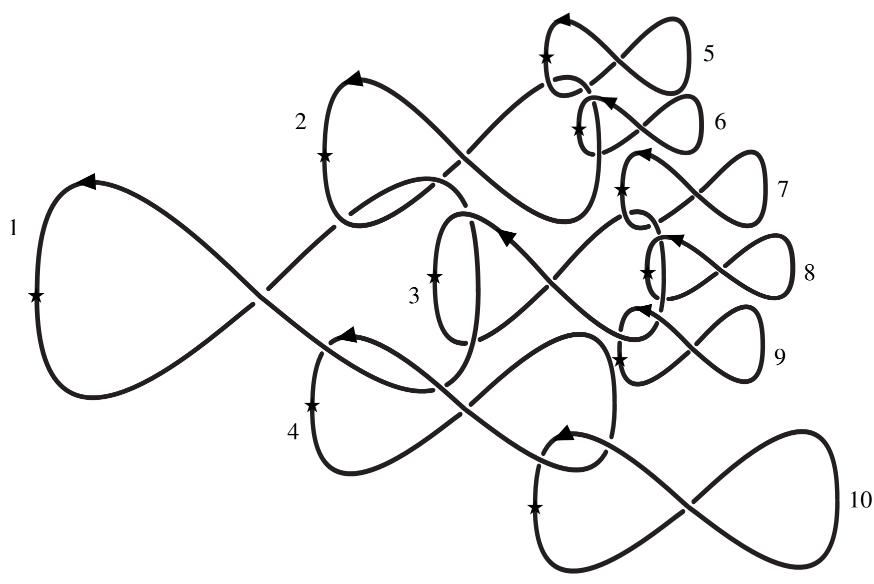


Figure 4: Lagrangian projection of Λ decorated with orientations and basepoints

the S_v. In particular, X_Γ is simply connected and the nonzero cohomology groups of X_Γ are given by

$$H^0(X_\Gamma;\mathbb{K}) = \mathbb{K}, \quad H^2(X_\Gamma;\mathbb{K}) = \bigoplus_v \mathbb{K}\cdot[S_v]^\vee.$$

The noncompact end of X_Γ is a symplectization of a contact 3–manifold Y_Γ which is topologically a plumbing of circle bundles over S^2 with Euler number -2. By abuse of notation, we will write $\partial X_\Gamma = Y_\Gamma$.

To equip our Lagrangians with a brane structure, so as to have $\mathbb{Z}$–gradings, we need:

Lemma 3 $$c_1(X_\Gamma,\omega) = 0.$$

Proof We have $\langle c_1(X_\Gamma), [S_v]\rangle = \mathrm{rot}(\Lambda_v)$ (see [40, Proposition 2.3]). Now, each Λ_v is an oriented Legendrian unknot in (S^3,ξ_{std}) and as such its rotation number can be computed to be $\mathrm{rot}(\Lambda_v) = 0$. □

Therefore, the canonical bundle $\mathcal{K} = \Lambda^2_{\mathbb{C}}(T^*X_\Gamma)$ representing $-c_1(X_\Gamma)$ is trivial. To define $\mathbb{Z}$–gradings in various Floer type invariants, one needs to fix a trivialization of $\mathcal{K}^{\otimes 2}$. Of course, since $H^1(X_\Gamma) = 0$, there is actually only one homotopy class of trivializations. We can induce a trivialization by picking a complexified volume form $\Omega \in \Lambda^2_{\mathbb{C}}(T^*X_\Gamma)$.

In this setup, a grading structure on a Lagrangian L can be thought of as a lift of the squared-phase map

$$\alpha_L\colon L \to S^1, \quad \alpha_L(x) = \frac{\Omega(T_xL)^2}{|\Omega(T_xL)^2|}$$

to a map $\tilde{\alpha}_L\colon L \to \mathbb{R}$. The fact that S_v and L_v are simply connected ensures that such a lift exists for our Lagrangians.

A grading structure allows one to associate an absolute Maslov index in $\mathbb{Z}$ to an intersection point $x \in S_v \cap S_w$ (see [59, Section 2d]). In our situation, all our Lagrangians S_v are simply connected, and if any two of them intersect they do so at a unique point. If x is the intersection point of S_v and S_w, then for any given $d \in \mathbb{Z}$ we can ensure that $x \in \mathrm{CF}^*(S_v, S_w)$ lies in degree d by shifting the grading structure on, say, S_w. When viewed as a generator of $\mathrm{CF}^*(S_w, S_v)$, the same intersection point would then be forced to have degree $2-d$ by Poincaré duality in Floer cohomology of compact Lagrangians (see [61, Section 12e]). Furthermore, since Γ is a tree, we can grade our Lagrangians inductively using the standard form of Γ as in Figure 2. Therefore we can grade all of our Lagrangians S_v at once such that for any pair of intersecting

Lagrangians S_v and S_w we are free to pick the gradings $(d, 2-d)$ as we would like. Collapsing a grading structure on a Lagrangian to a $\mathbb{Z}_2$–grading, we get an orientation of the underlying Lagrangian. To be compatible with the above choice of orientations for the Lagrangian spheres S_v, we will need to demand that the gradings d be odd. Throughout, a convenient choice will be to simply demand that $d = 1$, that is,

$$\mathrm{CF}^*(S_v, S_w) = \mathbb{K}[-1] \quad \text{if } v, w \text{ are adjacent.}$$

Having graded the Lagrangian spheres S_v for all v, we now pick grading structures for the noncompact Lagrangians L_v. As L_v is simply connected as well, we have the freedom to choose a grading such that

$$\mathrm{CF}^*(L_v, S_v) = \mathbb{K}[0].$$

This is compatible with our choice of orientations on L_v and S_v as given before.

These considerations fix the orientations and the grading data up to an overall shift (which does not change the degrees of intersection points) on our Lagrangians. (Note that there is a unique choice of Spin structures as our Lagrangians are simply connected.)

Somewhat more nontrivially, these choices force that if v and w are adjacent vertices, then we have the following.

Lemma 4 *For v and w adjacent vertices of the tree Γ, the shortest Reeb chord between L_v and L_w lies in the degree 0 part of $\mathrm{CW}^*(L_v, L_w)$. Furthermore, for any pair v, w, the complex $\mathrm{CW}^*(L_v, L_w)$ is supported in nonpositive degrees.*

Proof The first claim follows from a rigidity of a certain holomorphic square that contributes to the higher multiplication

$$\mu^3\colon \mathrm{HF}^0(L_v, S_v) \otimes \mathrm{HW}^0(L_w, L_v) \otimes \mathrm{HF}^2(S_w, L_w) \to \mathrm{HF}^1(S_w, S_v),$$

as explained in [7, Section 4.2]. The second claim is a consequence of the first by additivity properties of the Maslov grading (see [7, Lemma 4.11]). □

We do not use the above result in our computations below. We have stated and proved it as it helps motivate various grading choices (see also Remark 10). Let us also note that Theorem 23 below provides an indirect check of this lemma.

3 Ginzburg DG-algebra of Γ and Legendrian cohomology DG-algebra of Λ_Γ

3.1 Ginzburg DG-algebra of Γ

A *quiver* Q is a directed graph with a vertex set Q_0 and an arrow set Q_1. A rooted tree Γ in a standard form, as in Figure 2, gives rise to a quiver by orienting the edges so that they point *away* from the root. We will denote this quiver again by Γ unless otherwise specified. Recall that the path algebra $\mathbb{K}\Gamma$ of quiver Γ is defined as a vector space having all the paths in the quiver as basis (including, for each vertex v of the quiver Γ, a trivial path e_v of length 0), and multiplication is given by concatenation of paths. As mentioned before, throughout we concatenate paths from right to left, when we express them as a product.

The cohomologically graded 2–Calabi–Yau *Ginzburg DG-algebra* $\mathscr{G}_\Gamma$ of Γ (with zero potential) is defined as follows (see [39; 12; 41]).

Definition 5 Consider the extended quiver $\widehat{\Gamma}$ with vertices $\widehat{\Gamma}_0 = \Gamma_0$ and arrows $\widehat{\Gamma}_1$ consisting of

- the original arrows g in Γ_1 in bidegree $(1,-1)$;
- the opposite arrows g^* to g in Γ_1 in bidegree $(1,-1)$;
- loops h_v at the vertex $v \in \Gamma_0$ of bidegree $(1,-2)$.

We define $\mathscr{G}_\Gamma$ to be the DG-algebra over the semisimple ring $\mathrm{k} = \bigoplus_{v\in\Gamma_0} \mathbb{K}e_v$ given by the path algebra $\mathbb{K}\widehat{\Gamma}$ with the differential d of bidegree $(1,0)$ defined as a k–bimodule map by

$$dg = dg^* = 0 \quad \text{and} \quad dh = \sum_{g\in\Gamma_1} g^*g - gg^*,$$

where $h = \sum_{v\in\Gamma_0} h_v$.

In the notation (r,s) for bigraded complexes, r corresponds to the path-length grading and as usual we will call $r+s$ the total degree. In particular, the notation $H^*(\mathscr{G})$ will stand for the cohomology graded by the total degree. Note also that with respect to the total grading $\mathscr{G}_\Gamma$ is supported in nonpositive degrees.

The way we chose to orient the edges of Γ has only a minor effect on $\mathscr{G}_\Gamma$. Namely, different choices change the signs in the formula for the differential. Our choice is to ensure the consistency with the choice of orientations of the Lagrangians L_v, as we shall see in the next section. In particular, let Γ^{op} be the quiver obtained from Γ by reversing the orientation of all edges of Γ. Then the associated Ginzburg algebra gives $\mathscr{G}_\Gamma^{\mathrm{op}}$, the opposite of the Ginzburg algebra $\mathscr{G}_\Gamma$ associated to the original quiver Γ. In other words,

$$\mathscr{G}_{\Gamma^{\mathrm{op}}} = \mathscr{G}_\Gamma^{\mathrm{op}}.$$

Definition 6 The cohomology in total degree 0 of $\mathscr{G}_\Gamma$ is called the *preprojective algebra* $\Pi_\Gamma := H^0(\mathscr{G}_\Gamma)$. It is the quotient of the path algebra $\mathbb{K}\mathrm{D}\Gamma$ by the ideal generated by

$$\sum_{g \in \Gamma_1} g^* g - g g^*,$$

where $\mathrm{D}\Gamma$ denotes the double of Γ, obtained by adding the opposite arrow g^* for every $g \in \Gamma_1$.

It turns out that the nature of the DG-algebra $\mathscr{G}_\Gamma$ depends on whether Γ is of Dynkin type or not, as shown in the following theorem. It was first proven by Hermes [41] under the assumption that $\mathbb{K}$ is algebraically closed and characteristic 0. In Corollary 26, we give a proof of the first part of the theorem over an arbitrary field.

Theorem 7 (Hermes [41] and also Corollary 26) (1) *Suppose Γ is non-Dynkin. Then $H^*(\mathscr{G}_\Gamma) = \Pi_\Gamma$ is supported in degree 0 and is quasi-isomorphic to $\mathscr{G}_\Gamma$. In other words, $\mathscr{G}_\Gamma$ is formal.*

(2) *Suppose Γ is Dynkin and $\mathbb{K}$ is characteristic 0 and algebraically closed. Then*

$$H^*(\mathscr{G}_\Gamma) \cong \Pi_\Gamma \rtimes_\nu \mathrm{k}[u], \quad |u| = -1$$

as a k*–algebra, where the multiplication is twisted by the Nakayama automorphism ν on Π_Γ. Furthermore, $\mathscr{G}_\Gamma$ is not formal and there is an A_∞–structure $(\mu^n)_{n\geq 2}$ on the twisted polynomial algebra $\Pi_\Gamma \rtimes_\nu \mathrm{k}[u]$ making it a minimal model of $\mathscr{G}_\Gamma$. Moreover, this A_∞–structure is u–equivariant, and $\mu^n = 0$ for $n \neq 2, 3$.*

The *Nakayama automorphism* $\nu\colon \Pi_\Gamma \to \Pi_\Gamma$ in the above theorem refers to the automorphism defined by

$$\nu(g_{wv}) = \begin{cases} g_{\rho(w)\rho(v)} & \text{if } g_{wv} \in \Gamma \text{ or } g_{\rho(w)\rho(v)} \in \Gamma, \\ -g_{\rho(w)\rho(v)} & \text{if } g_{vw}, g_{\rho(v)\rho(w)} \in \Gamma, \end{cases}$$

where g_{wv} denotes the arrow from the vertex v to w in Π_Γ, and ρ denotes either the natural involution of the Dynkin graph (precisely when Γ is of type A_n, D_{2n+1} or E_6) or the identity. We will abuse the notation and always denote the arrow from v to w by g_{wv} regardless of where it is considered, in the quiver Γ, its double $D\Gamma$, the extended quiver $\widehat{\Gamma}$ or in the algebras $\mathscr{G}_\Gamma$ and Π_Γ, for that matter. In particular, $g_{vw} = g^*_{wv}$ if g_{wv} belongs to Γ. Note that ν has order at most 2 and it is the identity if and only if Γ is of type A_1 or it is of type D_{2n}, E_7 or E_8 and the base field $\mathbb{K}$ is of characteristic 2 (see [18, Definition 4.6]).

3.2 Legendrian cohomology DG-algebra of Λ_Γ

We recall the definition of the $\mathbb{Z}$–graded Chekanov–Eliashberg DG-algebra of the Legendrian link $\Lambda_\Gamma = \bigcup \Lambda_v$ following [17, Section 4], where it is denoted as $\mathrm{LHA}(\Lambda_\Gamma)$. It was originally introduced in [30; 19].

Let $\mathcal{R}$ denote the finite set of Reeb chords from Λ_Γ to itself. Recall from Section 2 that $\mathcal{R}$ is in bijection with the set of crossings in the Lagrangian projection of Λ_Γ (Figure 4). We endow the vector space $\mathbb{K}\langle\mathcal{R}\rangle$ with a k–bimodule structure by declaring

$$e_w \mathcal{R} e_v$$

to be the set of Reeb chords from Λ_w to Λ_v. As a k–module, $\mathrm{LHA}(\Lambda)$ is the tensor algebra over the semisimple ring k given by

$$\mathrm{LHA}_*(\Lambda_\Gamma) := \bigoplus_{i=0}^{\infty} \mathbb{K}\langle\mathcal{R}\rangle^{\otimes_{\mathrm{k}} i}.$$

After decorating Λ_Γ with extra data by orienting each component and picking a basepoint at each component as in Figure 4, the chords $c \in \mathcal{R}$ acquire a kind of Conley–Zehnder grading by $\mathbb{Z}$ which we denote by $|c|$. The subscript in the notation of $\mathrm{LHA}_*(\Lambda_\Gamma)$ denotes the induced grading on the tensor algebra. Elements $e_v \in \mathrm{k}$ have degree 0; however, in general there may also be Reeb chords which have degree 0. The differential $D\colon \mathrm{LHA}_*(\Lambda_\Gamma) \to \mathrm{LHA}_{*-1}(\Lambda_\Gamma)$ is defined as a map $D\colon \mathbb{K}\langle\mathcal{R}\rangle_* \to \mathrm{LHA}_{*-1}(\Lambda_\Gamma)$ and extended by the graded Leibniz rule to $\mathrm{LHA}_*(\Lambda)$.

Note that in general the differential is not compatible with the path-length grading corresponding to the index i in the definition of $\mathrm{LHA}_*(\Lambda)$.

As we follow the cohomological convention to be consistent with the literature on Fukaya categories, instead of $\mathrm{LHA}_*(\Lambda)$ we will use the cohomologically graded DG-algebra $\mathrm{LCA}^*(\Lambda)$. As a k–module, it is given by

$$\mathrm{LCA}^*(\Lambda_\Gamma) := \mathrm{LHA}_{-*}(\Lambda_\Gamma).$$

The differential $D\colon \mathrm{LCA}^*(\Lambda_\Gamma) \to \mathrm{LCA}^{*+1}(\Lambda_\Gamma)$ is just carried over from the one on $\mathrm{LHA}_*(\Lambda_\Gamma)$.

Let us describe the Legendrian cohomology DG-algebra of Λ_Γ more explicitly. The underlying algebra of $\mathrm{LCA}^*(\Lambda_\Gamma)$ is the tensor algebra of the k–bimodule $\mathbb{K}\langle\mathcal{R}\rangle$ generated by the Reeb chords (ie crossings in Figure 4):

$$\mathcal{R} = \{c_{wv}, c_{vw} : g_{wv} \in \Gamma_1\} \cup \{c_v : v \in \Gamma_0\},$$

where c_v is the Reeb chord at the unique self-crossing of the component Λ_v, and for every two adjacent vertices v and w of the tree Γ, c_{wv} corresponds to the unique Reeb chord from Λ_w to Λ_v, ie the chord at the unique crossing between Λ_v and Λ_w where Λ_w is the undercrossing component.

Notice the remarkable coincidence of the k–bimodule structure on $\mathrm{LCA}^*(\Lambda_\Gamma)$ and the k–bimodule structure on $\mathscr{G}_\Gamma$ from Definition 5. Next, we will see that the differentials do not agree in general. Nonetheless the Legendrian cohomology DG-algebra is isomorphic to a deformation of the Ginzburg algebra.

Theorem 8 *If Λ_Γ is the Legendrian link in the standard form associated to the tree Γ with Lagrangian projection in Figure 4 with the grading decoration as indicated, then there is an isomorphism between $(\mathrm{LCA}^*(\Lambda_\Gamma), D)$ and a deformation of $(\mathscr{G}_\Gamma, d)$ as DG-algebras. More precisely, there is a graded derivation $\mathfrak{d}\colon \mathscr{G}_\Gamma \to \mathscr{G}_\Gamma$ with homogeneous components $\mathfrak{d} = d_3 + d_5 + \cdots + d_{2m-1}$ for some $m \geq 1$, d_{2i-1} having bidegree $(2i-1, 2-2i)$, and there is an isomorphism of DG-algebras*

$$(\mathrm{LCA}^*(\Lambda_\Gamma), D) \simeq (\mathscr{G}_\Gamma, d + \mathfrak{d})$$

such that the Conley–Zehnder degree on the left-hand side agrees with the total degree on the right-hand side.

Proof **Generators** The natural one-to-one correspondence, ie $g_{wv} \leftrightarrow c_{wv}$, $h_v \leftrightarrow c_v$, between the arrow set $\widehat{\Gamma}_1$ of the extended quiver $\widehat{\Gamma}$ and the set $\mathcal{R}$ of Reeb chords provides the isomorphism of the underlying k–algebras, the path algebra $\mathbb{K}\widehat{\Gamma}$ and the tensor algebra of $\mathbb{K}\langle\mathcal{R}\rangle$. Note that the Reeb orientation of the chord c_{wv} is from Λ_w to Λ_v, whereas the arrow g_{wv} goes from the vertex v to w.

Gradings It suffices to identify the gradings of the generators. We first recall the definition for an arbitrary Legendrian link $\Lambda \subset (S^3, \xi_{\mathrm{std}})$.

According to the original combinatorial description [19], LCA has a $\mathbb{Z}/r\mathbb{Z}$–grading, where r is the gcd of the rotation numbers of the components. In our case, each component of Λ_Γ is an unknot with rotation number 0, providing a $\mathbb{Z}$–grading on $\mathrm{LCA}^*(\Lambda_\Gamma)$.

Let $z_\pm$ be the endpoints of a Reeb chord c of an oriented Legendrian link Λ equipped with basepoints on every component, z_+ being the one with the greater z–coordinate. Let $\gamma_\pm$ be the shortest paths in Λ, from $z_\pm$ to the basepoint of the corresponding component, in the direction of the orientation of Λ. The grading of c in LCA is defined to be $2r_- - 2r_+ + \frac{1}{2}$, where $r_\pm \in \mathbb{Q}$ is the number of counterclockwise rotations the tangent vector of $\gamma_\pm$ makes (in the xy–plane). It is straightforward to verify that the grading of every generator of the form c_v of $\mathrm{LCA}(\Lambda_\Gamma)$ is -1 and that of the form c_{wv} is 0.

Differential We briefly recall the definition of the differential of LCA for any Legendrian link in the standard contact S^3, and then compute the differentials on the set $\mathcal{R}$ of generators of $\mathrm{LCA}^*(\Lambda_\Gamma)$. The rest will be determined by the Leibniz rule.

To simplify the definition, we arrange that at every crossing of the Lagrangian projection, the understrand and the overstrand have slopes $+1$ and -1, respectively. We also use the same notation for a crossing in the Lagrangian projection as the corresponding Reeb chord.

First of all, each quadrant around a crossing in the Lagrangian projection is decorated with a Reeb sign. The right and left quadrants at a crossing have positive signs whereas the top and bottom quadrants have negative signs.

There is also a second set of signs, orientation signs, for these quadrants. Every quadrant has orientation sign $+1$ except for the bottom and right quadrants at an even-graded crossing, which are decorated with -1, as in Figure 5. In fact, the choice of orientation signs for a given diagram depends on an isotopy of the diagram near the crossing so that the strand with a positive slope goes under the strand with a negative slope, as in Figure 5. We indicated our choice in the upper left diagram of Figure 6. This affects the signs, but different choices give isomorphic DG-algebras (see [28, page 80]).

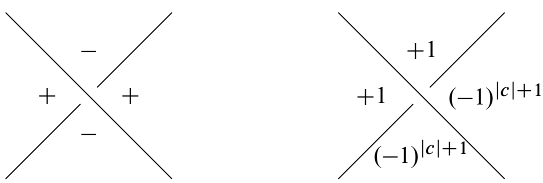


Figure 5: Reeb signs (left) and orientation signs (right) at a crossing c

On a generator, the differential is given by a count of immersed polygons and it is extended by the graded Leibniz rule. The polygons taken into account are in the xy–plane with boundary on the Lagrangian projection of the link and vertices at the crossings. It is also required that at all but one vertex of the polygon, the quadrant included in the polygon should have a negative *Reeb* sign. Suppose that Δ is such an immersed polygon whose positive vertex is at c and the negative vertices $c_1, c_2, \dots, c_m$ are in order as we traverse the boundary of Δ counterclockwise starting at c. Note that m may be 0 and the c_i are not necessarily distinct. If b is the total number of times the boundary of Δ passes through basepoints of the Legendrian link, the orientation sign ϵ_Δ is defined to be $(-1)^b$ times the product of the *orientation* signs at the vertices.

With this setup, we have

$$dc = \sum_{\Delta} \epsilon_\Delta c_m c_{m-1} \cdots c_1$$

for any generator c. Observe that the differential of a generator of the form c_{wv} vanishes since it has grading 0 and $\mathrm{LCA}^*(\Lambda_\Gamma)$ is nonpositively graded. Again for grading reasons, any negative vertex of an immersed polygon which contributes to the differential of a generator c_v is of type c_{uw}.

In the rest of the proof we will show that

$$D(c_v) = -\sum_{u:g_{vu}\in\Gamma_1} c_{vu}c_{uv} + \sum_{i\geq 1}\sum_{\substack{w_1,\dots,w_i\\ g_{w_j v}\in\Gamma_1\\ w_1<\dots<w_i}} c_{vw_1}c_{w_1v}\cdots c_{vw_i}c_{w_iv},$$

where the ordering in the last summation refers to the clockwise ordering of the components of Λ_Γ which are linked to v from the right in the Lagrangian projection in Figure 4, eg the natural ordering of the integers associated to components in Figure 4. Note that the second sum not only corresponds to higher-order terms in the length filtration, it also contributes terms of word-length 2 of the form $c_{vw_1}c_{w_1v}$. Indeed, all the terms of word-length 2 that appear in the image of $D(c_v)$ precisely correspond to $d(c_v)$ in $\mathscr{G}_\Gamma$. In particular, the first sum has at most one term as long as our Legendrian link is associated to a tree in the standard form.

We will prove that all the terms in the above differential are induced by *embedded* polygons as indicated in Figure 6, the relevant piece of the Lagrangian projection

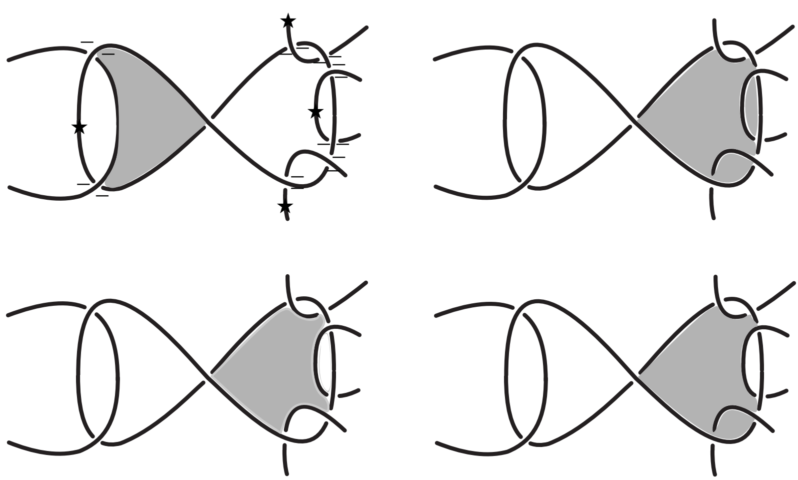


Figure 6: The polygons which correspond to the words in the differential $D(c_v)$: (from top left in clockwise order) a triangle (with a negative orientation sign), a triangle, a pentagon, and a heptagon (all with positive orientation signs). The quadrants with negative orientation signs and the basepoints are indicated in the upper left diagram.

given in Figure 4, together with the orientation signs at the crossings. There are also two unigons with a unique vertex at c_v, one to the left and the other to the right with canceling contributions to the differential $D(c_v)$ since they come with opposite signs.

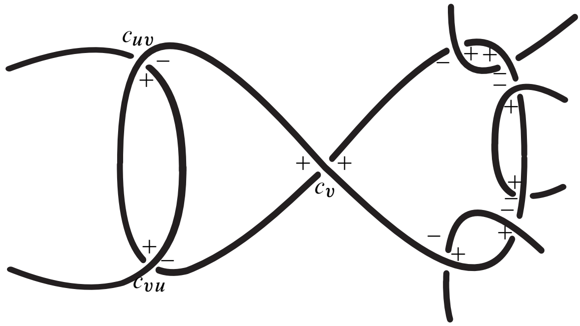


Figure 7: Reeb signs

We now prove that there are no other immersed polygons which contribute to the differential $D(c_v)$. To begin with, any such polygon has a (Reeb-) positive vertex at c_v (see Figure 7 for the Reeb signs at the relevant crossings). Start traversing its boundary in the counterclockwise direction assuming that the polygon includes the left quadrant at c_v. If it has a vertex other than c_v, ie if it is not the unigon canceled by a similar unigon to the right, then the only option for an initial negative vertex is at c_{uv} because of the configuration of the Reeb signs. Moreover, this vertex has to be followed (as we continue traversing the boundary) by a vertex at c_{vu} since otherwise the polygon would intersect the region outside the Lagrangian projection, which is prohibited. Similar considerations imply that a polygon which includes the right quadrant at c_v can only have vertices at the crossings of Λ_v with other components of Λ as shown in Figure 6 above so as not to intersect the noncompact region. □

Remark 9 A relation between Ginzburg's construction of CY3 DG-algebras associated with quivers (with potentials) and Fukaya categories of certain quasiprojective 3–folds also appears in the work of Smith [69].

Remark 10 Recall that $\mathrm{LCA}^*(\Lambda_\Gamma)$ is associated to the Legendrian attaching spheres Λ_v of Weinstein 2–handles. Stated results of [17] provide a dual picture given in terms of the wrapped Floer cohomology of the cocores L_v of these handles induced by cobordism maps associated to the handle attachments. Namely, there is a grading-preserving quasi-isomorphism of A_∞–algebras

$$\mathrm{LCA}^*(\Lambda_\Gamma) \simeq \bigoplus_{v,w} \mathrm{CW}^*(L_v, L_w).$$

A rigorous justification of the equivalence of these two dual pictures is not fully established at this time. However, a detailed sketch of proof based on the results of [17] has recently appeared in [27, Theorem 2]. We must emphasize that we do not make use of this correspondence anywhere in our computations. Rather, this appealing geometric picture serves us as a guide to find the correct algebraic statement to be proven rigorously.

3.2.1 Recourse to deformation theory of DG-algebras As a consequence of the explicit computation given above we can see the Legendrian cohomology DG-algebra $\mathrm{LCA}^*(\Lambda_\Gamma)$ as a deformation of the Ginzburg DG-algebra $\mathscr{G}_\Gamma$. Therefore, it is natural to check whether this deformation is trivial or not (up to equivalence). We recall here the basics of deformation theory of DG-algebras and exploit it to determine the relationship between our computation of $\mathrm{LCA}^*(\Lambda_\Gamma)$ and the Ginzburg DG-algebra $\mathscr{G}_\Gamma$. A classical reference for this material is [38]. A recent exposition close to our purpose appears in [65, Appendix A].

Unfortunately, these methods do not help directly as they apply in the setting of formal deformations (such as a deformation over $\mathrm{k}[\![t]\!]$) whereas here we have that $\mathrm{LCA}^*(\Lambda_\Gamma)$ is a global deformation of $\mathscr{G}_\Gamma$ (over $\mathrm{k}[t]$). Nonetheless, it is helpful to start at the formal level and observe that we can arrange for a globalization in certain cases.

There is a decreasing, exhaustive, bounded-above filtration on the complex $\mathrm{LCA}^*(\Lambda_\Gamma)$:

$$\mathcal{F}^0 := \mathrm{LCA}^*(\Lambda_\Gamma) \supset \mathcal{F}^1 := \bigoplus_{i=1}^{\infty} \mathbb{K}\langle\mathcal{R}\rangle^{\otimes_{\mathrm{k}}^i} \supset \cdots \supset \mathcal{F}^p := \bigoplus_{i=p}^{\infty} \mathbb{K}\langle\mathcal{R}\rangle^{\otimes_{\mathrm{k}}^i} \supset \cdots .$$

Let us write $(\mathrm{LCA}^*(\Lambda_\Gamma), D) = (\mathscr{G}_\Gamma, d_1 + d_2 + \cdots + d_m)$, for some finite m, where $d_i\colon \mathcal{F}^p \to \mathcal{F}^{p+i}$ is the i^{th} homogeneous piece of the differential. Observe that $d_1 = d$ can be identified as the differential in the Ginzburg DG-algebra. It follows from k–linearity of the differential that in fact d_i is identically zero for even i. Note also that since $\mathscr{G}_\Gamma$ is bigraded, this complex is doubly graded. Denoting the second grading by s, we have $s(d_{2i-1}) = 2-2i$.

Now, the first nontrivial d_i for $i > 1$ is possibly d_3. Because $D^2 = 0$, using the filtration, we deduce that

$$d_1 d_3 + d_3 d_1 = 0.$$

Recall that the reduced bar complex $(\hom_{\mathrm{k}}(\mathrm{T}\overline{\mathscr{G}_\Gamma}, \mathscr{G}_\Gamma), \delta = \delta_0 + \delta_1)$ can be used to compute Hochschild cohomology of $\mathscr{G}_\Gamma$. Here, we only need the explicit form of the Hochschild differential for elements $\phi \in \hom_{\mathrm{k}}(\overline{\mathscr{G}_\Gamma}, \mathscr{G}_\Gamma)$ (see formula in [61, Equation (1.8)], which we adapted using DG-algebra conventions given in the introduction).

For such ϕ, we have

$$(-1)^{|\phi|+|b|}(\delta_0\phi)(a\otimes_{\Bbbk} b) = a\phi(b) + (-1)^{|\phi||b|}\phi(a)b - \phi(ab),$$
$$(-1)^{|\phi|+|a|}(\delta_1\phi)(a) = d\phi(a) - \phi(da).$$

By definition, $\mathscr{G}_\Gamma$ is bigraded and its differential has bidegree $(1,0)$, so the Hochschild cochain complex $\mathrm{CC}^*(\mathscr{G}_\Gamma,\mathscr{G}_\Gamma) = \hom_{\Bbbk}(\mathrm{T}\overline{\mathscr{G}_\Gamma},\mathscr{G}_\Gamma)$ has three gradings: the cohomological degree, the degree induced by the total degree $r+s$ on $\mathscr{G}_\Gamma$ and the internal grading induced by the second grading s on $\mathscr{G}_\Gamma$. However, the Hochschild differential $\delta = \delta_0+\delta_1$ is homogeneous (of degree 1) with respect to the sum of the first two gradings and it also preserves the internal degree, hence we get a bigrading on $\mathrm{HH}^*(\mathscr{G}_\Gamma,\mathscr{G}_\Gamma)$, which we write as

$$\mathrm{HH}^*(\mathscr{G}_\Gamma,\mathscr{G}_\Gamma) \cong \bigoplus_{r,s} \mathrm{HH}^r(\mathscr{G}_\Gamma,\mathscr{G}_\Gamma[s]), \tag{4}$$

where r is the total degree (the sum of the cohomological degree and the degree induced by the total degree on $\mathscr{G}_\Gamma$) and s is the internal grading induced by the internal grading on $\mathscr{G}_\Gamma$.

Now, the fact that d_3 is a degree-1 derivation which anticommutes with d_1 means that the sign-modified map $\tilde{d}_3 \in \hom^1_{\Bbbk}(\overline{\mathscr{G}_\Gamma},\mathscr{G}_\Gamma)$, defined by

$$\tilde{d}_3 a = (-1)^{|a|} d_3 a,$$

is closed under the Hochschild differential. This yields the first obstruction class of the deformation:

$$[\tilde{d}_3] \in \mathrm{HH}^2(\mathscr{G}_\Gamma,\mathscr{G}_\Gamma[-2]).$$

If this class is trivial, choosing a trivializing class $\phi_2 \in \hom^0_{\Bbbk}(\overline{\mathscr{G}_\Gamma},\mathscr{G}_\Gamma[-2])$, we get a map ϕ_2 for which we have

$$d_3 = d\phi_2 - \phi_2 d.$$

Note that ϕ_2 is induced by a map $\mathbb{K}\langle\mathcal{R}\rangle \to \mathbb{K}\langle\mathcal{R}\rangle^{\otimes_{\Bbbk} 3}$. Therefore, we can consider an *algebra* map

$$\Phi_2 = \mathrm{Id} + \phi_2 \colon \mathscr{G}_\Gamma \to \mathscr{G}_\Gamma$$

defined initially as a map on $\mathbb{K}\langle\mathcal{R}\rangle \to \mathscr{G}_\Gamma$ and then extended to an algebra map.

Then, we would like to define a new differential D' on $\mathscr{G}_\Gamma$ of the form

$$D' = d + d'_5 + \cdots$$

so that $\Phi_2\colon (\mathscr{G}_\Gamma, D') \to (\mathscr{G}_\Gamma, D)$ is a chain map (in addition to being an algebra map). The obvious candidate for D' is given by

$$D' = (\mathrm{Id} - \phi_2 + \phi_2^2 - \cdots) \circ D \circ (\mathrm{Id} + \phi_2).$$

However, the alternating sum $(\mathrm{Id} - \phi_2 + \phi_2^2 - \cdots)$ will in general be an infinite series; therefore, to make sense of this we need to consider the completion of $\mathscr{G}_\Gamma$ with respect to the length filtration $\mathcal{F}^\bullet$:

$$\widehat{\mathscr{G}_\Gamma} = \varprojlim_p \mathscr{G}_\Gamma / \mathcal{F}^p \mathscr{G}_\Gamma.$$

The differential D of $\mathrm{LCA}^*(\Lambda_\Gamma)$ extends naturally to $\widehat{\mathscr{G}_\Gamma}$. We write the resulting complex as

$$\widehat{\mathrm{LCA}}(\Lambda_\Gamma) = (\widehat{\mathscr{G}_\Gamma}, D)$$

Concretely, we can write the underlying k–bimodule as $\widehat{\mathrm{LCA}}(\Lambda_\Gamma) = \mathbb{K}\langle \mathcal{R} \rangle [\![t]\!]$, where t is a formal parameter in degree 0. In other words, we now allow formal power series in Reeb chords.

We can now proceed with the construction mentioned above. Notice that since ϕ_2 increases the length by 2, there is no convergence issue for the series $(\mathrm{Id} - \phi_2 + \phi_2^2 - \cdots)$ on $\widehat{\mathscr{G}_\Gamma}$. Therefore, we have a filtered DG-algebra map

$$\Phi_2\colon (\widehat{\mathscr{G}_\Gamma}, D') \to (\widehat{\mathscr{G}_\Gamma}, D)$$

which by construction is a chain map with an inverse, hence is in particular a quasi-isomorphism.

We can then focus on the complex $(\widehat{\mathscr{G}_\Gamma}, D' = d + d_5' + \cdots)$. As before, we have that d_5' is a derivation which anticommutes with d, hence the sign-twisted map $\widetilde{d}_5'$ leads to an obstruction class $[\widetilde{d}_5'] \in \mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma[-4])$. If this vanishes we can continue along and find a quasi-isomorphism of the form $\mathrm{Id} + \phi_4$. Iterating this argument infinitely many times (which we can do as each quasi-isomorphism increases the length), we obtain the following lemma (cf [65, Lemma A.5]).

Lemma 11 *Suppose that* $\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma[s]) = 0$ *for all* $s < 0$*. Then there exists a quasi-isomorphism of completed DG-algebras*

$$(\widehat{\mathscr{G}_\Gamma}, d) \simeq (\widehat{\mathrm{LCA}}(\Lambda_\Gamma), D).$$

We next apply these ideas to the case where $\Gamma = D_n$ and show that all the obstructions vanish in this case. Furthermore, we prove that one can truncate the above quasi-isomorphism, eliminating the need for completions. Here, we make use of the results of Section 6.2.3, where $\mathrm{HH}^*(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma)$ is computed for $\Gamma = D_n$. We would like to point out that the computation given there is independent of the conclusions we are drawing here.

The following lemma is the key technical result that we will use to truncate the quasi-isomorphism given on completions by the above deformation theory argument.

Lemma 12 *Let $\mathcal{F}^\bullet$ denote the length filtration on* $\mathrm{LCA}^*(\Lambda_{D_n})$. *For each grading k, there exists a $p(k)$ such that for all $p \geq p(k)$ we have that*

$$\mathcal{F}^p H^k(\mathrm{LCA}(\Lambda_{D_n})) = \mathrm{Im}\big(H^k(\mathcal{F}^p \,\mathrm{LCA}(\Lambda_{D_n})) \to H^k(\mathrm{LCA}(\Lambda_{D_n}))\big) = 0.$$

In particular, for all k, the filtration on $H^k(\mathrm{LCA}(\Lambda_{D_n}))$ induced by $\mathcal{F}^\bullet$ is complete and Hausdorff.

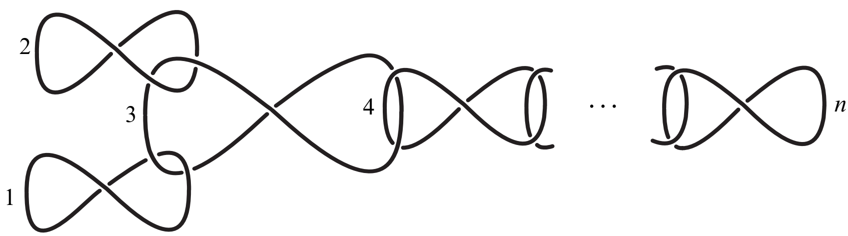


Figure 8: Lagrangian projection of a Legendrian link associated to the D_n tree

Proof Consider the Lagrangian projection in Figure 8. The proof of Theorem 8 gives us the following description of the differential on $(\mathrm{LCA}^*(\Lambda_{D_n}), D)$:

$$\begin{aligned}
Dc_1 &= c_{13}c_{31}, \\
Dc_2 &= c_{23}c_{32}, \\
Dc_3 &= -c_{31}c_{13} - c_{32}c_{23} + c_{34}c_{43} - c_{31}c_{13}c_{32}c_{23}, \\
Dc_4 &= -c_{43}c_{34} + c_{45}c_{54}, \\
&\ \ \vdots \\
Dc_{n-1} &= -c_{(n-1)(n-2)}c_{(n-2)(n-1)} + c_{(n-1)n}c_{n(n-1)}, \\
Dc_n &= -c_{n(n-1)}c_{(n-1)n},
\end{aligned}$$

where the gradings are given by $|c_i| = -1$ and $|c_{ij}| = 0$. In particular, $H^*(\mathrm{LCA}(\Lambda_{D_n}))$ is supported in nonpositive degrees.

Notice that $D = d_1 + d_3$, where d_1 is the differential on the Ginzburg DG-algebra $\mathscr{G}_{D_n}$ and d_3 is zero on all the generators except c_3, and we have

$$d_3(c_3) = -c_{31}c_{13}c_{32}c_{23}.$$

We shall first establish the result for $H^0(\mathrm{LCA}(\Lambda_{D_n}))$ by direct computation. The goal here is to take any word in c_{ij} and prove that if the word is long enough, then it is actually null-homologous.

Note that we have a decomposition

$$H^0(\mathrm{LCA}(\Lambda_{D_n})) \cong \bigoplus_{i,j=1}^{n} e_i H^0(\mathrm{LCA}(\Lambda_{D_n}))e_j.$$

Letting $x = c_{31}c_{13}$, $y = c_{32}c_{23}$ and $z = c_{34}c_{43}$ we obtain

$$e_3 H^0(\mathrm{LCA}(\Lambda_{D_n}))e_3 \cong \mathbb{K}\langle x, y, z\rangle/(x^2, y^2, z^{n-2}, x + y + xy - z)$$

(cf [57, Proposition 11.3.2(i)]). Indeed, we have

$$x^2 = D(c_{31}c_1c_{13}), \quad y^2 = D(c_{32}c_2c_{23}), \quad x + y + xy - z = D(-c_3).$$

Next, observe that for $4 \le i \le n-1$, we have $c_{i(i-1)}c_{(i-1)i} = c_{i(i+1)}c_{(i+1)i} \in H^0(\mathrm{LCA}(\Lambda_{D_n}))$ since their difference is precisely Dc_i. Consequently, we get

$$\begin{aligned} z^{n-2} &= c_{34}(c_{43}c_{34})^{n-3}c_{43} = c_{34}(c_{45}c_{54})^{n-3}c_{43} = c_{34}c_{45}(c_{56}c_{65})^{n-4}c_{54}c_{43} = \cdots \\ &= c_{34}c_{45}\cdots c_{(n-1)n}c_{n(n-1)}c_{(n-1)n}c_{n(n-1)}\cdots c_{54}c_{43} \\ &= D(-c_{34}c_{45}\ldots c_{(n-1)n}c_nc_{n(n-1)}\ldots c_{54}c_{43}). \end{aligned}$$

Furthermore, any word in $e_3 H^0(\mathrm{LCA}(\Lambda_{D_n}))e_3$ is cohomologous to a word in x, y, z which is of the same length (note that the lengths of x, y and z are 2). Namely, whenever a word w has terms which goes along the long branch of the D_n tree, it has to return back at some point, hence it will include a subword of the form $c_{i(i+1)}c_{(i+1)i}$ which can be replaced with $c_{i(i-1)}c_{(i-1)i}$ applying the relation Dc_i. This can be repeated until we replace each subword that lies in the long branch by a power of z.

Arguing similarly, one can see why it suffices to consider $e_3 H^0(\mathrm{LCA}(\Lambda_{D_n}))e_3$ to prove the statement in the lemma for the zeroth cohomology. Indeed, the relations given by $Dc_4, Dc_5, \ldots, Dc_n$ can be used to show that any sufficiently long word in $\mathrm{LCA}^0(\Lambda_{D_n})$ can be replaced by a word which contains a sufficiently long subword in $e_3\,\mathrm{LCA}^0(\Lambda_{D_n})e_3$. More precisely, for any word $w \in \langle c_{ij} \mid i, j = 1, n\rangle$ we can write

$$w = \alpha v \beta + \langle \mathrm{Im}\, D\rangle$$

such that v lies in $e_3\,\mathrm{LCA}^0(\Lambda_{D_n})e_3$ and is sufficiently long. In fact, since we only use the preprojective relations, Dc_i for $i \ne 3$, one can show that the analogue of [57, Proposition 11.3.2(ii)] holds in this case.

We can simplify the presentation of $e_3 H^0(\mathrm{LCA}(\Lambda_{D_n}))e_3$ further by eliminating the z variable and write

$$e_3 H^0(\mathrm{LCA}(\Lambda_{D_n}))e_3 \cong \mathbb{K}\langle x, y\rangle/(x^2, y^2, (x+y+xy)^{n-2}).$$

Let us define two-sided ideals

$$I_n = (x^2, y^2, (x+y+xy)^{n-2}) \quad \text{and} \quad J_n = (x^2, y^2, (x+y)^{n-2})$$

in $\mathbb{K}\langle x, y\rangle$ and claim that they are equal for $n \geq 4$. Note that in $\mathbb{K}\langle x, y\rangle/J_n$ any word that is long enough is trivial; in particular, this is a finite-dimensional vector space. This is because the only words that are not killed by the relations $x^2 = y^2 = 0$ are words alternating in x and y, and sufficiently long such words are killed by $x(x+y)^{n-2}y$ and $y(x+y)^{n-2}x$. Therefore the result for $H^0(\mathrm{LCA}(\Lambda_{D_n}))$ follows from the claim $I_n = J_n$.

To prove this claim, first observe that $A = x+y$ and $B = x+y+xy$ satisfy

$$B^2 = (1+x)A^2(1+y) \in \mathbb{K}\langle x, y\rangle/(x^2, y^2).$$

Moreover, since $(1+x)(1-x) = 1 = (1+y)(1-y)$ the above identity leads to $A^2 = (1-x)B^2(1-y)$ and together they show $I_4 = J_4$. We similarly obtain $I_5 = J_5$, using the observation

$$B^3 = (1+x)A^3(1+x)(1+y) \in \mathbb{K}\langle x, y\rangle/(x^2, y^2).$$

The fact that A^2 is in the center of $\mathbb{K}\langle x, y\rangle/(x^2, y^2)$ implies

$$\begin{aligned} B^{2k} &= (B^2)^k = (1+x)A^{2k}(1+y)(1+x)\cdots(1+y), \\ B^{2k+1} &= B^3(B^2)^{k-1} = (1+x)A^{2k+1}(1+y)(1+x)\cdots(1+y), \end{aligned}$$

proving $I_n = J_n$ for every $n \geq 4$.

Alternatively, one can check that a noncommutative Gröbner basis (with respect to the lexicographical order) for both I_n and J_n is given by the collection of the following three elements:

$$\{x^2, y^2, xyxy\cdots + yxyx\cdots\}$$

where the lengths of the words in the last element are $n-2$.

This completes the proof of the lemma for $H^0(\mathrm{LCA}(\Lambda_{D_n}))$. It is much harder to directly compute $H^i(\mathrm{LCA}(\Lambda_{D_n}))$ for $i < 0$ and verify Hausdorffness of the length filtration. Fortunately, there is an alternative way to go about this, making use of a recent result of Dimitroglou Rizell [26] which in turn exploits the weak division algorithm in free noncommutative algebras due to P M Cohn [22]. This is a general result about Legendrian cohomology DG-algebras which states that the natural algebra homomorphism

$$H^*(\mathrm{LCA}(\Lambda_\Gamma)) \to \mathrm{LCA}^*(\Lambda_\Gamma)/\langle \mathrm{Im}\, D\rangle$$

induced by inclusion is injective, where $\langle \mathrm{Im}\, D\rangle$ denotes the two-sided ideal in the tensor algebra $\mathrm{LCA}^*(\Lambda_\Gamma)$ generated by the image of the differential. In view of this, it suffices to show that for each k there exists a $p(k)$ such that if w is a word in c_{ij} of length greater than $p(k)$ containing exactly k instances of c_i, then w is in $\langle \mathrm{Im}\, D\rangle$.

This is, however, quite straightforward given what we have already proven. Namely, in any such word, since the number of degree -1 generators, c_i, is precisely k as soon as the length is sufficiently large, we can find a sufficiently long subword consisting of degree 0 generators c_{ij} only. Now, we proved above that any sufficiently long word in the degree 0 generators c_{ij} is in the image of D. Thus, the result follows. □

Note that the corresponding result also holds true for $\mathscr{G}_{D_n}$ but this is much simpler. The cohomology $H^*(\mathscr{G}_{D_n})$ is a graded filtered algebra, where the filtered subalgebras $\mathcal{F}^p H^*(\mathscr{G}_{D_n})$ for $p \geq 0$ are induced by the length filtration on $\mathscr{G}_{D_n}$. We claim that this filtration on $H^*(\mathscr{G}_{D_n})$ is complete and Hausdorff. To see this, observe the image of the differential of $\mathscr{G}_{D_n}$ consists of homogeneous terms (with respect to length filtration), hence the filtration is Hausdorff. The filtration is complete because $H^*(\mathscr{G}_{D_n})$ is finite-dimensional at each degree. To see this, when $\mathbb{K}$ is algebraically closed and of characteristic 0, one can use the result by Hermes (see Theorem 7) that $H^i(\mathscr{G}_{D_n}) \cong \Pi_{D_n}$ for every $i \geq 0$, and the well-known fact that the preprojective algebra of a Dynkin quiver is finite-dimensional. Alternatively, for any field, $H^0(\mathscr{G}_{D_n}) = \Pi_{D_n}$ by definition, hence we can appeal to the argument given in the last part of the above lemma to conclude. (Note that the result of [26] requires an action filtration on the chain complex respected by the differential. This is automatic for $\mathrm{LCA}^*(\Lambda_\Gamma)$ as the relevant filtration is given by the geometric action functional. On the other hand, if the complex is supported in nonpositive (or nonnegative) degrees, then one can easily construct an action filtration of the required type inductively, hence the main result of [26] is applicable to $\mathscr{G}_\Gamma$ as well for any Γ.)

We are now ready to prove the main result of this section:

Theorem 13 *Let $\Gamma = A_n$ or D_n, and assume that* $\mathrm{char}\, \mathbb{K} \neq 2$ *if $\Gamma = D_n$. Then there exists a quasi-isomorphism*

$$\mathrm{LCA}^*(\Lambda_\Gamma) \simeq \mathscr{G}_\Gamma.$$

Furthermore, if $\mathrm{char}\, \mathbb{K} = 2$ *and $\Gamma = D_n$, then $\mathrm{LCA}^*(\Lambda)$ and $\mathscr{G}_\Gamma$ are not quasi-isomorphic.*

(We conjecture that $\mathrm{LCA}^*(\Lambda) \simeq \mathscr{G}_\Gamma$ for $\Gamma = E_6, E_7$ if $\mathrm{char}\,\mathbb{K} \neq 2, 3$ and for $\Gamma = E_8$ if $\mathrm{char}\,\mathbb{K} \neq 2, 3, 5$.)

Proof The case when $\Gamma = A_n$ is immediate since $\mathrm{LCA}^*(\Lambda_\Gamma)$ and $\mathscr{G}_\Gamma$ are identical in this case. So we will focus on the case $\Gamma = D_n$.

When $\mathrm{char}\,\mathbb{K} \neq 2$, we will construct a chain map $\Phi\colon \mathscr{G}_\Gamma \to \mathrm{LCA}^*(\Lambda_\Gamma)$ which is of the form

$$\Phi = \mathrm{Id} + \text{h.o.t.},$$

where h.o.t. stands for higher-order terms in terms of the length filtration $\mathcal{F}^\bullet$ on $\mathrm{LCA}^*(\Lambda_\Gamma)$.

In Section 6.2.3, we computed

$$\mathrm{HH}^*(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma) \cong \mathrm{HH}^*(A_\Gamma, A_\Gamma),$$

where A_Γ is the Koszul dual to $\mathscr{G}_\Gamma$ as proven in Theorem 23. Note that the isomorphism between the Hochschild cohomologies of $\mathscr{G}_\Gamma$ and A_Γ is a consequence of the Koszul duality given by Theorem 23, which also states that the Koszul duality functor sends the internal grading of $\mathscr{G}_\Gamma$ to those of A_Γ, implying that the internal gradings on their Hochschild cohomologies match as well. In particular, we have

$$\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma[s]) \cong \mathrm{HH}^{2-s}(A_\Gamma, A_\Gamma[s]).$$

Let us warn the reader of a potentially confusing point in our notation. On the right-hand side, $r = 2 - s$ refers to the length grading in Hochschild cohomology, and s refers to the internal grading induced from the internal grading of the algebra A_Γ. This group is a summand of $\mathrm{HH}^2(A_\Gamma, A_\Gamma)$ where $2 = r + s$ is the total degree. On the other hand, $\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma[s])$ is a summand of $\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma)$ where s refers to the second grading on $\mathscr{G}_\Gamma$ (as was explained after (4)).

The computation given in Section 6.2.3 implies that for $\Gamma = D_n$ and when $\mathrm{char}\,\mathbb{K} \neq 2$, we have

$$\mathrm{HH}^2(\mathscr{G}_\Gamma, \mathscr{G}_\Gamma[s]) = 0 \quad \text{for } s < 0.$$

Therefore, from Lemma 11, we deduce that there exists a quasi-isomorphism

$$\Phi\colon \widehat{\mathscr{G}_\Gamma} \to \widehat{\mathrm{LCA}}^*(\Lambda_\Gamma).$$

Now, let N be an integer large enough that $\mathcal{F}^N H^0(\mathrm{LCA}(\Lambda_\Gamma)) = 0$; such an N exists, as we proved above in Lemma 12. We then consider the truncation of Φ at length N to define an algebra map between *uncompleted* algebras

$$\Phi^N\colon \mathscr{G}_\Gamma \to \mathrm{LCA}^*(\Lambda_\Gamma).$$

The apparent problem with Φ^N is that it is not a chain map, though it fails to be a chain map only at large length. So, we can correct it as follows. For each vertex v, let us find a chain α_v such that

$$D\Phi^N(h_v) - \Phi^N(dh_v) = D\alpha_v.$$

Note that the left-hand side is automatically D–closed since it lies in $\mathrm{LCA}^0(\Lambda_\Gamma)$.

We then define a new algebra map by setting

$$\Psi(h_v) := \Phi^N(h_v) + \alpha_v, \quad \Psi(g_{vw}) := \Phi^N(g_{vw}).$$

We now have a filtered chain map $\mathscr{G}_\Gamma \to \mathrm{LCA}^*(\Lambda_\Gamma)$ which respects the length filtrations on each side. Note that the E_2–pages of the associated spectral sequences are identical:

$$E_2^{p,q} \cong \mathcal{F}^p\mathscr{G}_\Gamma / \mathcal{F}^{p+2}, \mathscr{G}_\Gamma$$

with the differential induced from the differential on the Ginzburg DG-algebra. Furthermore, the length filtration is not only complete and Hausdorff on both sides by Lemma 12 and the discussion following its proof, but also easily seen to be weakly convergent. Therefore the spectral sequences converge *strongly* to $H^*(\mathscr{G}_{D_n})$ and $H^*(\mathrm{LCA}(\Lambda_{D_n}))$, respectively. Moreover, since

$$\Psi = \mathrm{Id} + \mathrm{h.o.t.},$$

where h.o.t. refers to a higher-order term that sends $\mathcal{F}^\bullet$ to $\mathcal{F}^{\bullet+2}$, it induces an isomorphism on the E_2–page, therefore we conclude that it induces a quasi-isomorphism of chain complexes by [15, Theorem 2.6]. This completes the proof that $\mathrm{LCA}^*(\Lambda_{D_n})$ and $\mathscr{G}_{D_n}$ are quasi-isomorphic over a field of characteristic $\neq 2$.

Next suppose that $\mathbb{K}$ is a field of characteristic 2. Let us write $D = d + d_3$ for the differential on $\mathrm{LCA}^*(\Lambda_{D_n})$ where, in the notation of Lemma 12, we have

$$d_3(c_3) = -c_{31}c_{13}c_{32}c_{23}.$$

We want to show that there is no degree 0 derivation ϕ_2 which increases length by 2 and solves $d_3 = d\phi_2 - \phi_2 d$. For $\Gamma = D_4$, this is equivalent to the following set of *linear* equations:

$$\begin{aligned} 0 &= d\phi_2(c_1) - \phi_2(c_{13})c_{31} - c_{13}\phi_2(c_{31}), \\ 0 &= d\phi_2(c_2) - \phi_2(c_{23})c_{32} - c_{23}\phi_2(c_{32}), \\ -c_{31}c_{13}c_{32}c_{23} &= d\phi_2(c_3) + \phi_2(c_{31}c_{13} + c_{32}c_{23} - c_{34}c_{43}), \\ 0 &= d\phi_2(c_4) + \phi_2(c_{43})c_{34} + c_{43}\phi_2(c_{34}). \end{aligned}$$

(Although we are working over characteristic 2 here, we have kept the signs in their general form for reference.)

Since ϕ_2 is supposed to preserve the degree and increase the length by 2, there are only a few possibilities. The general form of the possibilities is as follows:

$$\begin{aligned}
\phi_2(c_1) &\in \mathbb{K}c_1c_{13}c_{31} \oplus \mathbb{K}c_{13}c_{31}c_1 \oplus \mathbb{K}c_{13}c_3c_{31},\\
\phi_2(c_2) &\in \mathbb{K}c_2c_{23}c_{32} \oplus \mathbb{K}c_{23}c_{32}c_2 \oplus \mathbb{K}c_{23}c_3c_{32},\\
\phi_2(c_3) &\in \mathbb{K}c_3c_{31}c_{13} \oplus \mathbb{K}c_{31}c_{13}c_3 \oplus \mathbb{K}c_3c_{32}c_{23} \oplus \mathbb{K}c_{32}c_{23}c_3 \oplus \mathbb{K}c_3c_{34}c_{43}\\
&\qquad \oplus \mathbb{K}c_{34}c_{43}c_3 \oplus \mathbb{K}c_{31}c_1c_{13} \oplus \mathbb{K}c_{32}c_2c_{23} \oplus \mathbb{K}c_{34}c_4c_{43},\\
\phi_2(c_4) &\in \mathbb{K}c_4c_{43}c_{34} \oplus \mathbb{K}c_{43}c_{34}c_4 \oplus \mathbb{K}c_{43}c_3c_{34},\\
\phi_2(c_{13}) &\in \mathbb{K}c_{13}c_{31}c_{13} \oplus \mathbb{K}c_{13}c_{32}c_{23} \oplus \mathbb{K}c_{13}c_{34}c_{43},\\
\phi_2(c_{31}) &\in \mathbb{K}c_{31}c_{13}c_{31} \oplus \mathbb{K}c_{32}c_{23}c_{31} \oplus \mathbb{K}c_{34}c_{43}c_{31},\\
\phi_2(c_{23}) &\in \mathbb{K}c_{23}c_{32}c_{23} \oplus \mathbb{K}c_{23}c_{31}c_{13} \oplus \mathbb{K}c_{23}c_{34}c_{43},\\
\phi_2(c_{32}) &\in \mathbb{K}c_{32}c_{23}c_{32} \oplus \mathbb{K}c_{31}c_{13}c_{32} \oplus \mathbb{K}c_{34}c_{43}c_{32},\\
\phi_2(c_{43}) &\in \mathbb{K}c_{43}c_{34}c_{43} \oplus \mathbb{K}c_{43}c_{31}c_{13} \oplus \mathbb{K}c_{43}c_{32}c_{23},\\
\phi_2(c_{34}) &\in \mathbb{K}c_{34}c_{43}c_{34} \oplus \mathbb{K}c_{31}c_{13}c_{34} \oplus \mathbb{K}c_{32}c_{23}c_{34}.
\end{aligned}$$

This leads to a system of 18 linear equations of 36 variables. It is straightforward, if tedious, to verify directly (or with the help of a computer) that none of the possibilities gives a solution when $\mathbb{K} = \mathbb{Z}_2$. This, in turn, implies that the class of $[\widetilde{d}_3]$ is nontrivial over any field $\mathbb{K}$ of characteristic 2 by the universal coefficient theorem.

This implies that there is a nonvanishing obstruction for constructing a chain map between $\mathscr{G}_{D_4}$ and $\mathrm{LCA}^*(\Lambda)$ over a field of characteristic 2 for D_4. In other words, the class $[\widetilde{d}_3] \in \mathrm{HH}^2(\mathscr{G}_{D_4}, \mathscr{G}_{D_4}[-2])$ is nontrivial. (Compare this with our computation of $\mathrm{HH}^2(\mathscr{G}_{D_4}, \mathscr{G}_{D_4}[-2])$ given later on in Table 4, where this group is shown to be nontrivial only in characteristic 2.) Now, the class of $[\widetilde{d}_3]$ for $\Gamma = D_n$ restricts to the class of $\Gamma = D_4$ under the restriction map. (Note that in general Hochschild cohomology does not have good functoriality properties; however, there is a full and faithful inclusion of the $\mathscr{G}_{D_4}$ to $\mathscr{G}_{D_n}$, and there is a restriction map on Hochschild cohomology in this case.) Hence, it cannot vanish for $\Gamma = D_n$ either. □

Remark 14 Over a field of characteristic $\neq 2$, and for $\Gamma = D_4$, we constructed an explicit chain map between $\mathscr{G}_{D_4}$ and $\mathrm{LCA}^*(\Lambda_{D_4})$ as a check on our arguments above. The complication in this also displays the effectiveness of the deformation theory argument given above. (Notice the factors of $\frac{1}{2}$, which are indeed necessary.) The map is given as follows:

$$
\begin{aligned}
h_1 &\mapsto c_1 - \tfrac{1}{2}(c_{13}c_{31}c_1 + c_{13}c_3c_{31} + c_1c_{13}c_{32}c_{23}c_{31}),\\
h_2 &\mapsto c_2 - \tfrac{1}{2}(c_{23}c_{32}c_2 + c_{23}c_3c_{32} + c_{23}c_{31}c_{13}c_{32}c_2)\\
&\qquad\qquad\qquad + \tfrac{1}{4}\big(c_{23}c_{34}c_{43}c_3c_{32} + c_{23}c_{34}c_4c_{43}c_{32}\\
&\qquad\qquad\qquad\qquad + c_{23}c_{34}c_{43}c_{32}c_2 + c_{23}c_{34}c_{43}c_{31}c_{13}c_{32}c_2\big),\\
h_3 &\mapsto c_3 - \tfrac{1}{4}\big(c_{31}c_{13}c_3c_{34}c_{43} + c_{31}c_1c_{13}c_{34}c_{43}\\
&\qquad\qquad\qquad + c_{31}c_{13}c_{34}c_4c_{43} + c_{31}c_1c_{13}c_{32}c_{23}c_{34}c_{43}\big),\\
h_4 &\mapsto c_4 - \tfrac{1}{2}\big(c_4c_{43}c_{34} + c_{43}c_3c_{34} - c_{43}c_3c_{32}c_{23}c_{34} - c_{43}c_{32}c_2c_{23}c_{34}\\
&\qquad\qquad\qquad - c_4c_{43}c_{32}c_{23}c_{34} - c_{43}c_{31}c_{13}c_{32}c_2c_{23}c_{34}\big),\\
g_{13} &\mapsto c_{13} + \tfrac{1}{2}(c_{13}c_{32}c_{23} - c_{13}c_{34}c_{43}),\\
g_{31} &\mapsto c_{31},\\
g_{23} &\mapsto c_{23} - \tfrac{1}{2}c_{23}c_{34}c_{43},\\
g_{32} &\mapsto c_{32} + \tfrac{1}{2}c_{31}c_{13}c_{32},\\
g_{34} &\mapsto c_{34} - \tfrac{1}{2}(c_{32}c_{23}c_{34} + c_{31}c_{13}c_{34}),\\
g_{43} &\mapsto c_{43}.
\end{aligned}
$$

Remark 15 One can deduce from the argument given in the last part of the proof of Theorem 13 that for any tree Γ which is not of type A_n, we have that $\mathscr{B}_\Gamma := \mathrm{LCA}^*(\Lambda_\Gamma)$ is a nontrivial deformation of $\mathscr{G}_\Gamma$ over a field of characteristic 2 since any such tree has a subtree of the form D_4 (see also Remark 33).

4 Floer cohomology algebra of the spheres in X_Γ

We next consider the A_∞–algebra over $\Bbbk$ given by the Floer cochain complexes:

$$\mathscr{A}_\Gamma := \bigoplus_{v,w} \mathrm{CF}^*(S_v, S_w).$$

Recall that the Lagrangian 2–spheres S_v and S_w intersect only if the vertices v and w are connected by an edge, in which case $S_v \cap S_w$ is a unique point. Recall also that we made choices of grading structures on the sphere S_v in Section 2 so that $\mathrm{CF}^*(S_v, S_w)$ is concentrated in degree 1 if v, w are adjacent vertices. On the other hand, the self-Floer cochain complex $\mathrm{CF}^*(S_v, S_v)$ is quasi-isomorphic to the singular chain complex $C^*(S_v)$ since S_v is an exact Lagrangian sphere in X_Γ. Therefore, we can take a model for $\mathscr{A}_\Gamma$ such that the differential on $\mathscr{A}_\Gamma$ necessarily vanishes for degree reasons.

Let us put $A_\Gamma = H^*(\mathscr{A}_\Gamma)$ for the corresponding associative algebra. We can think of $\mathscr{A}_\Gamma$ as a minimal A_∞–structure $(\mu^n)_{n\geq 2}$ on the associative algebra A_Γ. As before, by choosing a root, we make Γ into a directed graph such that oriented edges point away from the root. Let $\mathrm{D}\Gamma$ denote the double of the quiver Γ, formed by introducing a new oriented edge a_{vw} from w to v for every oriented edge a_{wv} from v to w.

Proposition 16 *Suppose $\Gamma \neq A_1$. The graded* k*–algebra A_Γ is isomorphic to the* ***zigzag algebra*** *of Γ given by the path algebra $\mathbb{K}\mathrm{D}\Gamma$ equipped with the path-length grading modulo the homogeneous ideal generated by the following elements:*

- *$a_{uv}a_{vw}$ such that $u \neq w$, where v is adjacent to both u, w.*
- *$a_{vw}a_{wv} - a_{vu}a_{uv}$, where v is adjacent to both u, w.*

If $\Gamma = A_1$, then $A_\Gamma \cong H^*(S^2) = \mathbb{K}[x]/(x^2)$ with $|x| = 2$.

Proof Note that S_v intersects S_w for $w \neq v$ if and only if v and w are adjacent vertices, in which case the intersection is transverse at a unique point. Furthermore, we have chosen the grading structures on the Lagrangians S_v so as to ensure that for v, w adjacent $\mathrm{CF}^*(S_v, S_w)$ is of rank 1 and concentrated in degree 1. We let a_{vw} be a generator for this 1–dimensional vector space. Finally, the algebra structure is determined by the general Poincaré duality property of Floer cohomology (see [61, Section 12e]). □

The algebra A_Γ only depends on the underlying tree Γ; different ways of orienting its edges results in the same algebra. We call the algebra A_Γ the *zigzag algebra* of Γ, following Khovanov and Huerfano [43], who studied properties of this algebra and its appearances in a variety of areas related to representation theory and categorification. On the other hand, the case where Γ is the A_n quiver appeared in an earlier paper of Seidel and Thomas [67] in the context of Floer cohomology (as it does here) and mirror symmetry. In the context of Koszul duality (see [54; 10]), the algebras A_Γ were studied much earlier by Martínez-Villa in [52]. This remarkable work is the first paper, as far as we know, which draws attention to the fact that A_Γ is a Koszul algebra if and only if Γ is not Dynkin or $\Gamma = A_1$.

We will next discuss formality of $\mathscr{A}_\Gamma$, ie the question of whether there is a quasi-isomorphism between $\mathscr{A}_\Gamma$ and $A_\Gamma = H^*(\mathscr{A}_\Gamma)$. In the case when Γ is the A_n quiver, the formality was proven by Seidel and Thomas [67, Lemma 4.21] based on the notion of *intrinsic formality*.

Definition 17 A graded algebra A is called intrinsically formal if any A_∞–algebra $\mathscr{A}$ with $H^*(\mathscr{A}) \cong A$ is quasi-isomorphic to A.

Furthermore, Seidel and Thomas give a useful method to recognize intrinsically formal algebras. Recall that for a graded algebra A, $\mathrm{HH}^*(A)$ has two gradings: the cohomological grading r and the grading s coming from the grading of the algebra A. To specify the decomposition into graded pieces, we write

$$\mathrm{HH}^*(A) = \bigoplus_{*=r+s} \mathrm{HH}^r(A, A[s]).$$

Notice that the superscript denotes the diagonal grading, as usual. It is also the grading that survives, if A is more generally a DG-algebra or an A_∞–algebra.

Theorem 18 (Kadeishvili [45]; see also Seidel and Thomas [67]) *Let A be an augmented graded algebra. If*

$$\mathrm{HH}^{2-s}(A, A[s]) = 0 \quad \textit{for all } s < 0,$$

then A is intrinsically formal.

As mentioned above, Seidel and Thomas proved intrinsic formality of A_Γ where Γ is the A_n quiver by showing the vanishing of $\mathrm{HH}^{2-s}(A_\Gamma, A_\Gamma[s])$ for $s < 0$. In a similar vein, we prove in Theorem 44 that A_Γ is intrinsically formal if Γ is the D_n quiver and the characteristic of the ground field is not 2.

We have the following conjecture for the remaining Dynkin types.

Conjecture 19 *Working over a ground field $\mathbb{K}$ of characteristic 0, let Γ be a tree of type E_6, E_7 or E_8. Then the corresponding zigzag algebra A_Γ is intrinsically formal.*

Unlike the A_n case, some restriction on the characteristic of $\mathbb{K}$ is necessary as we have checked that the zigzag algebras are not intrinsically formal in type D_n, $n \geq 4$, over characteristic 2, in type E_6 and E_7 over characteristic 2 or 3, and in the type E_8, over characteristic 2, 3 or 5. It is very likely that these are the only "bad" characteristics (cf [57]).

5 Koszul duality

By combining the work of Bourgeois, Ekholm and Eliashberg [17] with Abouzaid's generation criteria [1], one might suspect that the Lagrangians L_v split-generate the wrapped Fukaya category $\mathcal{W}(X_\Gamma)$. Now, there exists a full and faithful embedding

$$\mathcal{F}(X_\Gamma) \to \mathcal{W}(X_\Gamma)$$

of the exact Fukaya category of compact Lagrangians. Therefore, in view of Remark 10, we would conclude that there is a quasi-isomorphism of DG-algebras

$$\mathrm{RHom}_{\mathscr{B}_\Gamma}(\mathrm{k},\mathrm{k}) \simeq \mathscr{A}_\Gamma. \tag{5}$$

The right-hand side is in turn quasi-isomorphic to A_Γ if one checks that $\mathscr{A}_\Gamma$ is formal (for example this is known if Γ is of type A_n [67] and we prove it in Theorem 44 for type D_n over a field of characteristic $\neq 2$). We will provide an alternative independent approach via a purely algebraic argument based on Koszul duality theory for DG- or A_∞–algebras (see [51]) to stay within the algebraic framework of this paper (and avoid the technicalities that go into the discussion in Remark 10).

In fact, as we shall see below, Koszul duality theory allows us to work directly with $A_\Gamma = H^*(\mathscr{A}_\Gamma)$, hence in this way we bypass formality questions for $\mathscr{A}_\Gamma$.

We now give a brief review of Koszul duality, first in the case of associative algebras and then for A_∞–algebras.

5.1 Quadratic duality and Koszul algebras

To begin with, we review quadratic duality for associative algebras following [64, Section 2.1] which has an explicit discussion of signs in the context relevant here. The original reference is [54], and see also the excellent exposition in [10].

Let $\mathrm{k} = \bigoplus_v \mathbb{K} e_v$ be the commutative semisimple ring of orthogonal primitive idempotents over the base field $\mathbb{K}$, as before. Let V be a finite-dimensional graded $\mathbb{K}$–vector space with a k–bimodule structure. We write

$$T_{\mathrm{k}} V := \bigoplus_{i=0}^{\infty} V^{\otimes_{\mathrm{k}} i}$$

for the tensor algebra over k. A quadratic graded algebra A is an associative unital graded k–algebra that is a quotient

$$A := T_{\mathrm{k}} V / J$$

of $T_{\mathrm{k}} V$ by the two-sided ideal generated by a graded k–submodule $J \subset V \otimes_{\mathrm{k}} V$. In fact, this makes A into a bigraded algebra: it has an internal grading coming from the graded vector space V, denoted by s or $|x|$ if for a specific element, and a length grading coming from the tensor algebra, denoted by r. The reference [51] refers to s as Adams grading.

Let $V^\vee = \mathrm{Hom}_{\mathbb{K}}(V, \mathbb{K})$ be the linear dual of V viewed naturally as a k–bimodule, ie $e_i V^\vee e_j$ is the dual of $e_j V e_i$. Next, we consider the orthogonal dual $J^\perp \subset V^\vee \otimes_{\mathrm{k}} V^\vee$

with respect to the canonical pairing given by

$$V^\vee \otimes_{\mathrm{k}} V^\vee \otimes_{\mathrm{k}} V \otimes_{\mathrm{k}} V \to \mathrm{k}, \quad v_2^\vee \otimes_{\mathrm{k}} v_1^\vee \otimes_{\mathrm{k}} v_1 \otimes_{\mathrm{k}} v_2 \mapsto (-1)^{|v_2|} v_2^\vee(v_2) v_1^\vee(v_1).$$

The quadratic dual to A is defined as

$$A^! = T_{\mathrm{k}}(V^\vee[-1])/J^\perp[-2].$$

As does A, the graded quadratic algebra $A^!$ has two natural gradings: one internal grading coming from the internal grading of the vector space $V^\vee[-1]$, denoted by s or $|x^!|$ for a specific element, and the length grading coming from the tensor algebra, denoted by r.

The Koszul complex of a quadratic algebra is the graded right A–module $A^! \otimes_{\mathrm{k}} A$ with the differential[3]

$$x^! \otimes_{\mathrm{k}} x \to \sum_i (-1)^{|x|} x^! a_i^\vee \otimes_{\mathrm{k}} a_i x, \tag{6}$$

where the sum is over a basis of $\{a_i\}$ of V, and $\{a_i^\vee\}$ is the dual basis in $V^\vee[-1]$. This should be thought of as an (r,s)–bigraded complex, where the grading r is the path-length grading in the $A^!$ factor and the total grading $r+s$ corresponds to the natural grading $|x^!|+|x|$. In particular, one has $|a_i^\vee|+|a_i| = 1$ for all i, hence the s grading is preserved by the differential.

A Koszul algebra A is a quadratic algebra for which the Koszul complex is acyclic (ie homology is isomorphic to $\mathrm{k}[0]$). Taking the dual by applying the left exact functor $\mathrm{Hom}_A(\,\cdot\,, A)$, we get a resolution of k as a graded right A^{op}–module (see [10, Section 2] for more details).

In fact, if A is Koszul, considering k as a simple module in the abelian category of graded *right* A^{op}–modules, one has a canonical isomorphism of bigraded rings

$$A^! \cong \mathrm{Ext}^*_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k}).$$

Since A is bigraded, a priori $\mathrm{Ext}^*_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$ is triply graded (by the cohomological degree and by the length and internal gradings, derived from the corresponding ones in A). One characterization of Koszulity is that the cohomological degree, which we denote by r, agrees with the grading induced by length. Finally, we denote the internal grading by s. With this understood, we have the graded identifications

$$A^!_{r,s} \cong \mathrm{Ext}^r_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k}[s]).$$

[3] [10] prefers to use the graded left module $A \otimes_{\mathrm{k}} {}^\vee(A^!)$; the two graded modules are related by the right module isomorphism $A^! \otimes A \simeq \mathrm{Hom}_A(A \otimes_{\mathrm{k}} {}^\vee(A^!), A)$ and the sign $(-1)^{|x|}$ coming from this dualization.

If A is Koszul, then its Koszul dual $A^!$ is also Koszul and $(A^!)^! \cong A$.

Finally, for a Koszul algebra A, the Hochschild cohomology can be computed via the Koszul bimodule resolution of A. The resulting complex which computes Hochschild cohomology is

$$(A^! \otimes_{\mathrm{k}} A)_{\mathrm{diag}} = \bigoplus_v e_v A^! \otimes_{\mathrm{k}} A e_v \tag{7}$$

with the differential

$$x^! \otimes_{\mathrm{k}} x \to \sum_i (-1)^{|x|} x^! a_i^\vee \otimes_{\mathrm{k}} a_i x - (-1)^{(|a_i|+1)(|x|+|x^!|)} a_i^\vee x^! \otimes_k x a_i.$$

It is often the case, as in this paper, that V is generated either by odd elements or even elements; this simplifies the signs in the above formula. For Koszul algebras, the homology of this complex coincides with the bigraded Hochschild cohomology groups $\mathrm{HH}^r(A, A[s])$, where $r+s$ corresponds to the natural grading on $(A^! \otimes A)_{\mathrm{diag}}$, that is, an element $x^! \otimes_{\mathrm{k}} x$ has grading $|x^!| + |x|$. The length grading r corresponds to the path-length grading in the $A^!$ factor.

Example 20 Let $A_\Gamma = \mathbb{K}[x]/(x^2)$ with $|x| = 2$ be the zigzag algebra associated with a single vertex, ie Γ is of type A_1. It is easy to see that this is a Koszul algebra and we have $A_\Gamma^! = \mathbb{K}[x^\vee]$, the free algebra with $|x^\vee| = -1$. One can compute Hochschild cohomology using the Koszul bimodule complex. This has generators $(x^\vee)^i \otimes 1$ and $(x^\vee)^i \otimes x$ for $i \geq 0$. The differential can be computed as

$$\begin{aligned} d((x^\vee)^i \otimes 1) &= (1 + (-1)^{i+1})(x^\vee)^{i+1} \otimes x, \\ d((x^\vee)^i \otimes x) &= 0. \end{aligned}$$

Therefore, whenever $\operatorname{char} \mathbb{K} = 2$, the differential vanishes, and as a consequence $\mathrm{HH}^*(A_\Gamma)$ has a basis $(x^\vee)^i \otimes 1$, for $i \geq 0$, in bigrading $(r,s) = (i, -2i)$ and $(x^\vee)^i \otimes x$, for $i \geq 0$, in bigrading $(r,s) = (i, 2-2i)$.

If $\operatorname{char} \mathbb{K} \neq 2$, then $\mathrm{HH}^*(A_\Gamma)$ has a basis $(x^\vee)^{2i} \otimes 1$, for $i \geq 0$, in bigrading $(r,s) = (2i, -4i)$ and $(x^\vee)^{2i+1} \otimes x$, for $i \geq 0$, in bigrading $(r,s) = (2i+1, -4i)$ and $1 \otimes x$ in bigrading $(0,2)$.

In view of the discussion in the introduction, this result computes $\mathrm{SH}^*(T^*S^2)$ for $* = r+s$. For convenient access, we record a finite portion of this calculation in Table 1.

By Viterbo's isomorphism [70; 5], this computation also gives $H_{2-*}(\mathcal{L}S^2)$, where $\mathcal{L}S^2$ is the free loop space of S^2. This was previously computed as a ring by Cohen, Jones and Yan [21] over $\mathbb{Z}$ to be

$$H_{2-*}(\mathcal{L}S^2;\mathbb{Z}) \cong (\Lambda b \otimes \mathbb{Z}[a,v])/(a^2,ab,2av), \quad |a|=2,\ |b|=1,\ |v|=-2$$

using the fibration $\Omega_x S^2 \to \mathcal{L}S^2 \to S^2$. From this, one can deduce that

$$H_{2-*}(\mathcal{L}S^2;\mathbb{K}) \cong \Lambda a \otimes \mathbb{K}[u], \quad |a|=2,\ |u|=-1$$

if $\operatorname{char}\mathbb{K}=2$, and

$$H_{2-*}(\mathcal{L}S^2;\mathbb{K}) \cong (\Lambda b \otimes \mathbb{K}[a,v])/(a^2,ab,av), \quad |a|=2,\ |b|=1,\ |v|=-2r$$

if $\operatorname{char}\mathbb{K}\neq 2$, in agreement with our computation.

$r+s\downarrow \quad s\rightarrow$	2	1	0	−1	−2	−3	−4	−5	−6	−7	−8
2	1	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	0	0	0	0	0	0	0
0	0	0	1	0	x	0	0	0	0	0	0
−1	0	0	0	0	x	0	1	0	0	0	0
−2	0	0	0	0	0	0	1	0	x	0	0

Table 1: $\Gamma = A_1$; x is 1 if $\operatorname{char}\mathbb{K}=2$, 0 otherwise

5.2 Koszul duality for A_∞–algebras

We now review Koszul duality for A_∞–algebras. Our primary reference for this material is [51]. The discussion in [51] is about A_∞–algebras over a field $\mathbb{K}$, but as in classical Koszul duality, the proofs extend readily to A_∞–algebras over a semisimple ring k (see also [58]). The extension of Koszul duality theory to DG- or A_∞–algebras has appeared earlier (see eg [46]).

Suppose $A=\bigoplus_{i\geq 0} A_i$ is a positively graded associative algebra over $A_0=\mathrm{k}$. Then, as before, the complex

$$\mathrm{RHom}_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$$

inherits a bigrading by cohomological and length gradings. However, it usually happens that at the level of homology these two gradings do not agree, that is, A is not Koszul as an associative algebra, and passing to the homology of this complex yields an associative algebra $\mathrm{Ext}^*_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$ from which one cannot recover A. In this case, the idea is that rather than passing to homology, one thinks of the DG-algebra $\mathrm{RHom}_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$ as the DG-Koszul dual of A. To be able to carry this out, one is led to work with DG- or A_∞–algebras from the beginning. So, let $\mathscr{A}$ be a $\mathbb{Z}$–graded A_∞–algebra over k

together with an augmentation $\epsilon\colon \mathscr{A} \to \mathrm{k}$, making k into a right A_∞–module over $\mathscr{A}^{\mathrm{op}}$. One defines

$$\mathscr{A}^! = \mathrm{RHom}_{\mathscr{A}^{\mathrm{op}}}(\mathrm{k}, \mathrm{k}).$$

Note that the Yoneda image of k given by $\mathrm{RHom}_{\mathscr{A}^{\mathrm{op}}}(\mathscr{A}^{\mathrm{op}}, \mathrm{k})$ makes k into a right $(\mathscr{A}^!)^{\mathrm{op}}$–module. Now, the obvious concern is whether $(\mathscr{A}^!)^!$ gets back to $\mathscr{A}$ (up to quasi-isomorphism). This is not quite the case in general; one recovers a certain completion of $\mathscr{A}$ (see [58] for a beautiful geometric description of this construction). However, suppose that $\mathscr{A}$ has an additional s grading (called Adams grading in [51]) which is required to be preserved by the A_∞–operations. Furthermore, assume that $\mathscr{A}$ is connected and locally finite with respect to this grading; this means that $\mathscr{A}$ is either nonnegatively or nonpositively graded and the s–homogeneous subspace of $\mathscr{A}$ is of finite dimension for each s (see [51, Definition 2.1]). Then it is true that $(\mathscr{A}^!)^!$ is quasi-isomorphic to $\mathscr{A}$. We state this as:

Theorem 21 (Lu, Palmieri, Wu and Zhang [51, Theorem 2.4][4]) *Suppose $\mathscr{A}$ is an augmented A_∞–algebra over the semisimple ring* k *with a bigrading for which μ^k has degree $(2-k, 0)$ and suppose $\mathscr{A}$ is connected and locally finite with respect to the second grading. Let*

$$\mathscr{A}^! = \mathrm{RHom}_{\mathscr{A}^{\mathrm{op}}}(\mathrm{k}, \mathrm{k})$$

be its Koszul dual as an A_∞–algebra. Then there is a quasi-isomorphism of A_∞–algebras

$$\mathscr{A} \simeq \mathrm{RHom}_{(\mathscr{A}^!)^{\mathrm{op}}}(\mathrm{k}, \mathrm{k}).$$

Below, we will apply this result for $\mathscr{A} = A_\Gamma$ viewed as a formal A_∞–algebra.

Example 22 To see the importance of the connectedness and finiteness assumptions, let us consider $A = \mathbb{K}[x, x^{-1}]$ with x in bigrading $(0,0)$, the (trivially graded) algebra of Laurent polynomials. Consider the augmentation $\epsilon\colon A^{\mathrm{op}} \to \mathbb{K}$ given by mapping x to $1 \in \mathbb{K}$, which makes $\mathbb{K}$ into a right A–module. Then one can check that $A^! = \mathrm{RHom}_{A^{\mathrm{op}}}(\mathbb{K}, \mathbb{K})$ is quasi-isomorphic to the exterior algebra $\mathbb{K}[x^!]/((x^!)^2)$ with $x^!$ in bigrading $(0,1)$. However, $\mathrm{RHom}_{(A^!)^{\mathrm{op}}}(\mathbb{K}, \mathbb{K}) \cong \mathbb{K}[\![y]\!]$ gives the power series ring with y in bigrading $(0,0)$. Hence, dualizing twice does not get us back in this case.

[4] The proof of [51, Theorem 2.4] uses [51, Lemma 1.15] which omits a necessary hypothesis. Namely, in the notation of [51, Lemma 1.15], one should further assume $B^\infty_{\mathrm{aug}} R$ is locally finite. By [51, Lemma 2.2], this requirement holds under our hypothesis.

5.3 Koszul dual of $\mathscr{G}_\Gamma$

We next prove that the DG-algebras $\mathscr{G}_\Gamma$ and A_Γ (viewed as a formal A_∞–algebra) are related by Koszul duality. We remind the reader that we always work with right modules (as we follow the sign conventions from [61]).

We have the following analogue of [39, Proposition 2.9.5] in our setting:

Theorem 23 *Consider* $\mathrm{k} = A_\Gamma^{\mathrm{op}}/(A_\Gamma^{\mathrm{op}})_{>0}$ *as a right* A_Γ^{op}*–module. There is a DG-algebra isomorphism*

$$\mathrm{RHom}_{A_\Gamma^{\mathrm{op}}}(\mathrm{k},\mathrm{k}) \simeq \mathscr{G}_{\Gamma^{\mathrm{op}}}$$

such that the cohomological (resp. internal) grading on the left-hand side agrees with the path-length (resp. internal) grading on the right-hand side.

Proof First, let us clarify the multiplication on A_Γ^{op}, which we view as a formal A_∞–algebra. We identify the elements of A_Γ^{op} with the elements of A_Γ which are given by the symbols a_{vw} and $a_{vw}a_{wv}$ as before. Since $|a_{wv}| = 1$ for all w adjacent to v, the product is given by

$$\mu^2_{A_\Gamma^{\mathrm{op}}}(a_{wv}, a_{vw}) = (-1)^{|a_{wv}|+|a_{vw}|}\mu^2_{A_\Gamma}(a_{vw}, a_{wv}) = (-1)^{|a_{vw}|}a_{vw}a_{wv} = -a_{vw}a_{wv}$$

for w adjacent to v (see [61, Section (1a)] for signs used in defining the opposite of an A_∞–algebra).

We use the reduced bar resolution of k as a right A_Γ^{op}–module to calculate $\mathrm{RHom}_{A_\Gamma^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$, which takes the form

$$\mathrm{RHom}_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k}) \simeq \hom_{A^{\mathrm{op}}}((A \otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}}, \mathrm{k}),$$

where $A = A_\Gamma$, $\bar{A} = A_\Gamma/\mathrm{k}$, and $T\bar{A}$ is the tensor algebra of $\bar{A}_\Gamma$ over k.

The fact that $k = A_0$ allows us to identify $\bar{A}$ with the positive graded subalgebra $A_1 \oplus A_2$ of A. We follow the conventions in [61, Section (1j)] for the DG-algebra structure of $\hom_{A^{\mathrm{op}}}((A \otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}}, \mathrm{k})$. However, we view $\hom_{A^{\mathrm{op}}}((A \otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}}, \mathrm{k})$ as a DG-algebra rather than an A_∞–algebra with $\mu^k = 0$ for $k > 2$ since $\mathscr{G}_\Gamma$ is always viewed as a DG-algebra. The difference is in the signs, and this was explained in the introduction (see (3)).

More precisely, a generator $t \in \hom_{A^{\mathrm{op}}}((A \otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}}, \mathrm{k})$ of bidegree (r,s) is an A^{op}–module homomorphism $t\colon A \otimes_{\mathrm{k}} \bar{A}^{\otimes r} \to \mathrm{k}$ of internal degree $|t| = s$. Observe that any such t maps an element $(a_{r+1}, a_r, \dots, a_1)$ to 0 unless $a_{r+1} \in A_0$ because of the A^{op}–module structure of k.

The differential and the product on the DG-algebra $\hom_{A^{\mathrm{op}}}((A\otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}},\mathrm{k})$ are defined by

$$(dt)(e_v,a_{r+1},\dots,a_1) = \sum_{n=1}^{r}(-1)^{\dagger+|t|}t(e_v,a_{r+1},\dots a_{n+2},\mu^2_{A^{\mathrm{op}}}(a_{n+1},a_n),a_{n-1},\dots,a_1)$$

and if t_1 and t_2 are two generators of lengths r_1,r_2, then

$$(t_2\cdot t_1)(e_v,a_{r_2+r_1},\dots,a_1) = (-1)^{\ddagger+|t_1|}t_2(t_1(e_v,a_{r_2+r_1},\dots,a_{r_2+1}),a_{r_2},\dots,a_1),$$

where $\dagger=\sum_{i=n}^{r+1}(|a_i|-1)$ and $\ddagger=\sum_{i=r_2+1}^{r_2+r_1}(|a_i|-1)$.

We now construct a chain map

$$\Phi\colon \mathscr{G}_{\Gamma^{\mathrm{op}}}\to \hom_{A^{\mathrm{op}}}((A\otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}},\mathrm{k})$$

that respects the cohomological and internal gradings, first by defining it on the generators g_{wv} and h_v of the underlying tensor algebra of $\mathscr{G}_{\Gamma^{\mathrm{op}}}$, and then extending by mapping the product p_2p_1 of two elements p_2 and p_1 in $\mathscr{G}_{\Gamma^{\mathrm{op}}}$ to the homomorphism $\Phi(p_2)\cdot\Phi(p_1)\in\hom_{A^{\mathrm{op}}}((A\otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}},\mathrm{k})$.

Indeed, let us define $\Phi(g_{wv})$ and $\Phi(h_v)$ to be A–module homomorphisms each of which is nonzero only on a 1–dimensional subspace of $A\otimes_{\mathrm{k}} T\bar{A}$, given by

$$\Phi(g_{wv})\colon (e_v,a_{wv})\mapsto \epsilon_{wv}e_w \quad\text{and}\quad \Phi(h_v)\colon (e_v,a_{vw}a_{wv})\mapsto \epsilon_v e_v,$$

for any vertex w adjacent to v in Γ. Here the signs ϵ_{wv}, ϵ_v are determined as follows. For a vertex $v\in\Gamma_0$, we set $\epsilon_v=(-1)^{\delta_v}$, where δ_v is the distance from the root of Γ to the vertex v. If g_{wv} is an arrow in the quiver Γ^{op}, then define $\epsilon_{wv}=\epsilon_v$ and $\epsilon_{vw}=+1$. Note that $\epsilon_{wv}\epsilon_{vw}/\epsilon_v$ is $+1$ if and only if g_{wv} is an arrow in the quiver Γ^{op}.

Observe that the internal gradings are

$$|\Phi(g_{wv})|=-|a_{wv}|=-1 \quad\text{and}\quad |\Phi(h_v)|=-|a_{vw}a_{wv}|=-2,$$

respectively. Note also that Φ takes the path-length grading on $\mathscr{G}_\Gamma$ to the cohomological grading on $\hom_{A^{\mathrm{op}}}((A\otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}},\mathrm{k})$, hence Φ respects the bigraded structure of both sides.

The differentials on the DG-algebras $\mathscr{G}_{\Gamma^{\mathrm{op}}}$ and $\hom_{A^{\mathrm{op}}}((A\otimes_{\mathrm{k}} T\bar{A})^{\mathrm{op}},\mathrm{k})$ obey the graded Leibniz rule, hence it suffices to check that

$$d(\Phi(g_{wv}))=\Phi(dg_{wv})=0 \quad\text{and}\quad d(\Phi(h_v))=\Phi(dh_v)$$

to verify that Φ is a DG-algebra homomorphism.

The first identity follows immediately since both g_{wv} and $\Phi(g_{wv})$ are in total degree 0 and the differential vanishes here. To check the second identity, observe that $d(\Phi(h_v))$ is nonzero only on the subspace of $A \otimes_{\mathrm{k}} T\bar{A}$ spanned by

$$\{(e_v, a_{wv}, a_{vw}) : w \text{ is adjacent to } v\},$$

and for every w adjacent to v,

$$\begin{aligned}(d(\Phi(h_v)))(e_v, a_{wv}, a_{vw}) &= (-1)^{|\Phi(h_v)|+(|a_{wv}|-1)+(|a_{vw}|-1)}\Phi(h_v)(e_v, -a_{vw}a_{wv}) \\ &= -\epsilon_v e_v.\end{aligned}$$

Note that the appearance of the extra sign here is precisely the point where the use of A_Γ^{op} rather than A_Γ takes effect.

On the other hand,

$$\Phi(dh_v) = \Phi\left(\sum_w \frac{\epsilon_{wv}\epsilon_{vw}}{\epsilon_v} g_{vw} g_{wv}\right) = \sum_w \frac{\epsilon_{wv}\epsilon_{vw}}{\epsilon_v} \Phi(g_{vw}) \cdot \Phi(g_{wv}).$$

For each w adjacent to v, $\Phi(g_{vw}) \cdot \Phi(g_{wv})$ is nonzero only on the subspace spanned by (e_v, a_{wv}, a_{vw}), and

$$\begin{aligned}&(\Phi(g_{vw}) \cdot \Phi(g_{wv}))(e_v, a_{wv}, a_{vw}) \\ &\qquad = (-1)^{|\Phi(g_{wv})|+(|a_{wv}|-1)}\Phi(g_{vw})((\Phi(g_{wv})(e_v, a_{wv})), a_{vw}) \\ &\qquad = -\epsilon_{wv}\epsilon_{vw} e_v.\end{aligned}$$

Indeed, we also have an extra sign here, and hence the second identity holds.

To prove the bijectivity of Φ, consider a generator $(e_v, a_r, \dots, a_1)$ of $A \otimes_{\mathrm{k}} \bar{A}^{\otimes r}$. Note that such a generator is uniquely determined by the initial and terminal points of a_i considered as paths in A_Γ which in turn determine a unique path $g_r \cdots g_1$ of length r in $\mathscr{G}_\Gamma$, so that the initial and terminal points of each arrow g_i in the extended quiver $\widehat{\Gamma}$ match those of a_{r+1-i}. It is straightforward to check that

$$(\Phi(g_r \cdots g_1))(e_v, a_r, \dots, a_1) = \pm e_w,$$

where w is the terminal point of a_1. This proves that Φ is injective since the algebra underlying $\mathscr{G}_\Gamma$ is the path algebra generated by the arrows in $\widehat{\Gamma}$. Moreover, the observation that $\Phi(g_r \cdots g_1)$ is nonzero only on the subspace of $A \otimes_{\mathrm{k}} T\bar{A}$ spanned by $(e_v, a_r, \dots, a_1)$ shows that Φ is surjective as well. □

Remark 24 As can be seen from the proof of Theorem 23, we could arrange the definition of the DG-algebra isomorphism Φ so as to obtain an isomorphism

$$\mathrm{RHom}_{A_\Gamma}(\mathrm{k}, \mathrm{k}) \simeq \mathscr{G}_\Gamma,$$

where $\mathrm{k} = A_\Gamma/(A_\Gamma)_{>0}$ is viewed as a right A_Γ–module. This is because there happens to be an isomorphism of algebras between A_Γ and A_Γ^{op}. We have opted to use A_Γ^{op} to be consistent with the general framework of Koszul duality (see [10, Theorem 2.10.1]).

The following corollary is immediate from Theorem 23 and Theorem 21:

Corollary 25 *Consider $\mathrm{k} = \mathscr{G}_\Gamma/(\mathscr{G}_\Gamma)_{r>0}$ as a right $\mathscr{G}_\Gamma$–module, and A_Γ as a DG-algebra with trivial differential. There is a quasi-isomorphism of DG-algebras*

$$\mathrm{RHom}_{\mathscr{G}_\Gamma}(\mathrm{k},\mathrm{k}) \simeq A_\Gamma$$

such that the cohomological and internal gradings on the left-hand side coincide with each other and they agree with the path-length grading on the right-hand side.

Proof In view of Theorem 23 and Theorem 21, we only need to check the hypothesis in Theorem 21, but this is straightforward. Certainly, A_Γ is positively graded and the local finiteness condition holds since A_Γ is finite-dimensional (see [51, Definition 2.1]). □

Since A_Γ is known to be Koszul in the classical sense for non-Dynkin Γ, we easily get an alternative proof of the formality result mentioned in Theorem 7(1).

Corollary 26 *For Γ non-Dynkin, $\mathscr{G}_\Gamma$ is formal, that is, it is quasi-isomorphic to the preprojective algebra $\Pi_\Gamma = H^0(\mathscr{G}_\Gamma)$.*

Proof Recall that the differential on the complex $\mathrm{RHom}_{A_\Gamma^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$ has bidegree $(1,0)$. Therefore, after applying the homological perturbation lemma, we obtain a minimal A_∞–structure on $\mathrm{Ext}^*_{A^{\mathrm{op}}}(\mathrm{k},\mathrm{k})$ such that μ^d has bidegree $(2-d,0)$. On the other hand, Koszulity of A_Γ means that the two gradings agree at the level of cohomology. Therefore, it is impossible to have a nontrivial μ^d for $d \neq 2$. □

Note that if Γ is a Dynkin-type graph, $\mathscr{G}_\Gamma$ is not quasi-isomorphic to the preprojective algebra Π_Γ. Our result above can be described as stating that $\mathscr{G}_\Gamma$ and A_Γ are A_∞–Koszul dual. This should be seen as the natural extension to all Γ of the classical Koszul duality between Π_Γ and A_Γ which only worked when Γ is non-Dynkin.

Finally, in view of the Theorem 23 and Corollary 25, we conclude from Keller's theorem [47] that there is an isomorphism of Hochschild cohomologies as Gerstenhaber algebras. Besides this isomorphism, the following theorem also uses the fact that $\mathrm{HH}_{2-*}(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(\mathscr{G}_\Gamma)$ by the Calabi–Yau property [39], together with [17] which applies over $\mathbb{K}$ of characteristic 0, and Theorem 13.

Theorem 27 *For any tree Γ, there is an isomorphism of Gerstenhaber algebras over $\mathbb{K}$*

$$\mathrm{HH}^*(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(A_\Gamma).$$

If Γ is of Dynkin type A_n or D_n (and conjecturally also for E_6, E_7, E_8) and $\mathbb{K}$ is of characteristic 0, then we have

$$\mathrm{SH}^*(X_\Gamma) \cong \mathrm{HH}^*(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(A_\Gamma).$$

Remark 28 Note that all of the Gerstenhaber algebras appearing in the above theorem are induced from a natural underlying Batalin–Vilkovisky (BV) algebra structure. In the case of symplectic cohomology, BV-algebra structure is given by a geometric construction reminiscent of the loop rotation in string topology and in the cases of $\mathscr{G}_\Gamma$ and A_Γ, it is induced by the underlying Calabi–Yau structure on these DG-algebras, which allows one to dualize the Connes differential B on Hochschild homology. However, the above theorem does not claim an isomorphism of the underlying Batalin–Vilkovisky structures. We believe that this can be achieved, however, it requires a finer investigation of Calabi–Yau structures. On the other hand, we explain in Remark 33 that for Γ non-Dynkin and non-extended Dynkin, we have an isomorphism of Batalin–Vilkovisky algebras between $\mathrm{HH}^*(\mathscr{G}_\Gamma)$ and $\mathrm{HH}^*(A_\Gamma)$ as it turns out that there is a unique way of equipping this Gerstenhaber algebra with a BV-algebra structure.

Remark 29 It is well-known that in the case when Γ is Dynkin, the exact Lagrangian spheres S_v split-generate the Fukaya category $\mathcal{F}(X_\Gamma)$ of compact exact Lagrangians — this follows for example by combining [59, Lemma 4.15] and [61, Corollary 5.8]. Furthermore, as mentioned in the beginning of Section 5, one expects that the noncompact Lagrangians L_v split-generate the wrapped Fukaya category. Hence, one could interpret the above result as showing that

$$\mathrm{HH}^*(\mathcal{F}(X_\Gamma)) \cong \mathrm{HH}^*(\mathcal{W}(X_\Gamma)).$$

On the other hand, it is by no means the case that $D^\pi\mathcal{F}(X_\Gamma)$ and $D^\pi\mathcal{W}(X_\Gamma)$ are equivalent as triangulated categories. (Here, we mean an equivalence between the Karoubi-completed triangulated closures of $\mathcal{F}(X_\Gamma)$ and $\mathcal{W}(X_\Gamma)$.) This is clear from the fact that the latter category has objects with infinite-dimensional endomorphisms (over $\mathbb{K}$) but every object in the former has finite-dimensional endomorphisms. More strikingly, the monotone Lagrangian tori studied in [49] give objects in $D^\pi\mathcal{W}(X_\Gamma)$ for $\Gamma = A_n$ with finite-dimensional endomorphisms and yet these do not belong to the category $D^\pi\mathcal{F}(X_\Gamma)$. One has to collapse the grading to $\mathbb{Z}_2$ in order to admit these objects in $\mathcal{F}(X_\Gamma)$.

In the next section, we compute $\mathrm{HH}^*(A_\Gamma)$ for all trees Γ except E_6, E_7, E_8.

6 Hochschild cohomology computations

6.1 Non-Dynkin case

In this section we assume that Γ is a non-Dynkin tree and describe the Hochschild cohomology $\mathrm{HH}^*(\mathscr{G}_\Gamma)$ of the associated Ginzburg DG-algebra. Note, however, that as explained in the introduction, when Γ is non-Dynkin, $\mathscr{B}_\Gamma$ is a nontrivial deformation of $\mathscr{G}_\Gamma$, and so this computation does not directly give enough information to compute $\mathrm{HH}^*(\mathscr{B}_\Gamma)$, and thus $\mathrm{SH}^*(X_\Gamma)$. However, at least away from characteristic 0, the computation of $\mathrm{HH}^*(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(A_\Gamma)$ is still of geometric significance as it controls the deformations of the compact Fukaya category $\mathcal{F}(X_\Gamma)$.

Recall that for non-Dynkin Γ, the cohomology $H^*(\mathscr{G}_\Gamma) \cong \Pi_\Gamma$ is supported in total degree 0 and moreover $\mathscr{G}_\Gamma$ is formal, ie it is quasi-isomorphic to the preprojective algebra Π_Γ. Therefore we have an isomorphism of Gerstenhaber algebras

$$\mathrm{HH}^*(\mathscr{G}_\Gamma) \cong \mathrm{HH}^*(\Pi_\Gamma),$$

where Π_Γ is to be considered as a trivially graded algebra. For any non-Dynkin quiver Γ, the Gerstenhaber structure of the Hochschild cohomology of $\Pi = \Pi_\Gamma$ is described in [57] (and previously in [23] when $\operatorname{char} \mathbb{K} = 0$). We do not have anything new to say here, we simply review some of the results of [23] and [57] to give a flavor of what's known. For an impressive amount of further information, see the comprehensive work of Schedler [57].

The Hochschild cohomology $\mathrm{HH}^*(\Pi_\Gamma)$ turns out to be trivial in every grading except for $0, 1$ and 2. A way to see this is to use the Koszul bimodule resolution given in (7). Recall that for Γ non-Dynkin, Π_Γ is Koszul in the classical sense with Koszul dual $A = A_\Gamma$. The latter has a decomposition into its graded pieces as $A = A_0 \oplus A_1 \oplus A_2$. Hence, the Koszul bimodule resolution takes the form

$$0 \to \bigoplus_v e_v \Pi e_v \to \bigoplus_v e_v A_1 \otimes_{\mathrm{k}} \Pi e_v \to \bigoplus_v e_v A_2 \otimes_{\mathrm{k}} \Pi e_v \to 0.$$

Moreover, it is well known that Π is Calabi–Yau of dimension 2 (see [39, Definition 3.2.3]), hence a duality result of Van den Bergh [11] applies and we have a canonical isomorphism

$$\mathrm{HH}^*(\Pi) \cong \mathrm{HH}_{2-*}(\Pi).$$

For the $\mathbb{K}$–vector space structure of the Hochschild cohomology let us recall some general facts (see eg [50]) which apply to any algebra (with trivial grading and differential). The zeroth cohomology $\mathrm{HH}^0(\Pi)$ is given by the center $Z(\Pi)$, and $\mathrm{HH}^1(\Pi)$ is given by *outer derivations* $\mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi)$. Recall that a derivation is a linear map

$D\colon \Pi \to \Pi$ satisfying the Leibniz rule, and each $a \in \Pi$ defines an inner derivation by $D_a(b) = ab - ba$. The zeroth homology $\mathrm{HH}_0(\Pi)$ is isomorphic to $\Pi_{\mathrm{cyc}} := \Pi/[\Pi, \Pi]$, where $[\Pi, \Pi] \subset \Pi$ is the $\mathbb{K}$–linear subspace spanned by the commutators.

Theorem 30 [56, Corollary 10.1.2; cf 23, Theorem 8.4.1] *The $\mathbb{K}$–vector space structure of the Hochschild cohomology $\mathrm{HH}^*(\Pi)$ of the preprojective algebra associated to a non-Dynkin quiver is as follows.*

(1) *If Γ is extended Dynkin, then $\mathrm{HH}^0(\Pi) \cong Z(\Pi) \cong e_{v_0}\Pi e_{v_0}$, where v_0 is a vertex in Γ whose complement is Dynkin. Otherwise the center $Z(\Pi)$ is isomorphic to $\mathbb{K}$.*

(2) $\mathrm{HH}^1(\Pi) \cong \mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi) \cong Z(\Pi) \oplus (F \otimes_{\mathbb{Z}} \mathbb{K}) \oplus \big(T \otimes_{\mathbb{Z}} \bigoplus_p \mathrm{Hom}_{\mathbb{Z}}(\mathbb{F}_p, \mathbb{K})\big)$, *where F and T are the free and torsion parts of $\Pi^{\mathbb{Z}}_{\mathrm{cyc}}$, respectively, and $\Pi^{\mathbb{Z}}$ is the preprojective $\mathbb{Z}$–algebra associated to Γ.*

(3) $\mathrm{HH}^2(\Pi) \cong \mathrm{HH}_0(\Pi) \cong \Pi_{\mathrm{cyc}}$.

Remark 31 In the extended Dynkin case, by the McKay correspondence $Z(\Pi)$ is isomorphic to the ring of invariant polynomials in $\mathbb{K}[x, y]$ under the action of the corresponding finite subgroup $G \subset \mathrm{SL}_2(\mathbb{K})$ as long as $\mathbb{K}$ has $|G|^{\mathrm{th}}$ roots of unity (see [56, Theorem 9.1.1]). Furthermore, in this case T is trivial and hence $\mathrm{HH}^*(\Pi)$ is determined by $Z(\Pi)$ and Π_{cyc}, unless the characteristic of $\mathbb{K}$ is a "bad prime" for Γ, ie 2 for $\widetilde{D}_n$, 2 or 3 for $\widetilde{E}_6$ and $\widetilde{E}_7$, and $2, 3$ or 5 for $\widetilde{E}_8$ [57]. Note that the Hilbert series of $Z(\Pi)$ and Π_{cyc}, as algebras graded by path-length, are given in [34] and [57].

The quotient Π_{cyc} can be considered as a graded Lie algebra with the path-length grading and the Lie bracket induced by the *necklace Lie bracket* $\{\cdot,\cdot\}$ on Π, given by

$$\{p, q\} = \sum_{g_{wv} \in \Gamma_1} (\partial_{vw} q)(\partial_{wv} p) - (\partial_{wv} q)(\partial_{vw} p).$$

Here, for any path $p \in \Pi$ and adjoint pair (v, w) in Γ, $\partial_{wv} p$ is given as the sum

$$\sum\nolimits_i g_{i-1} \cdots g_1 g_l \cdots g_{i+1},$$

taken over all i for which the i^{th} arrow g_i in the path $p = g_l \cdots g_1$ is g_{wv}.

Note that the Lie bracket $[D, D'] = D \circ D' - D' \circ D$ on $\mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi)$ coincides with the Gerstenhaber bracket on $\mathrm{HH}^1(\Pi)$ in favorable cases, eg if $\operatorname{char}\mathbb{K} = 0$ and Γ is not extended Dynkin.

The Lie brackets above are used to describe the (cup) product as well as the Gerstenhaber bracket on $\mathrm{HH}^*(\Pi)$ in [23], when $\operatorname{char}\mathbb{K} = 0$. We now recall the description of the

Gerstenhaber algebra structure of $\mathrm{HH}^*(\Pi)$ in [57], for arbitrary $\operatorname{char}\mathbb{K}$, using the BV operator Δ dual to the Connes differential (see eg [50]) on $\mathrm{HH}_*(\Pi)$. The *Euler derivation* eu on Π_{cyc} is defined as multiplication by l on each path of length l, and the derivation u, called *half Euler derivation* in [57], multiplies each path by the number of edges from Γ that it contains. Note that we have $\mathrm{eu} = 2u$ as elements of $\mathrm{HH}^1(\Pi)$. In other words, their difference is an inner derivation. The first summand of $\mathrm{HH}^1(\Pi)$ in Theorem 30 consists of multiples of u by $Z(\Pi)$.

Theorem 32 [56, Theorem 10.3.1] *As a BV-algebra, $\mathrm{HH}^*(\Pi)$ is determined by the following properties.*

(1) *The graded-commutative product*

$$\cup\colon \mathrm{HH}^i(\Pi)\otimes \mathrm{HH}^j(\Pi)\to \mathrm{HH}^{i+j}(\Pi)$$

is given as follows:

(a) *If $\theta,\theta'\in \mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi)\cong \mathrm{HH}^1(\Pi)$ and θ' belongs to the $F\otimes_{\mathbb{Z}}\mathbb{K}$ summand of $\mathrm{HH}^1(\Pi)$, then $\theta\cup\theta'$ is obtained by considering θ' as an element of Π_{cyc} and applying the derivation θ to it.*

(b) *If none of $\theta,\theta'\in \mathrm{HH}^1(\Pi)$ belongs to the $F\otimes_{\mathbb{Z}}\mathbb{K}$ summand, then $\theta\cup\theta'=0$.*

(c) *If $ij=0$, then $\cup$ is given by multiplication in Π.*

(2) *The BV-operator*

$$\Delta\colon \mathrm{HH}^i(\Pi)\to \mathrm{HH}^{i-1}(\Pi)$$

dual to the Connes differential is given as follows.

(a) *We have*

$$\Delta(u)=1,\qquad \Delta(z\cup\theta)=\theta(z)+z\Delta(\theta)$$

for every $z\in \mathrm{HH}^0(\Pi)\cong Z(\Pi)$, $\theta\in \mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi)\cong \mathrm{HH}^1(\Pi)$. The BV-operator vanishes on the $\big(T\otimes_{\mathbb{Z}}\bigoplus_p \mathrm{Hom}_{\mathbb{Z}}(\mathbb{F}_p,\mathbb{K})\big)$ summand of $\mathrm{HH}^1(\Pi)$.

(b) *The operator $\Delta\colon \mathrm{HH}^2(\Pi)\cong \Pi_{\mathrm{cyc}}\to \mathrm{Der}(\Pi)/\mathrm{Inn}(\Pi)\cong \mathrm{HH}^1(\Pi)$ maps to the $F\otimes_{\mathbb{Z}}\mathbb{K}$ summand and it is given by*

$$\Delta(g_l\cdots g_1)=\sum\nolimits_{i=1}^{l}\pm\partial_{g_i^*}(\cdot)g_{i-1}\cdots g_1 g_l\cdots g_{i+1},$$

where each g_i is an arrow in the double of the quiver Γ and the sign is positive if and only if $g_i\in\Gamma$.

Remark 33 A word of caution is in order. For Γ non-Dynkin, the BV-algebra structure on $\mathrm{HH}^*(\Pi_\Gamma)$ is induced by the 2–Calabi–Yau structure (in the sense of Ginzburg [39], also known as smooth Calabi–Yau structure) on the homologically smooth algebra Π_Γ. This means that we have an isomorphism of Π_Γ–bimodules

$$\Pi_\Gamma \simeq \mathrm{RHom}_{\Pi_\Gamma - \Pi_\Gamma}(\Pi_\Gamma, \Pi_\Gamma \otimes \Pi_\Gamma)[2],$$

where the bimodule structure on the right is with respect to the inner bimodule structure on $\Pi_\Gamma \otimes \Pi_\Gamma$ and RHom is taken with respect to the outer bimodule structure on $\Pi_\Gamma \otimes \Pi_\Gamma$. Two such 2–Calabi–Yau structures differ by an invertible element in $\mathrm{HH}^0(\Pi_\Gamma)$. The effect by such an invertible ϕ is to replace Δ by $\phi^{-1}\Delta\phi$ [66, Remark 4.8].

We can consider the Koszul dual notion. Namely, by Koszul duality, for Γ non-Dynkin, we have $\mathrm{HH}^*(\Pi_\Gamma) \cong \mathrm{HH}^*(A_\Gamma)$ and then the BV-algebra structure can be seen as naturally arising from a weak Calabi–Yau structure on A_Γ. Recall that a weak Calabi–Yau structure (also known as Frobenius structure or compact Calabi–Yau structure) of dimension 2 on the finite-dimensional algebra A_Γ is a quasi-isomorphism of A_Γ–bimodules

$$A_\Gamma \simeq A_\Gamma^\vee[-2],$$

where $A_\Gamma^\vee$ is the $\mathbb{K}$–linear dual of A_Γ. Two such Calabi–Yau structures again differ by an invertible element in $\mathrm{HH}^0(A_\Gamma)$.

In any case, if Γ is non-Dynkin and non-extended Dynkin, then by Theorem 30, $\mathrm{HH}^0(\Pi_\Gamma) \cong \mathrm{HH}^0(A_\Gamma) \cong \mathbb{K}$ is rank-1 generated by the identity, hence there exists (up to scaling) at most one (Ginzburg) Calabi–Yau structure on Π_Γ and at most one (weak) Calabi–Yau structure on A_Γ. These Calabi–Yau structures can either be constructed algebraically as in [39] or symplectically as a manifestation of Poincaré duality for the Fukaya category of compact Lagrangians or the open Calabi–Yau property of the wrapped Fukaya category.

Now, *suppose* $\mathscr{B}_\Gamma \simeq \mathscr{G}_\Gamma$. Then, since $\mathscr{G}_\Gamma$ is formal, we would have an isomorphism $\mathrm{SH}^*(X_\Gamma) \cong \mathrm{HH}^*(\mathscr{B}_\Gamma) \cong \mathrm{HH}^*(\Pi_\Gamma)$. Under this isomorphism, the natural BV-algebra structure on $\mathrm{SH}^*(X_\Gamma)$ given by the loop rotation operator $\Delta\colon \mathrm{SH}^*(X_\Gamma) \to \mathrm{SH}^{*-1}(X_\Gamma)$ has to coincide with the algebraically constructed BV-algebra structure on $\mathrm{HH}^*(\Pi_\Gamma)$ in the case that Γ is non-Dynkin and non-extended Dynkin.

On the other hand, combining the results from [53] and [5] one deduces that

$$\mathrm{SH}^*(T^*S^2) \cong \mathrm{HH}^*(C_{2-*}(\Omega S^2)) \cong \mathrm{HH}^*(C^*(S^2))$$

does not admit a *dilation* over a field of characteristic 2.[5] Recall that a dilation is an element $b \in \mathrm{SH}^1(X_\Gamma)$ such that

$$\Delta b = 1,$$

where $\Delta\colon \mathrm{SH}^*(X_\Gamma) \to \mathrm{SH}^{*-1}(X_\Gamma)$ is the BV-operator in symplectic cohomology. Furthermore, since T^*S^2 can be embedded as a Liouville subdomain of X_Γ, one has a restriction map, $\mathrm{SH}^*(X_\Gamma) \to \mathrm{SH}^*(T^*S^2)$ which is a map of BV-algebras. Therefore, a dilation on X_Γ can be restricted to a dilation on T^*S^2. On the other hand, we see from the above theorem that there is a class $u \in \mathrm{HH}^1(\Pi_\Gamma)$ that is sent to the identity by the BV-operator induced from the Calabi–Yau structure on Π_Γ. Hence, we arrive at a contradiction.

This is in agreement with Remark 15 where we have seen that $\mathscr{B}_\Gamma$ is a nontrivial deformation of $\mathscr{G}_\Gamma$ over a field of characteristic 2.

6.2 Dynkin case

In this section we compute the Hochschild cohomology of the zigzag algebra A_Γ associated with a *Dynkin* tree. If the underlying tree Γ is of type A_1, ie a single vertex, then $A_\Gamma = \mathbb{K}[x]/(x^2)$ with $|x| = 2$ and it is a Koszul algebra. Its Hochschild cohomology was computed in Example 20 above. Thus, hereafter we assume $\Gamma \neq A_1$. It turns out that if the underlying tree Γ is of Dynkin type but not a single vertex, then A_Γ is an almost-Koszul algebra (in the sense of [18]). In this situation, the Koszul complex leads to a construction of a minimal *periodic* resolution. We first review the basics of quadratic algebras and the associated Koszul complexes.

6.2.1 Zigzag algebra A_Γ as a trivial extension Recall that for any Γ, the zigzag algebra A_Γ is defined as the quotient of the path algebra $\mathbb{K}\mathrm{D}\Gamma$ of the double quiver $\mathrm{D}\Gamma$ by the ideal J generated by the elements

- $a_{uv}a_{vw}$ such that $u \neq w$, where v is adjacent to both u, w, and
- $a_{vw}a_{wv} - a_{vu}a_{uv}$ where v is adjacent to both u, w.

Clearly, this is an example of a quadratic algebra over k where V is the $\mathbb{K}$–vector space generated by the edges a_{wv} of $\mathrm{D}\Gamma$ and supported in grading 1. The path-length grading on $\mathbb{K}\mathrm{D}\Gamma$ descends to A_Γ where it is supported in degrees $0, 1$ and 2. It is straightforward to verify that:

Proposition 34 *For any tree Γ the quadratic dual $A_\Gamma^!$ of the zigzag algebra A_Γ is the preprojective algebra Π_Γ, when both are equipped with path-length grading.* □

[5] An independent verification of this fact based on a Morse–Bott computation of BV-operator on $\mathrm{SH}^*(T^*S^2)$ was communicated to us by P. Seidel.

As mentioned before, when Γ is a single vertex, or not a Dynkin-type tree, A_Γ is a Koszul algebra. For these cases, we have already computed $\mathrm{HH}^*(A_\Gamma)$ above (see Section 6.1 and Example 20). Henceforth, we will assume that Γ is *Dynkin*, but not a single vertex. These are the only cases when $A_\Gamma^! = \Pi_\Gamma$ is finite-dimensional.

Let us drop Γ from the notation for the moment and write

$$A = A_0 \oplus A_1 \oplus A_2 \quad \text{and} \quad \Pi = \Pi_0 \oplus \Pi_1 \oplus \cdots \oplus \Pi_{h-2}$$

for the graded pieces of A and Π. Here h stands for the Coxeter number of the Dynkin tree and it is equal to $n+1$, $2n-2$, 12, 18 and 30, for A_n, D_n, E_6, E_7 and E_8, respectively [18].

It turns out that, in this case, A_Γ is not Koszul and its Koszul complex (6) is not acyclic. Indeed, the Koszul complex is given by

$$0 \to A_\Gamma \to \Pi_1 \otimes_{\mathrm{k}} A_\Gamma \to \cdots \to \Pi_{h-2} \otimes_{\mathrm{k}} A_\Gamma \to 0 \tag{8}$$

and it fails to be exact at the right end but only there [18]. Nonetheless, in [18] the authors are able to modify the Koszul bimodule complex to obtain a $(2h-2)$–periodic complex that computes Hochschild cohomology of A_Γ. Indeed, the algebras A_Γ belong to a class of periodic algebras which are *almost Koszul*.

We will, however, now turn to a slightly different approach, which makes use of the fact that A_Γ is isomorphic to a *trivial extension algebra*.

Definition 35 Let B be a finite-dimensional algebra over the field $\mathbb{K}$. Let $B^\vee := \mathrm{Hom}_{\mathbb{K}}(B, \mathbb{K})$ be the linear dual of B, viewed naturally as a B–bimodule. The trivial extension algebra of B, denoted by $\mathcal{T}(B)$, is the vector space $B \oplus B^\vee$ equipped with the multiplication

$$(x, f) \cdot (y, g) = (xy, xg + fy).$$

If B is graded, to get a CY2 algebra, we grade $\mathcal{T}(B)$ so that $\mathcal{T}(B) = B \oplus B^\vee[-2]$.

Let $A^\to = \mathbb{K}\Gamma/J$ be the quotient of the path algebra of a quiver with respect to an arbitrary orientation of the edges modulo the ideal generated by paths of length 2. The following proposition appears in [43, Proposition 9] and results from an easy computation.

Proposition 36 *A_Γ is isomorphic to the trivial extension algebra $\mathcal{T}(A^\to)$.* □

In particular, if we orient Γ so that each vertex is either a sink or a source, then there are no paths of length 2, hence A_Γ is a trivial extension algebra of the path algebra $\mathbb{K}\Gamma$ in the bipartite orientation.

Remark 37 There is a way to understand the above proposition in terms of symplectic topology. Namely, one can consider a Lefschetz fibration $f\colon \mathbb{C}^3 \to \mathbb{C}$, $(x,y,z) \mapsto f(x,y,z)$ given by perturbing the simple singularities

$$\begin{aligned} A_n: &\quad x^2+y^2+z^{n+1} \qquad \text{for } n \geq 1,\\ D_n: &\quad x^2+zy^2+z^{n-1} \qquad \text{for } n \geq 4,\\ E_6: &\quad x^2+y^3+z^4,\\ E_7: &\quad x^2+y^3+yz^3,\\ E_8: &\quad x^2+y^3+z^5. \end{aligned}$$

One can then identify the surface X_Γ with a regular fiber of these fibrations, ie the Milnor fiber of the singularity. The spheres S_v can be identified with the vanishing spheres and the corresponding thimbles generate the Fukaya–Seidel category of f by a famous result of Seidel [61]. For a suitable choice of grading structures and ordering of objects, the Floer endomorphism algebra $A^{\to}$ of these thimbles in the Fukaya–Seidel category of f coincides with the path algebra of $\mathbb{K}\Gamma$ modulo the ideal generated by length 2 paths. The algebra isomorphism

$$A_\Gamma = A^{\to} \oplus A^{\to}[-2]$$

follows from the general relationship between the Fukaya–Seidel category of a Lefschetz fibration and the Fukaya category of its fiber (see [62, Section 4]).

We next recall the following theorem about trivial extension algebras, which we will apply to path algebras of quivers whose underlying graph is a tree. Note that by a well-known result of Bernšteĭn, Gel'fand and Ponomarev [13], the path algebras $\mathbb{K}Q$ of quivers Q obtained by orienting edges of the same *tree* in different ways are derived equivalent algebras.

Theorem 38 (Rickard [55]) *Suppose C and D are derived equivalent algebras. Then their trivial extensions $\mathcal{T}(C)$ and $\mathcal{T}(D)$ are also derived equivalent. In particular,* $\mathrm{HH}^*(\mathcal{T}(C))$ *and* $\mathrm{HH}^*(\mathcal{T}(D))$ *are isomorphic as Gerstenhaber algebras.*

Our strategy will be to apply the above theorem to $\mathcal{T}(A^{\to}) = A_\Gamma$ to pass to another algebra whose Hochschild cohomology is previously computed. However, it is important to note that the above theorem is for trivially graded algebras. On the other hand, we need to compute $\mathrm{HH}^*(A_\Gamma)$ as a bigraded algebra. What's worse, since A_Γ has elements in both even and odd degrees, we cannot simply forget about the grading and reinstate it afterwards, as in a graded resolution, odd elements affect the signs.

We next explain how to deal with this tricky point. Namely, recall from Proposition 16 that A_Γ is the graded algebra obtained as

$$A_\Gamma = \bigoplus_{v,w} \mathrm{HF}^*(S_v, S_w).$$

On the other hand, given integers $\sigma_v \in \mathbb{Z}$ for every vertex v, we can define another graded algebra

$$\widetilde{A}_\Gamma = \bigoplus_{v,w} \mathrm{Hom}(S_v[\sigma_v], S_w[\sigma_w]) = \bigoplus_{v,w} \mathrm{HF}^*(S_v, S_w)[\sigma_w - \sigma_v],$$

where $S_v[n_v]$ denotes a graded object whose grading is shifted down by n_v. Clearly, A_Γ and $\widetilde{A}_\Gamma$ are graded Morita equivalent (in particular, derived equivalent). Therefore, the (graded) Hochschild cohomologies of A_Γ and $\widetilde{A}_\Gamma$ are canonically isomorphic (see for example [64, Section (1c)]). Hence, for the purpose of computing Hochschild cohomology of A_Γ, we can choose the shifts σ_v so that the shifted algebra is supported in even degrees. In fact, using the standard tree form of Γ as in Figure 2, we simply shift the object S_v up $S_v[-\delta_v]$, where δ_v is the distance from the root to the vertex v. In this way, any arrow in the double $\mathrm{D}\Gamma$ is in degree 0 or 2 according to whether it points towards or away from the root.

Summary To compute $\mathrm{HH}^*(A_\Gamma)$ as a graded Gerstenhaber algebra, we follow this procedure:

- First check that it is possible to shift gradings so that A_Γ is supported in even degrees.
- Forget the grading altogether and treat A_Γ as an ungraded algebra.
- Compute the algebra structure of the Hochschild cohomology of the ungraded algebra by relating it to previous computations using derived equivalences of ungraded algebras in Theorem 38. This algebra will have only the cohomological grading r.
- Finally, reinstate the s–grading on $\mathrm{HH}^*(A_\Gamma)$ by finding explicit (graded) cocycles for the generators of Hochschild cohomology as an algebra.

6.2.2 Type *A* Throughout this section, Γ is the Dynkin tree A_n, $n > 1$. We describe the Hochschild cohomology ring of the zigzag algebra A_Γ in detail. We follow the strategy outlined in the previous section. Namely, we first determine the Hochschild cohomology of A_Γ as an ungraded algebra. The result will be singly graded with the cohomological grading r. We then reinstate the s–grading by explicitly identifying generators.

As was mentioned in Proposition 36, A_Γ is isomorphic to the trivial extension algebra of the path algebra $\mathbb{K}Q$ of the quiver Q with the underlying tree $\Gamma = A_n$ and oriented with the bipartite orientation (see Figure 9). Furthermore, as explained above, the derived equivalence class of a path algebra of a quiver, and hence by Theorem 38, the derived equivalence class of trivial extensions of $\mathbb{K}Q$, does not depend on the choice of the orientation of the edges of the underlying tree.

$$\bullet \longrightarrow \bullet \longleftarrow \bullet \longrightarrow \quad \cdots \quad \longleftarrow \bullet$$

Figure 9: A_n quiver in bipartite orientation

Let B_Γ be the trivial extension algebra of the path algebra of $\Gamma = A_n$ where the underlying quiver is now oriented in the linear orientation (see Figure 10).

$$\bullet \longrightarrow \bullet \longrightarrow \bullet \longrightarrow \quad \cdots \quad \longrightarrow \bullet$$

Figure 10: A_n quiver in linear orientation

Let $\widetilde{A}_{n-1}$ be the extended Dynkin quiver of type A_{n-1}, namely the quiver with cyclic orientation whose underlying graph is a simple cycle with n vertices and n edges (see Figure 11), and let us denote the ideal generated by paths of length $\geq n+1$ by J_{n+1}.

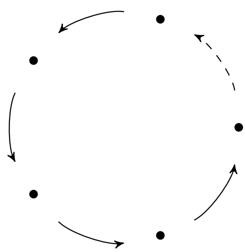


Figure 11: Cyclic quiver $\widetilde{A}_{n-1}$

The following well-known fact (cf [18]) can be verified by identifying $\mathbb{K}\Gamma$ with its image under the natural inclusion $\mathbb{K}\Gamma \to \mathbb{K}\widetilde{A}_{n-1}/J_{n+1}$, and observing that the subspace of $\mathbb{K}\widetilde{A}_{n-1}/J_{n+1}$ spanned by paths containing the unique arrow in the complement of Γ in $\widetilde{A}_{n-1}$ is canonically isomorphic to the linear dual of $\mathbb{K}\Gamma$ as a $\mathbb{K}\Gamma$–bimodule.

Lemma 39 *B_Γ is isomorphic to the truncated algebra $\mathbb{K}\widetilde{A}_{n-1}/J_{n+1}$.*

The derived equivalence between A_Γ and B_Γ implies an isomorphism between the Hochschild cohomology rings. On the other hand, the Hochschild cohomology of the (trivially graded) algebra B_Γ is studied in [42; 32; 9]. In particular, the algebra structure of $\mathrm{HH}^*(B_\Gamma)$ over a field of arbitrary characteristic was already known. Our contribution is to determine the internal s–grading coming from the grading of A_Γ. We have the following result:

Theorem 40 *As a (graded) commutative $\mathbb{K}$–algebra, the (r,s)–bigraded Hochschild cohomology algebra*

$$\mathrm{HH}^*(A_\Gamma) = \bigoplus_{r+s=*} \mathrm{HH}^r(A_\Gamma, A_\Gamma[s]),$$

of the graded k*–algebra A_Γ is given by the following generators and relations.* (The subscripts of the generators, except for s_i, refer to total degrees.)

- *Suppose* $\operatorname{char}\mathbb{K} \nmid n+1$. *We have generators labeled along with their bidegrees (r,s) given by*

$$\begin{aligned} s_1,\dots,s_n &\quad (0,2),\\ t_1 &\quad (1,0),\\ t_0 &\quad (2,-2),\\ t_{-2} &\quad (2n,-2n-2) \end{aligned}$$

  and relations

$$s_i s_j = s_i t_j = t_1^2 = t_0^n = 0.$$

- *Suppose* $\operatorname{char}\mathbb{K} \mid n+1$. *We have generators labeled along with their bidegrees (r,s) given by*

$$\begin{aligned} s_1,\dots,s_n &\quad (0,2),\\ t_1 &\quad (1,0),\\ t_0 &\quad (2,-2),\\ u_{-1} &\quad (2n-1,-2n),\\ t_{-2} &\quad (2n,-2n-2) \end{aligned}$$

  and relations

$$\begin{aligned} s_i s_j = s_i t_1 = s_i t_0 = t_1^2 &= 0,\\ s_i u_{-1} &= t_1 t_0^{n-1},\\ s_i t_{-2} &= t_0^n,\\ t_0 u_{-1} &= t_1 t_{-2},\\ t_1 u_{-1} &= \alpha t_0^n,\\ u_{-1}^2 &= \beta t_0^{n-1} t_{-2}, \end{aligned}$$

  where $\alpha=\beta=1$ if $\operatorname{char}\mathbb{K}=2$ *and* $4 \nmid n+1$, *otherwise $\alpha=\beta=0$.*

Proof The presentation of $\mathrm{HH}^*(A_\Gamma)$ given above is adapted from the presentation of $\mathrm{HH}^*(B_\Gamma)$ as a $\mathbb{K}$–algebra graded by the cohomological grading, which was calculated in [42, Theorems 8.1 and 8.2] and [32, Theorem 5.19]. In view of the isomorphism between $\mathrm{HH}^*(A_\Gamma)$ and $\mathrm{HH}^*(B_\Gamma)$ as $\mathbb{K}$–algebras graded with respect to the cohomological r–gradings, it remains to determine the s–gradings. In particular, the rank of $\mathrm{HH}^r(B_\Gamma) \cong \bigoplus_s \mathrm{HH}^r(A_\Gamma, A_\Gamma[s])$ is given explicitly in [42; 32] for each r and it can be recovered from the presentations in the statement. We will make extensive use of this information in the following arguments.

In what follows, we describe generators as elements of the reduced bar-resolution

$$\mathrm{CC}^*(A,A) := \hom_{\mathrm{k}}(T\bar{A}, A), \tag{9}$$

where $A = A_\Gamma$ and $\bar{A} = A/\mathrm{k}$. The grading on A gives a decomposition

$$\mathrm{CC}^*(A,A) = \bigoplus_{*=r+s} \mathrm{CC}^r(A, A[s]),$$

where the Hochschild differential δ is of bidegree $(1,0)$. We find explicit cocycles for $r = 0,1,2$ and show that the s–gradings of other generators are determined by the relations given above.

As a graded algebra, $A_\Gamma = A_0 \oplus A_1 \oplus A_2$, with components given by

$$A_0 = \bigoplus_{i=1}^{n} \mathbb{K} e_i, \quad A_1 = \bigoplus_{i=1}^{n-1} \mathbb{K} a_i \oplus \bigoplus_{i=1}^{n-1} \mathbb{K} b_i, \quad A_2 = \bigoplus_{i=1}^{n} \mathbb{K} s_i,$$

where $e_{i+1} a_i e_i = a_i$, $e_i b_i e_{i+1} = b_i$ and $s_{i+1} = a_i b_i = b_{i+1} a_{i+1}$.

The Hochschild differential δ in the complex (9) is given by the formula in [61, Equation (1.8)] (recall also the convention in (3)). We will only need the differentials on $\mathrm{CC}^r(A, A[s])$ for $r = 0,1,2$. These are given by

$$\begin{aligned}
\delta(c)(x_1) &= \mu^2(x_1, c) + (-1)^{(s-1)(|x_1|-1)} \mu^2(c, x_1), \\
\delta(c)(x_2, x_1) &= \mu^2(x_2, c(x_1)) + (-1)^{(s-1)(|x_1|-1)} \mu^2(c(x_2), x_1) \\
&\qquad + (-1)^s c(\mu^2(x_2, x_1)), \\
\delta(c)(x_3, x_2, x_1) &= \mu^2(x_3, c(x_2, x_1)) + (-1)^{(s-1)(|x_1|-1)} \mu^2(c(x_3, x_2), x_1) \\
&\qquad + (-1)^s c(x_3, \mu^2(x_2, x_1)) + (-1)^{s+|x_1|-1} c(\mu^2(x_3, x_2), x_1)
\end{aligned}$$

for $c \in \mathrm{CC}^0(A, A[s])$, for $c \in \mathrm{CC}^1(A, A[s])$, and for $c \in \mathrm{CC}^2(A, A[s])$, respectively.

$r = 0$ The 0–cocycles are given by central elements. The identity element

$$\sum_j e_j \in \mathrm{CC}^0(A, A[0])$$

and the elements

$$s_i \in \mathrm{CC}^0(A, A[2]) \quad \text{for } i = 1, \dots, n$$

give a basis of the center of A over $\mathbb{K}$.

$r = 1$ The 1–cocycles are given by derivations. We define a 1–cocycle $\tau_1 \in \mathrm{CC}^1(A, A[0])$ by

$$\tau_1(a_i) = -a_i, \quad \tau_1(b_i) = 0, \quad \tau_1(s_i) = s_i$$

for all $i = 1, \dots, n$. It is straightforward to check that τ_1 is a derivation but not an inner derivation, so it is a nontrivial element of $\bigoplus_s \mathrm{HH}^1(A, A[s])$, which is 1–dimensional for any $\mathbb{K}$. Therefore, any generator of this group, in particular t_1, must have the same s–grading as τ_1.

$r = 2$ We define a 2–cocycle $\tau_0 \in \mathrm{CC}^2(A, A[-2])$ given by

$$\begin{aligned}
\tau_0(a_i, b_i) &= (-1)^i e_{i+1}, \\
\tau_0(a_i, s_i) &= (-1)^{i+1} a_i, \\
\tau_0(s_i, b_i) &= (-1)^i b_i, \\
\tau_0(s_i, s_i) &= (-1)^{i+1} s_i
\end{aligned}$$

for all $i = 1, \dots, n$. Applying the Hochschild differential we get

$$\begin{aligned}
(\delta(\tau_0))(x_3, x_2, x_1) = (-1)^{|x_1|+|x_2|} & x_3 \tau_0(x_2, x_1) - \tau_0(x_3, x_2) x_1 \\
&+ (-1)^{|x_1|} \tau_0(x_3, x_2 x_1) - (-1)^{|x_1|+|x_2|} \tau_0(x_3 x_2, x_1).
\end{aligned}$$

It is straightforward (if tedious) to check that this expression vanishes identically on $\bar{A}^{\otimes 3}$. On the other hand, τ_0 cannot be a coboundary, since any $\kappa \in \mathrm{CC}^1(A, A[-2])$ has to be of the form

$$\kappa(s_i) = m_i e_i \quad \text{for some } m_i \in \mathbb{K}$$

and the Hochschild differential takes the form

$$(-1)^{|x_1|} (\delta(\kappa))(x_2, x_1) = x_2 \kappa(x_1) + \kappa(x_2 x_1) - (-1)^{|x_1|} \kappa(x_2) x_1,$$

which gives, in particular, that $\delta(\kappa)(s_i, s_i) = 0$ and $\delta(\kappa)(a_i, s_i) = m_i a_i$.

Hence, τ_0 cannot be of the form $\delta(\kappa)$ and therefore it represents a nontrivial element of the group $\bigoplus_s \mathrm{HH}^2(A, A[s])$. But we know that this group is 1–dimensional over any field $\mathbb{K}$, consequently any generator of this group over an arbitrary field $\mathbb{K}$ must have the same s–grading as τ_0.

It is harder to find explicit cocycles representing the elements u_{-1} and t_{-2} given in the statement of the theorem. Fortunately, for the purpose of determining the s–gradings we do not need explicit cocycles for these.

The element u_{-1} appears only if $\operatorname{char}\mathbb{K} \mid n+1$, and it satisfies the equation

$$s_i u_{-1} = t_1 t_0^{n-1}.$$

Since the s–gradings of s_i, t_1 and t_0 are $2, 0$ and -2, respectively, it follows that the projection u'_{-1} of u_{-1} to $\mathrm{HH}^{2n-1}(A, A[-2n])$ must be nonzero. A priori u_{-1} is not necessarily homogeneous with respect to the s–grading, but it has r–grading $2n-1$, and $\bigoplus_s \mathrm{HH}^{2n-1}(A, A[s])$ is 2–dimensional with generators u_{-1} and $t_1 t_0^{n-1}$. Therefore, u_{-1} has a decomposition $u'_{-1} + \lambda t_1 t_0^{n-1}$ into (r,s)–homogeneous elements for some $\lambda \in \mathbb{K}$. On the other hand, the relations in the statement of the theorem which involve u_{-1} are satisfied by u_{-1} if and only if they are satisfied by $u'_{-1} = u_{-1} - \lambda t_1 t_0^{n-1}$. Therefore, we may freely replace u_{-1} by u'_{-1} and hence assume that it is homogeneous with s–grading $-2n$.

Similarly, if $\operatorname{char}\mathbb{K} \mid n+1$, then $t_{-2} \in \bigoplus_s \mathrm{HH}^{2n}(A, A[s])$ appears in the relation

$$s_i t_{-2} = t_0^n$$

and $\bigoplus_s \mathrm{HH}^{2n}(A, A[s])$ is 2–dimensional with generators t_{-2} and t_0^n. As a consequence, t_{-2} has a decomposition $t_{-2} = t'_{-2} + \lambda t_0^n$ into (r,s)–homogeneous elements for some $\lambda \in \mathbb{K}$ and $t'_{-2} \neq 0$. The argument we used for u_{-1} applies here as well and we may assume that t_{-2} is homogeneous with s–grading $-2n-2$.

Finally, we need to determine the s–grading of t_{-2} over a field $\mathbb{K}$ for which $\operatorname{char}\mathbb{K} \nmid n+1$. Since A can be defined over $\mathbb{Z}$, its Hochschild cohomology groups can also be defined over $\mathbb{Z}$. Furthermore, since A has finite rank as a $\mathbb{Z}$–module, the bar-complex over $\mathbb{Z}$ is just a chain complex of finitely generated free abelian groups. So we can apply the universal coefficient theorem

$$0 \to \bigoplus_s \mathrm{HH}^r_{\mathbb{Z}}(A, A[s]) \otimes \mathbb{K} \to \bigoplus_s \mathrm{HH}^r_{\mathbb{K}}(A \otimes \mathbb{K}, A[s] \otimes \mathbb{K}) \to \mathrm{Tor}\Big(\bigoplus_s \mathrm{HH}^{r+1}_{\mathbb{Z}}(A, A[s]), \mathbb{K}\Big) \to 0. \tag{10}$$

Now, it follows from the presentation given in the statement that the middle group for $r = 2n+1$ has rank 1 for any field $\mathbb{K}$ and we know that it is supported in internal degree

$s = -2n-2$ if $\operatorname{char}\mathbb{K} \mid n+1$. Therefore, we deduce from the universal coefficient theorem (by testing $\mathbb{K} = \mathbb{F}_p$ for infinitely many primes p) that

$$\bigoplus_s \mathrm{HH}^{2n+1}_{\mathbb{Z}}(A, A[s]) = \mathbb{Z}[2n+2],$$

hence, in particular,

$$\bigoplus_s \mathrm{HH}^{2n+1}_{\mathbb{K}}(A, A[s]) = \mathbb{K}[2n+2].$$

Finally, observe that the element

$$t_1 t_{-2} \in \bigoplus_s \mathrm{HH}^{2n+1}(A, A[s]) = \mathbb{K}[2n+2]$$

is a generator of the Hochschild cohomology group in grading $r = 2n+1$ over an arbitrary field $\mathbb{K}$, and hence t_{-2} must have s–grading $-2n-2$ over an arbitrary field $\mathbb{K}$. □

Remark 41 Over the finite field $\mathbb{F}_3$ of characteristic 3, the group algebra $\mathbb{F}_3\mathfrak{S}_3$ of the symmetric group in three letters is isomorphic to the algebra A_Γ for $\Gamma = A_2$. A presentation for the Hochschild cohomology ring of this group algebra was given in [68, Theorem 7.1]. This agrees with the presentation given above.

As a consequence of Theorem 40 we conclude that the group $\bigoplus_{r+s=*} \mathrm{HH}^r(A_\Gamma, A_\Gamma[s])$ is nontrivial if and only if $* \le 2$. If $\operatorname{char}\mathbb{K} \nmid n+1$, the rank is n at each $* \le 2$, otherwise the rank is n for $* = 2, 1$ and $n+1$ for $* \le 0$.

Recall that we have proved in Theorem 27 that there is an isomorphism of Gerstenhaber algebras

$$\mathrm{SH}^*(X_\Gamma) \cong \mathrm{HH}^*(A_\Gamma)$$

over a field $\mathbb{K}$ of characteristic 0, where the Conley–Zehnder grading on the left corresponds to the total grading $r+s$ on the right. Having computed $\mathrm{HH}^*(A_\Gamma)$ as a bigraded algebra, we immediately get a description of the algebra structure of the symplectic cohomology. Let us also record its rank.

Corollary 42 *The symplectic cohomology group $\mathrm{SH}^*(X_\Gamma)$ over a field $\mathbb{K}$ of characteristic 0 is of rank n if $* \le 2$ and it is trivial otherwise.*

We have also performed computer-aided checks on our calculations. Tables 2 and 3 list the ranks (of a finite portion) for the cases A_2 and A_3.

$r+s \downarrow$ $\quad s \rightarrow$	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8
2	2	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	0	0	0	0	0	0
0	0	0	1	0	1	0	x	0	0	0	0
-1	0	0	0	0	0	0	x	0	1	0	1
-2	0	0	0	0	0	0	0	0	1	0	1

Table 2: $\Gamma = A_2$; x is 1 if $\operatorname{char} \mathbb{K} = 3$, 0 otherwise

$r+s \downarrow$ $\quad s \rightarrow$	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8
2	3	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	1	0	1	0	0	0	0
0	0	0	1	0	1	0	1	0	x	0	0
-1	0	0	0	0	0	0	0	0	x	0	1
-2	0	0	0	0	0	0	0	0	0	0	1

Table 3: $\Gamma = A_3$; x is 1 if $\operatorname{char} \mathbb{K} = 2$, 0 otherwise

6.2.3 Type *D* In this section we consider the case where Γ is the Dynkin tree D_n, $n \geq 4$. Most of the arguments in the previous section apply verbatim or with minor modifications. So we will focus on the differences and provide details as necessary.

Considering the quiver based on Γ with the orientation of the arrows given by Figure 12, we obtain the following result.

$$\begin{tikzcd}[row sep=small]
 & & & & & \bullet \\
\bullet \arrow[r, "a_1"] & \bullet \arrow[r, "a_2"] & \bullet \arrow[r, "a_3"] & \cdots \arrow[r, "a_{n-3}"] & \bullet \arrow[ur, "a_{n-2}"] \arrow[dr, "a_{n-1}"'] & \\
 & & & & & \bullet
\end{tikzcd}$$

Figure 12: D_n quiver

Lemma 43 *The trivial extension algebra B_Γ of the path algebra $\mathbb{K}\Gamma$ is isomorphic to the quotient $\mathbb{K}Q/I$, where Q is the quiver given in Figure 13 and I is the ideal generated by the elements*

$$\beta_{n-1}\gamma_{n-1} - \beta_n\gamma_n, \quad \alpha_i \cdots \alpha_1 \beta_n \gamma_n \alpha_{n-3} \cdots \alpha_i,$$
$$\gamma_n \alpha_{n-3} \cdots \alpha_1 \beta_{n-1}, \quad \gamma_{n-1}\alpha_{n-3} \cdots \alpha_1 \beta_n.$$

Proof Using the identifications $a_i \leftrightarrow \alpha_i$ for $1 \leq i \leq n-3$ and $a_j \leftrightarrow \gamma_{j+1}$ for $j = n-2$ and $n-1$, we can consider $\mathbb{K}\Gamma$ as a subalgebra of $\mathbb{K}Q/I$. Observe that $\mathbb{K}Q/I$ decomposes as a direct sum $\mathbb{K}\Gamma \oplus V$ and V is generated by β_{n-1} and β_n as a

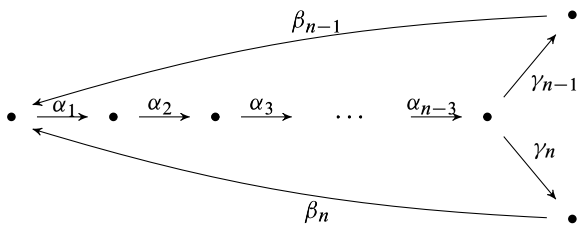


Figure 13: The quiver Q

$\mathbb{K}\Gamma$–bimodule. Moreover, as $\mathbb{K}\Gamma$–bimodules, V and the dual of $\mathbb{K}\Gamma$ are isomorphic via

$$\begin{aligned}\psi\colon V &\to (\mathbb{K}\Gamma)^\vee,\\ \beta_{n-1} &\mapsto (a_{n-2}a_{n-3}\cdots a_2a_1)^\vee,\\ \beta_n &\mapsto (a_{n-1}a_{n-3}\cdots a_2a_1)^\vee.\end{aligned}$$

It is straightforward to check that this map is a well-defined isomorphism.

In fact, $(\mathbb{K}\Gamma)^\vee$ can also be considered as a subalgebra of $\mathbb{K}Q/I$ by identifying the dual $p^\vee$ of a path $p \in \mathbb{K}\Gamma$ with the path $q \in \mathbb{K}Q/I$ such that

$$q \cdot p = \tau^t(\beta_n\gamma_n\alpha_{n-3}\cdots\alpha_1) = \tau^t(\beta_{n-1}\gamma_{n-1}\alpha_{n-3}\cdots\alpha_1) \in \mathbb{K}Q/I,$$

where τ denotes the simple rotation action on the cycles and t is the distance between the initial points of p and α_1. □

As a consequence of this lemma and the discussions in the previous section, there is an isomorphism between the Hochschild cohomology rings of the zigzag algebra A_Γ and B_Γ. On the other hand, the Hochschild cohomology of B_Γ as a trivially graded algebra was described in detail in [72; 71]. As in the case of $\Gamma = A_n$ (see Theorem 40), we determine the internal grading s induced by the zigzag algebra and obtain the following result. This extra information does not appear in [72; 71] and the determination of this grading is the main contribution given in the following theorem.

Theorem 44 *Let $\Gamma = D_n$, $n \geq 4$. The (r,s)–bigraded Hochschild cohomology algebra*

$$\mathrm{HH}^*(A_\Gamma) = \bigoplus_{r+s=*} \mathrm{HH}^r(A_\Gamma, A_\Gamma[s])$$

of the graded k*–algebra A_Γ is (graded) commutative and given by the following generators and relations. (The subscripts of the generators, except for the s_i, refer to total degrees.)*

(1) *Suppose* $\operatorname{char}\mathbb{K} \neq 2$. *We have generators labeled along with their bidegrees* (r,s) *given by*

$$\begin{array}{rl} s_1,\dots,s_n & (0,2), \\ t_1 & (1,0), \\ r_1 & (2n-3,-2n+4), \\ t_0 & (4,-4), \\ r_0 & (2n-4,-2n+4), \\ t_{-2} & (4n-6,-4n+4) \end{array}$$

and relations

$$s_is_j = s_it_j = s_ir_j = t_1^2 = t_1r_1 = r_1^2 = t_0^{n-1} = 0,$$

together with

	if n is even	if n is odd
$t_1r_0 =$	$\binom{n}{2}t_1t_0^{(n-2)/2} - (n-1)r_1$	$\binom{n-1}{2}r_1$
$2t_0r_1 =$	$t_1t_0^{n/2}$	0
$2r_1r_0 =$	0	$t_1t_0^{n-2}$
$2t_0r_0 =$	$t_0^{n/2}$	0
$2r_0^2 =$	$\binom{n}{2}t_0^{n-2}$	$\binom{n-1}{2}t_0^{n-2}$

(2) *Suppose* $\operatorname{char}\mathbb{K} = 2$. *We have generators labeled along with their bidegrees* (r,s) *given by*

$$\begin{array}{rl} s_1,\dots,s_n & (0,2), \\ t_1 & (1,0), \\ u_1 & (3,-2), \\ t_0 & (4,-4), \\ r_0 & (2n-4,-2n+4), \\ u_0 & \left(4\left\lfloor \frac{n}{2}\right\rfloor, -4\left\lfloor \frac{n}{2}\right\rfloor\right), \\ u_{-1} & \left(4\left\lfloor \frac{n-1}{2}\right\rfloor + 1, -4\left\lfloor \frac{n-1}{2}\right\rfloor - 2\right), \\ t_{-2} & (4n-6,-4n+4) \end{array}$$

and relations

$$\begin{gathered} s_i s_j = s_i t_1 = s_i u_1 = s_i u_0 = 0, \\ t_1^2 = u_1^2 = u_0^2 = u_1 u_0 = 0, \\ t_0^{\lfloor \frac{n}{2} \rfloor} = u_1 t_0^{\lfloor \frac{n-1}{2} \rfloor} = 0, \\ r_0^2 = \lfloor \tfrac{n}{2} \rfloor u_0 t_0^{\lfloor \frac{n-3}{2} \rfloor}, \\ s_j t_0 = t_1 u_1 \end{gathered}$$

together with

	if n is even	if n is odd
$u_{-1}^2 =$	t_{-2}	$t_{-2}t_0$
$u_1 u_{-1} =$	u_0	$u_0 t_0$
$t_0 r_0 =$	$u_1 u_{-1}$	$t_1 u_{-1}$
$u_1 r_0 =$	0	$t_1 u_0$
$s_j u_{-1} =$	$\begin{cases} \left(\frac{n-2}{2}\right) t_1 t_0^{(n-2)/2} + t_1 r_0 & \textit{if } j \le n-1, \\ \left(\frac{n}{2}\right) t_1 t_0^{(n-2)/2} + t_1 r_0 & \textit{if } j = n \end{cases}$	$u_1 r_0$
$s_j r_0 =$	$\begin{cases} t_1 u_1 t_0^{(n-4)/2} & \textit{if } j \le n-1, \\ 0 & \textit{if } j = n \end{cases}$	0
$u_{-1} r_0 =$		$t_1 t_{-2}$
$t_1 r_0 =$		$\left(\frac{n-1}{2}\right) u_1 t_0^{(n-3)/2}$
$s_j t_{-2} =$		$r_0 u_0$

Proof The presentation of the algebra structure of $\mathrm{HH}^*(B_\Gamma)$ in [71, Theorem 4] provides all the generators with their r–gradings and relations. The derived equivalence between A_Γ and B_Γ gives

$$\mathrm{HH}^r(B_\Gamma) \cong \bigoplus_s \mathrm{HH}^r(A_\Gamma, A_\Gamma[s]).$$

Therefore it suffices to determine the s–gradings of the generators in the statement. Extending the notation in Figure 12, we consider the decomposition of the graded algebra A_Γ into homogeneous $\mathbb{K}$–subspaces A_0, A_1 and A_2, spanned by

$$\{e_1, \dots, e_n\}, \quad \{a_1, b_1, \dots, a_{n-1}, b_{n-1}\} \quad \text{and} \quad \{s_1, \dots, s_n\},$$

respectively, where

$$e_{i+1}a_ie_i = a_i, \quad e_ib_ie_{i+1} = b_i, \quad e_na_{n-1}e_{n-2} = a_{n-1}, \quad e_{n-2}b_{n-1}e_n = b_{n-1},$$
$$s_1 = b_1a_1, \quad s_{i+1} = a_ib_i = b_{i+1}a_{i+1}, \quad s_{n-2} = a_{n-3}b_{n-3} = b_ja_j, \quad s_{j+1} = a_jb_j$$

for $1 \le i \le n-4$ and $j = n-2, n-1$.

As in the proof of Theorem 40, we will again use the reduced bar-resolution associated to $A = A_\Gamma$ and denote the Hochschild differential by δ. Consequently, the discussion for $r = 0, 1$ is exactly the same as in the proof of Theorem 40. We identify the s–gradings of $s_1, \dots, s_n$ and t_1 as in the statement.

For every nonnegative integer r, the dimension of $\bigoplus_s \mathrm{HH}^r(A, A[s]) \cong \mathrm{HH}^r(B_\Gamma)$ can be deduced from the presentation in the statement and it is explicitly given in [71, Theorem 3]. We will make extensive use of this information. To begin with, note that $\bigoplus_s \mathrm{HH}^2(A, A[s])$ is trivial over any field $\mathbb{K}$, and $\bigoplus_s \mathrm{HH}^3(A, A[s])$ is 1–dimensional if $\operatorname{char}\mathbb{K} = 2$ and trivial otherwise. Over a field $\mathbb{K}$ of characteristic 2, for $c \in \mathrm{CC}^3(A, A[s])$, the Hochschild differential δ is given by

$$\begin{aligned}\delta(c)(x_4, x_3, x_2, x_1) = x_4c(x_3, x_2, x_1) + c(x_4, x_3, x_2)x_1 + c(x_4x_3, x_2, x_1)\\ + c(x_4, x_3x_2, x_1) + c(x_4, x_3, x_2x_1).\end{aligned}$$

We claim that, if $\operatorname{char}\mathbb{K} = 2$, there is a cocycle $\upsilon_1 \in \mathrm{CC}^3(A, A[-2])$ which is not the coboundary of any $\kappa \in \mathrm{CC}^2(A, A[s])$. This and the fact that $\bigoplus_s \mathrm{HH}^3(A, A[s])$ is 1–dimensional imply that the s–grading of u_1 must be -2, the same as υ_1. To describe the graded homomorphism $\upsilon_1\colon \bar{A}^{\otimes 3} \to A[-2]$ uniquely, it suffices to list the generators of $\bar{A}^{\otimes 3}$ on which υ_1 is nonzero. It necessarily vanishes on any element of degree 5 or 6 in $\bar{A}^{\otimes 3}$ since A is supported in gradings between 0 and 2. We declare υ_1 to be nonzero exactly on those nontrivial elements $(x_3, x_2, x_1) \in \bar{A}^{\otimes 3}$ which satisfy one of the following conditions:

- One of x_1, x_2 and x_3 is of the form a_i and the other two is of the form b_i, possibly with different indices, and $(x_3, x_2, x_1) \neq (b_{n-1}, a_{n-1}, b_{n-2})$.
- Exactly one of x_1, x_2 and x_3 is of the form s_k, and the initial point of x_1 matches the terminal point of x_3.
- $(x_3, x_2, x_1) = (a_{n-2}, b_{n-1}, a_{n-1})$.

It is straightforward to check that υ_1 is a cocycle. To see that it is not a coboundary, suppose that $c \in \mathrm{CC}^2(A, A[-2])$. Then

$$\begin{aligned}&\delta(\kappa)((b_2, a_2, s_2) + (a_2, s_2, b_2) + (s_2, b_2, a_2))\\ &\qquad = b_2\kappa(a_2, s_2) + a_2\kappa(s_2, b_2) + \kappa(a_2, s_2)b_2 + \kappa(s_2, b_2)a_2\end{aligned}$$

after cancellations. Observe that the right-hand side is either s_2+s_3 or 0, depending on the values of $\kappa(a_2,s_2)$ and $\kappa(s_2,b_2)$. Since

$$\upsilon_1((b_2,a_2,s_2)+(a_2,s_2,b_2)+(s_2,b_2,a_2))=s_3,$$

υ_1 cannot be a coboundary.

Next we determine the s–grading of t_0. Consider the case $\operatorname{char}\mathbb{K}=2$. If $n=4$, then $\bigoplus_s \mathrm{HH}^4(A,A[s])$ has generators t_0, r_0 and t_1u_1. Note that any relation satisfied by t_0 and r_0 is also satisfied by $t_0-\gamma t_1u_1$ and $r_0-\gamma t_1u_1$, respectively, for any $\gamma\in\mathbb{K}$. Therefore, without loss of generality, we may assume that there are s–homogeneous generators t_0', r_0' and constants $\alpha,\beta\in\mathbb{K}$ such that

$$t_0=t_0'+\alpha r_0' \quad \text{and} \quad r_0=r_0'+\beta t_0'.$$

From the relations regarding s_nr_0 and s_nt_0 we obtain

$$0=s_nr_0=s_nr_0'+\beta s_nt_0' \quad \text{and} \quad 0\neq u_1t_1=s_nt_0=s_nt_0'+\alpha s_nr_0'.$$

Since the gradings of u_1, t_1 and s_n are established above, the second equation implies that at least one of t_0' and r_0' has s–grading -4; in fact they both do, as the following arguments show. If $s_nr_0'\neq 0$, then the first equation proves that r_0' and t_0' have the same s–grading, which is necessarily -4. So suppose $s_nr_0'=0$. Now the second equation gives $s_nt_0'\neq 0$. Moreover, the first equation implies $\beta=0$, which means $r_0=r_0'$; in particular, r_0 is s–homogeneous. So we can use the relation $s_1r_0=t_1u_1$ to establish the s–grading of r_0' as -4. On the other hand, under the assumption $s_nr_0'=0$, the second equation becomes $s_nt_0'=u_1t_1$, implying that t_0' has s–grading -4 as well. Therefore, regardless of the value of s_nr_0', the s–gradings of t_0 and r_0 are both -4.

If $n>4$ and $\operatorname{char}\mathbb{K}=2$, then $\bigoplus_s \mathrm{HH}^4(A,A[s])$ has rank 2 with generators t_0 and t_1u_1, hence we may assume that there is an s–homogeneous generator t_0' and $\alpha\in\mathbb{K}$ such that $t_0=t_0'+\alpha t_1u_1$. The relation $s_nt_0=t_1u_1$ implies that the s–grading of t_0' is -4. The s–grading of t_1u_1 is -2 by previous computations. If n is even, then any relation in the statement holds for t_0 if and only if it holds for t_0'. Therefore, without loss of generality, we may assume that $t_0=t_0'$ is s–homogeneous with grading -4, at least when n is even. The same conclusion holds for odd n as well, but we will not prove (nor use) it until Case 3 below.

Let us now consider the s–grading of t_0 when $\operatorname{char}\mathbb{K}\neq 2$. Regardless of whether $n=4$ or not, the argument uses the universal coefficient theorem (10) as in the proof of Theorem 40. First of all, considering that $\bigoplus_s \mathrm{HH}^2(A,A[s])$ is trivial for any field $\mathbb{K}$ and using (10) for $r=2$, we conclude that $\bigoplus_s \mathrm{HH}^3_{\mathbb{Z}}(A,A[s])$ has no torsion. Since $\bigoplus_s \mathrm{HH}^3(A,A[s])$ is trivial when $\operatorname{char}\mathbb{K}\neq 2$, applying the universal coefficient theorem (10) for $r = 3$ implies that $\bigoplus_s \mathrm{HH}^3_{\mathbb{Z}}(A, A[s])$ has no free component either, hence it is trivial. Moreover, the same exact sequence and the fact that for $\operatorname{char}\mathbb{K} = 2$, $\bigoplus_s \mathrm{HH}^3(A, A[s])$ is generated by u_1 whose s–grading is computed as -2 above, establish the torsion of $\bigoplus_s \mathrm{HH}^4_{\mathbb{Z}}(A, A[s])$ as $\mathbb{Z}_2[2]$.

The argument above for $\operatorname{char}\mathbb{K} = 2$ shows that $\bigoplus_s \mathrm{HH}^4(A, A[s]) \cong \mathbb{K}^d[4] \oplus \mathbb{K}[2]$, where $d = 2$ if $n = 4$ and $d = 1$ otherwise. Using the fact that $\bigoplus_s \mathrm{HH}^4(A, A[s])$ is d–dimensional for any field $\mathbb{K}$ with $\operatorname{char}\mathbb{K} \neq 2$, and applying the universal coefficient theorem (10) for $r = 4$ to infinitely many characteristics, we conclude that $\bigoplus_s \mathrm{HH}^4_{\mathbb{Z}}(A, A[s])$ is in fact $\mathbb{Z}^d[4] \oplus \mathbb{Z}_2[2]$. In particular, $\bigoplus_s \mathrm{HH}^4(A, A[s])$ is supported in s–grading -4 whenever $\operatorname{char}\mathbb{K} \neq 2$, and the s–grading of t_0 is -4 unless n is odd and $\operatorname{char}\mathbb{K} = 2$.

The rest of the argument varies slightly according to the parity of n and the characteristic of the base field.

Case 1 (n even and $\operatorname{char}\mathbb{K} = 2$)

We need to determine the s–gradings of the rest of the generators, namely u_{-1}, t_{-2}, u_0 and r_0. Since

$$\{u_{-1}, t_1 r_0, t_1 t_0^{(n-2)/2}\}$$

forms a basis of $\bigoplus_s \mathrm{HH}^{2n-3}(A, A[s])$,

$$u_{-1} = u'_{-1} + \alpha t_1 r_0 + \beta t_1 t_0^{(n-2)/2}$$

for some s–homogeneous $u'_{-1} \neq 0$ and some $\alpha, \beta \in \mathbb{K}$. Observe that any relation satisfied by u_{-1} is satisfied by u'_{-1} as well. Therefore, without loss of generality, we may assume that $u_{-1} = u'_{-1}$ and its s–grading is $-2n + 2$ as a result of the relation

$$s_n u_{-1} - s_1 u_{-1} = t_1 t_0^{(n-2)/2}.$$

Moreover, by the relations $u_0 = u_1 u_{-1}$ and $t_{-2} = u_{-1}^2$, both u_0 and t_{-2} are s–homogeneous with gradings $-2n$ and $-4n + 4$, respectively. Regarding r_0, note that

$$\{r_0, t_0^{(n-2)/2}, t_1 u_1 t_0^{(n-4)/2}\}$$

forms a basis of $\bigoplus_s \mathrm{HH}^{2n-4}(A, A[s])$. Hence

$$r_0 = r'_0 + \alpha t_0^{(n-2)/2} + \beta t_1 u_1 t_0^{(n-4)/2}$$

for some s–homogeneous $r'_0 \neq 0$ and some $\alpha, \beta \in \mathbb{K}$. It is straightforward to check that any relation satisfied by r_0 is also satisfied by $r_0 - \beta t_1 u_1 t_0^{(n-4)/2}$, so we may assume that $r_0 = r'_0 + \alpha t_0^{(n-2)/2}$. Moreover, the relation $u_0 = t_0 r_0 = t_0 r'_0$ implies that the s–grading of r_0' is $-2n+4$, the same as that of $t_0^{(n-2)/2}$. Therefore, r_0 is s–homogeneous with this grading as well.

Case 2 (n even and $\operatorname{char}\mathbb{K} \neq 2$)

We have a single argument for the s–grading of r_0 and r_1 which belong to 2–dimensional spaces $\bigoplus_s \mathrm{HH}^{2n-4}(A, A[s])$ and $\bigoplus_s \mathrm{HH}^{2n-3}(A, A[s])$, respectively. We take s–homogeneous elements $r_0' \neq 0$ and $r_1' \neq 0$ such that

$$r_0 = r_0' + \alpha t_0^{(n-2)/2} \quad \text{and} \quad r_1 = r_1' + \beta t_1 t_0^{(n-2)/2}.$$

Suppose that $\operatorname{char}\mathbb{K} \nmid n-1$. By way of contradiction, assume that r_0 is not s–homogeneous, ie $\alpha \neq 0$ and the s–grading of r_0' is not $-2n+4$. Then $t_1 r_0 = \left(\frac{n}{2}\right) t_1 t_0^{(n-2)/2} - (n-1) r_1$ implies that $-(n-1) r_1' = t_1 r_0'$ for grading reasons. Consequently, the s–gradings of r_0' and r_1' should match. Moreover, since $2 t_0 r_1 = t_1 t_0^{n/2}$, and again for grading reasons, $\beta \neq 0$. But then, $\alpha\beta t_1 t_0^{n-2} \neq 0$ and its s–grading does not match with the s–grading of any other term in the product $r_1 r_0$ contradicting with $r_1 r_0 = 0$. Therefore r_0 is s–homogeneous, and so is r_1, in fact with the same s–grading, as a consequence of

$$t_1 r_0 = \left(\tfrac{1}{2} n\right) t_1 t_0^{(n-2)/2} - (n-1) r_1.$$

In order to account for the possibility that $\operatorname{char}\mathbb{K} \mid (n/2)$, instead of the relation above we use the relation $2 t_0 r_0 = t_0^{n/2}$ to obtain the common s–grading of r_0 and r_1.

For a field $\mathbb{K}$ with $\operatorname{char}\mathbb{K} \neq 2$, both $\bigoplus_s \mathrm{HH}^{2n-3}(A, A[s])$ and $\bigoplus_s \mathrm{HH}^{2n-4}(A, A[s])$ are 2–dimensional, and moreover we just proved that when $\operatorname{char}\mathbb{K} \nmid n-1$, each of these spaces are supported in $s = -2n+4$. By using the universal coefficient theorem (10) for $r = 2n-4$ we conclude that, as long as $\operatorname{char}\mathbb{K} \neq 2$ (even if $\operatorname{char}\mathbb{K}$ divides $n-1$) both $\bigoplus_s \mathrm{HH}^{2n-3}(A, A[s])$ and $\bigoplus_s \mathrm{HH}^{2n-4}(A, A[s])$ are supported in $s = -2n+4$. In particular, the common s–grading of r_0 and r_1 is $-2n+4$.

The s–grading of the remaining generator t_{-2} is obtained by the following argument, which applies to odd n as well. First of all, t_{-2} is s–homogeneous as it belongs to the 1–dimensional space $\bigoplus_s \mathrm{HH}^{4n-6}(A, A[s])$. On the other hand, $\bigoplus_s \mathrm{HH}^{4n-5}(A, A[s])$ is 1–dimensional over any field $\mathbb{K}$ and it is generated by $t_1 t_{-2}$. Since we already have the s–grading of $t_1 t_{-2}$ for $\operatorname{char}\mathbb{K} = 2$ from the previous case, we obtain the s–grading of t_{-2} over any field using the universal coefficient theorem (10) for $r = 4n-5$.

Case 3 (n odd and $\operatorname{char}\mathbb{K} = 2$)

In this case, the s–grading of r_0 can be obtained by an argument which works regardless of $\operatorname{char}\mathbb{K}$. Over any $\mathbb{K}$, $\bigoplus_s \mathrm{HH}^{2n-4}(A, A[s])$ is 1–dimensional and generated by r_0, which is therefore s–homogeneous. Applying the universal coefficient theorem

(10) for $r = 2n-4$ and infinitely many different characteristics, we conclude that $\bigoplus_s \mathrm{HH}^{2n-4}_{\mathbb{Z}}(A, A[s]) \cong \mathbb{Z}$ and to establish the s–grading of this group, it suffices to use the relation $2r_0^2 = \binom{n-1}{2} t_0^{n-2}$ over a field of characteristic 0. In particular, r_0 has s–grading $-2n+4$ for any field $\mathbb{K}$.

The generator u_0 belongs to the 1–dimensional space $\bigoplus_s \mathrm{HH}^{2n-2}(A, A[s])$, hence it is s–homogeneous, and its s–grading is determined by the relation $u_1 r_0 = t_1 u_0$.

Next we consider u_{-1}. It belongs to $\bigoplus_s \mathrm{HH}^{2n-1}(A, A[s])$ which is generated by u_{-1} and $u_1 r_0$. So $u_{-1} = u'_{-1} + \alpha u_1 r_0$ for some $\alpha \in \mathbb{K}$ and s–homogeneous $u'_{-1} \neq 0$. Observe that any relation which involves u_{-1} is satisfied by u'_{-1} as well. Hence we may assume that u_{-1} is s–homogeneous. Its s–grading is obtained from

$$t_1 u_{-1} = t_0 r_0 = t'_0 r_0.$$

Note that we have not established the s–homogeneity of t_0 in this case yet, and that is why we had to refer to t'_0 in the relation above and use the fact that $t_0 r_0 - t'_0 r_0 = 0$ since it is a multiple of $t_1 u_1 r_0 = s_j t_0 r_0 = 0$.

Finally, we determine the s–gradings of t_0 and t_{-2} simultaneously. In the case we consider, they belong to 2–dimensional spaces

$$\bigoplus_s \mathrm{HH}^4(A, A[s]) \quad \text{and} \quad \bigoplus_s \mathrm{HH}^{4n-6}(A, A[s]),$$

with respective bases $\{t_0, t_1 u_1\}$ and $\{t_{-2}, r_0 u_0\}$. So there are s–homogeneous elements t'_0 and t'_{-2} with constants $\alpha, \beta \in \mathbb{K}$ such that

$$t_0 = t'_0 + \alpha t_1 u_1 \quad \text{and} \quad t_{-2} = t'_{-2} + \beta r_0 u_0 .$$

In fact, the s–gradings of t'_0 and t'_{-2} are -4 and $-4n+4$, respectively, since $s_j t'_0 = t_1 u_1$ and $s_j t'_{-2} = r_0 u_0$. It is straightforward to check that any relation in the statement, except for $u_{-1}^2 = t_{-2} t_0$, holds for t_0 and t_{-2} if and only if it holds for t'_0 and t'_{-2}. To check that the remaining relation holds, we use

$$u_{-1}^2 = t_{-2} t_0 = t'_{-2} t'_0 + \alpha t'_{-2} t_1 u_1 + \beta t'_0 r_0 u_0 + \alpha\beta t_1 u_1 r_0 u_0$$

and observe that the only term on the right-hand side of the above relation whose s–grading matches that of u_{-1}^2 is $t'_{-2} t'_0$. Therefore, without loss of generality, we may assume that $t_0 = t'_0$ and $t_{-2} = t'_{-2}$.

Case 4 (n is odd and $\operatorname{char} \mathbb{K} \neq 2$)

The s–gradings of t_{-2} and r_0 are already obtained in Cases 2 and 3 above.

The remaining generator r_1 is s–homogeneous since it belongs to the 1–dimensional space $\bigoplus_s \mathrm{HH}^{2n-3}(A, A[s])$ and its s–grading is determined by the relation $2r_1 r_0 = t_1 t_0^{n-2}$. □

Using Theorem 18, which is due to Seidel and Thomas, one gets the following consequence of the computation above.

Corollary 45 *If* $\operatorname{char}\mathbb{K} \neq 2$ *and* Γ *is of type* D_n, $n \geq 4$, *then the zigzag algebra* A_Γ *is intrinsically formal.*

One can write explicit bases for the relevant $\mathbb{K}$–vector subspaces of $\mathrm{HH}^*(A_\Gamma)$ as follows.

If $\operatorname{char}\mathbb{K} \neq 2$, then $\bigoplus_{r+s=2} \mathrm{HH}^r(A, A[s])$ is spanned by $\{s_1, \ldots, s_n\}$, and for any nonnegative integer m and $i = 0, 1$, a basis of $\bigoplus_{r+s=i-2m} \mathrm{HH}^r(A, A[s])$ is given by

$$\{r_i t_{-2}^m, t_1^i t_0^k t_{-2}^m : 0 \leq k \leq n-2\}.$$

When $\operatorname{char}\mathbb{K} = 2$, the increase in the dimensions of these spaces is immediate from the statement of Theorem 44. The subspace $\bigoplus_{r+s=2} \mathrm{HH}^r(A, A[s])$ is spanned by

$$\left\{s_j, t_1 u_1 t_0^k = s_n t_0^{k+1} : 1 \leq j \leq n,\ 0 \leq k \leq \left\lfloor \tfrac{n-4}{2} \right\rfloor\right\},$$

and depending on the parity of n, $\bigoplus_{r+s=1} \mathrm{HH}^r(A, A[s])$ is spanned by

$$\left\{u_1 t_0^k, t_1 t_0^l, t_1 r_0 t_0^l : 0 \leq k \leq \tfrac{n-4}{2},\ 0 \leq l \leq \tfrac{n-2}{2}\right\}$$

if n is even, and by

$$\left\{u_1 t_0^l, t_1 t_0^l, t_1 u_0 t_0^l : 0 \leq l \leq \tfrac{n-3}{2}\right\}$$

if n is odd.

If n is even and m is nonnegative, then a basis of $\bigoplus_{r+s=-m} \mathrm{HH}^r(A, A[s])$ can be given as

$$\left\{t_0^l u_{-1}^m, r_0 t_0^l u_{-1}^m, t_1 t_0^l u_{-1}^{m+1}, r_0 t_1 t_0^l u_{-1}^{m+1} : 0 \leq l \leq \tfrac{n-2}{2}\right\}.$$

If n is odd and m is nonnegative, then

$$\bigoplus_{r+s=-2m} \mathrm{HH}^r(A, A[s]) \quad \text{and} \quad \bigoplus_{r+s=-2m-1} \mathrm{HH}^r(A, A[s])$$

are spanned by

$$\left\{t_0^l t_{-2}^m, r_0 t_0^l t_{-2}^m, u_0 t_0^l t_{-2}^m, u_0 r_0 t_0^l t_{-2}^m : 0 \leq l \leq \tfrac{n-3}{2}\right\}$$

and

$$\left\{u_{-1} t_0^l t_{-2}^m, u_{-1} r_0 t_0^l t_{-2}^m, u_{-1} u_0 t_0^l t_{-2}^m, u_{-1} u_0 r_0 t_0^l t_{-2}^m :\ 0 \leq l \leq \tfrac{n-3}{2}\right\},$$

respectively.

Therefore, the group $\bigoplus_{r+s=*} \mathrm{HH}^r(A_\Gamma, A_\Gamma[s])$ is nontrivial if and only if $* \leq 2$. If the ground field has characteristic 2, the rank is $n + \left\lfloor \frac{n-2}{2} \right\rfloor$ for $* = 2, 1$ and $4\left\lfloor \frac{n}{2} \right\rfloor$ for

$* \le 0$. Otherwise the rank is n at each $* \le 2$. Therefore, it follows from Theorem 27 that we have:

Corollary 46 *The symplectic cohomology group $\mathrm{SH}^*(X_\Gamma)$ over a field of characteristic 0 is of rank n if $* \le 2$ and it is trivial otherwise.*

As before, for convenient access, we give tables listing the ranks of a truncated piece of our calculation. As mentioned in Section 6.2.1, A_Γ has a graded periodic resolution as a graded bimodule, from which it follows easily that for $\Gamma = D_n$, $n \ge 4$, the ranks of the Hochschild cohomology groups obeys the following periodicity:

$$\operatorname{rank} \mathrm{HH}^r(A, A[s]) = \operatorname{rank} \mathrm{HH}^{r+(4n-6)}(A, A[s-(4n-4)]) \quad \text{for } r > 0.$$

In this presentation, multiplication by the generator t_{-2} gives rise to this periodicity. The tables below give the truncation, which includes a fundamental domain of the period in the cases $\Gamma = D_4, D_5, D_6$. We have also performed computer-aided checks in these cases.

$r+s \downarrow$ $s \to$	2	1	0	−1	−2	−3	−4	−5	−6	−7	−8	−9	−10
2	4	0	0	0	x	0	0	0	0	0	0	0	0
1	0	0	1	0	x	0	2	0	0	0	1	0	0
0	0	0	1	0	0	0	2	0	x	0	1	0	$2x$
−1	0	0	0	0	0	0	0	0	x	0	0	0	$2x$

Table 4: $\Gamma = D_4$; x is 1 if $\operatorname{char} \mathbb{K} = 2$, 0 otherwise

$r+s \downarrow$ $s \to$	2	1	0	−1	−2	−3	−4	−5	−6	−7	−8	−9	−10	−11	−12	−13	−14
2	5	0	0	0	x	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	1	0	x	0	1	0	1	0	1	0	0	0	1	0	0
0	0	0	1	0	0	0	1	0	1	0	1	0	x	0	1	0	x
−1	0	0	0	0	0	0	0	0	0	0	0	0	x	0	0	0	x

Table 5: $\Gamma = D_5$; x is 1 if $\operatorname{char} \mathbb{K} = 2$, 0 otherwise

Remark 47 As a result of the computation for $\Gamma = D_n$, we have $\mathrm{HH}^2(A_\Gamma, A_\Gamma[s]) = 0$ for all s over any field $\mathbb{K}$. This rigidity has a useful implication in Floer theory: namely, if one has a D_n–configuration of Lagrangian spheres S_v in a symplectic 4–manifold M, then the Floer cohomology algebra $\bigoplus_{v,w} \mathrm{HF}^*_M(S_v, S_w)$ is isomorphic to A_Γ, ie it is independent of the symplectic manifold M. Furthermore, if $\operatorname{char} \mathbb{K} \ne 2$, intrinsic formality implies that in fact the A_∞–algebra $\bigoplus_{v,w} \mathrm{CF}^*_M(S_v, S_w)$ is quasi-isomorphic to A_Γ.

7 Conclusion

7.1 Comparison with geometric viewpoint

We would like to discuss the algebraic computations given in Section 6.2.2 in terms of the symplectic geometry of the Milnor fiber X_Γ. We shall omit some of the details, but the geometric setup that we are about to lay out is taken from [59]. Consider $\mathbb{C}^3$ with its standard symplectic form $d\alpha$, where

$$\alpha = -\tfrac{1}{4} d^c(|z_1|^2 + |z_2|^2 + |z_3|^2).$$

Let $p\colon \mathbb{C}^3 \to \mathbb{C}$ be the polynomial

$$p(z_1, z_2, z_3) = z_1^{n+1} + z_2^2 + z_3^2,$$

which has an isolated singularity at the origin of type A_n. Consider also the Hamiltonian function $H\colon \mathbb{C}^3 \to \mathbb{R}$ given by

$$H(z_1, z_2, z_3) = 2|z_1|^2 + (n+1)|z_2|^2 + (n+1)|z_3|^2.$$

Let ψ be a cutoff function such that $\psi(t^2) = 1$ for $t \leq \frac{1}{3}$ and $\psi(t^2) = 0$ for $t \geq \frac{2}{3}$. For $u \in \mathbb{C} \backslash \{0\}$ with $0 < |u| < \epsilon$ for sufficiently small ϵ, we consider the Milnor fiber

$$\{z \in \mathbb{C}^3 : p(z) = \psi(H(z))u\}.$$

For sufficiently small ϵ, this is a symplectic submanifold of $\mathbb{C}^3$ and can be symplectically identified with X_Γ. For $r \geq \frac{2}{3}$, we let $L_r = F \cap \{H = r\}$ be the link of the singularity. In other words, for such r, we have

$$L_r = \{z \in \mathbb{C}^3 : 2|z_1|^2 + (n+1)|z_2|^2 + (n+1)|z_3|^2 = r,\ p(z) = 0\}.$$

$r+s \downarrow$ $s \to$	2	1	0	−1	−2	−3	−4	−5	−6	−7	−8
2	6	0	0	0	x	0	0	0	x	0	0
1	0	0	1	0	x	0	1	0	x	0	2
0	0	0	1	0	0	0	1	0	0	0	2
−1	0	0	0	0	0	0	0	0	0	0	0

$r+s \downarrow$ $s \to$	−9	−10	−11	−12	−13	−14	−15	−16	−17	−18
2	0	0	0	0	0	0	0	0	0	0
1	0	0	0	1	0	0	0	1	0	0
0	0	x	0	1	0	x	0	1	0	$2x$
−1	0	x	0	0	0	x	0	0	0	$2x$

Table 6: $\Gamma = D_6$; x is 1 if $\operatorname{char} \mathbb{K} = 2$, 0 otherwise

For $r > 0$, L_r inherits a contact structure $\alpha|_{L_r}$ and outside of a compact set X_Γ can be identified with the positive symplectization of L_r. The appealing feature of this setup is that the Reeb vector field R_r on L_r has a periodic flow given by

$$t \cdot (z_1, z_2, z_3) = (e^{4it/r} z_1, e^{2(n+1)it/r} z_2, e^{2(n+1)it/r} z_3).$$

Thus, all the Reeb orbits are along the circle direction of a Seifert fibered structure on the lens space $L_r \cong L(n+1, n)$. Furthermore, since the Reeb flow is explicit, we can actually write down all the orbits. Let us take $Y_\Gamma = L_1$ as our contact boundary. There are two types of simple orbits:

- Generic simple orbits of period $\frac{\pi}{2n+2}\,\mathrm{lcm}(2, n+1)$. These are orbits through points $(z_1, z_2, z_3) \in Y_\Gamma$ such that $z_1 \neq 0$. The N^{th} multiple cover of these orbits have Conley–Zehnder index $2N$ if n is odd, $4N$ if n is even.
- Exceptional simple orbits of period $\frac{\pi}{n+1}$. These are orbits through points $(0, z_2, z_3) \in Y_\Gamma$. The N^{th} multiple cover of this orbit has Conley–Zehnder index $2\lfloor\frac{2N}{n+1}\rfloor + 1$ except when $2N = M(n+1)$ for some $M \in \mathbb{Z}$, in which case the index is $2M$.

For each $N \in \mathbb{Z}_+$, we can consider N–fold multiple covers of generic simple orbits together with $(n+1)N$–fold $\big($resp. $\frac{(n+1)N}{2}$–fold$\big)$ for n even (resp. n odd) multiple covers of exceptional orbits as parametrized by the manifold $L(n+1, n)$ and the N–fold cover of exceptional orbits for each $N \in \mathbb{Z}_+$ not divisible by $n+1$ $\big($resp. $\frac{n+1}{2}\big)$ for n even (resp. n odd) as parametrized by $S^1 \sqcup S^1$. This leads to a standard Morse–Bott-type spectral sequence converging to $\mathrm{SH}^*(X_\Gamma)$ (see [60] and/or [48] for a more recent exposition). For example, for $n = 2$, the E_1 page is given by

$$(11)\quad E_1^{pq} = \begin{cases} H^q(X_\Gamma; \mathbb{K}) & \text{if } p = 0, \\ H^{q-p-2}((S^1 \sqcup S^1); \mathbb{K}) & \text{if } p = 2l+1 < 0, \\ H^{q-p}(L(3,2); \mathbb{K}) \oplus H^{q-p-2}((S^1 \sqcup S^1); \mathbb{K}) & \text{if } p = 2l < 0, \\ 0 & \text{if } p > 0. \end{cases}$$

The higher differentials come from contributions of holomorphic cylinders counted in the differential of symplectic cohomology. A finite truncation of the E_1 page of this spectral sequence is shown in Table 7.

Comparing this with our results from Section 6.2.2, which correspond to a calculation of the total complex at the E_∞ page of the spectral sequence, gives us information about the holomorphic cylinders contributing to the differential of symplectic cohomology. For example, if $\operatorname{char}\mathbb{K} = 3$, the spectral sequence has to be degenerate but otherwise there has to be a nontrivial differential. See also the appendix of [48] for a similar

$r+s \downarrow \quad s \rightarrow$	2	1	0	-1	-2	-3	-4
2	2	0	0	0	0	0	0
1	2	0	0	0	0	0	0
0	0	2	1	0	0	0	0
-1	0	3	0	0	0	0	0
-2	0	0	$x+2$	0	0	0	0
-3	0	0	2	x	0	0	0
-4	0	0	0	2	1	0	0
-5	0	0	0	3	0	0	0
-6	0	0	0	0	$x+2$	0	0
-7	0	0	0	0	2	x	0
-8	0	0	0	0	0	2	1

Table 7: E_1 page of the Morse–Bott spectral sequence for $\Gamma = A_2$; x is 1 if $\operatorname{char} \mathbb{K} = 3$, 0 otherwise.

spectral sequence obtained via another natural choice of a contact form on the lens space $L(n+1, n)$.

In conclusion, even though this geometric point of view leads to an appealing description of the generators of the chain complex, it seems harder to determine the cohomology this way, let alone its multiplicative structure. However, it is reassuring that the algebraic approach taken in this paper and the geometric picture just outlined are compatible.

7.2 Generalizations

In this paper, we have studied Legendrian links $\Lambda \subset (S^3, \xi_{\mathrm{std}})$ which are obtained by plumbing Legendrian unknots according to a plumbing tree Γ. One might wonder what Koszul duality has to say when Λ is a more general Legendrian submanifold. Of course, one can study this plumbing construction in higher dimensions. Both the Ginzburg DG-algebra and the zigzag algebra have analogues corresponding to higher-dimensional plumbings, and we expect that our calculations can be extended in a straightforward way.

Perhaps a more interesting direction to pursue is the following. One of our main observations was that the Legendrian cohomology DG-algebra of Λ admits a certain natural augmentation $\epsilon\colon \mathrm{LCA}^*(\Lambda) \to \mathrm{k}$ such that

$$\mathrm{RHom}_{\mathrm{LCA}^*(\Lambda)^{\mathrm{op}}}(\mathrm{k}, \mathrm{k}) \tag{12}$$

is quasi-isomorphic to a finite-dimensional associative algebra A, whose Hochschild complex is isomorphic to that of $\mathrm{LCA}^*(\Lambda)$ by an A_∞–version of Koszul duality.

One could contemplate generalizing this construction to an arbitrary Legendrian link Λ whose $\mathrm{LCA}^*(\Lambda)$ admits an augmentation ϵ. In general, one cannot expect to have the connectedness and the finiteness conditions required in Theorem 21. Furthermore, in general, $\mathrm{LCA}^*(\Lambda)$ is not graded over $\mathbb{Z}$ but over $\mathbb{Z}/N$ for some $N > 0$. These pose important restrictions, analogous to the assumption of simple connectedness that appears in the classical story discussed in the introduction. One could partially extend Koszul duality to these more general situations if one takes completions with respect to the augmentation ideal.

References

[1] **M Abouzaid**, *A geometric criterion for generating the Fukaya category*, Publ. Math. Inst. Hautes Études Sci. 112 (2010) 191–240 MR

[2] **M Abouzaid**, *A cotangent fibre generates the Fukaya category*, Adv. Math. 228 (2011) 894–939 MR

[3] **M Abouzaid**, *A topological model for the Fukaya categories of plumbings*, J. Differential Geom. 87 (2011) 1–80 MR

[4] **M Abouzaid**, *On the wrapped Fukaya category and based loops*, J. Symplectic Geom. 10 (2012) 27–79 MR

[5] **M Abouzaid**, *Symplectic cohomology and Viterbo's theorem*, from "Free loop spaces in geometry and topology" (J Latschev, A Oancea, editors), IRMA Lect. Math. Theor. Phys. 24, Eur. Math. Soc., Zürich (2015) 271–485 MR

[6] **M Abouzaid**, **P Seidel**, *An open string analogue of Viterbo functoriality*, Geom. Topol. 14 (2010) 627–718 MR

[7] **M Abouzaid**, **I Smith**, *Exact Lagrangians in plumbings*, Geom. Funct. Anal. 22 (2012) 785–831 MR

[8] **J F Adams**, *On the cobar construction*, Proc. Nat. Acad. Sci. USA 42 (1956) 409–412 MR

[9] **M J Bardzell**, **A C Locateli**, **E N Marcos**, *On the Hochschild cohomology of truncated cycle algebras*, Comm. Algebra 28 (2000) 1615–1639 MR

[10] **A Beilinson**, **V Ginzburg**, **W Soergel**, *Koszul duality patterns in representation theory*, J. Amer. Math. Soc. 9 (1996) 473–527 MR

[11] **M van den Bergh**, *A relation between Hochschild homology and cohomology for Gorenstein rings*, Proc. Amer. Math. Soc. 126 (1998) 1345–1348 MR Correction in 130 (2002) 2809–2810

[12] **M Van den Bergh**, *Calabi–Yau algebras and superpotentials*, Selecta Math. 21 (2015) 555–603 MR

[13] **I N Bernšteĭn**, **I M Gel'fand**, **V A Ponomarev**, *Coxeter functors and Gabriel's theorem*, Uspekhi Mat. Nauk 28 (1973) 19–33 MR In Russian; translated in Russian Math. Surveys 28 (1973) 17–32

[14] **A J Blumberg**, **R L Cohen**, **C Teleman**, *Open-closed field theories, string topology, and Hochschild homology*, from "Alpine perspectives on algebraic topology" (C Ausoni, K Hess, J Scherer, editors), Contemp. Math. 504, Amer. Math. Soc., Providence, RI (2009) 53–76 MR

[15] **J M Boardman**, *Conditionally convergent spectral sequences*, from "Homotopy invariant algebraic structures" (J-P Meyer, J Morava, W S Wilson, editors), Contemp. Math. 239, Amer. Math. Soc., Providence, RI (1999) 49–84 MR

[16] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Symplectic homology product via Legendrian surgery*, Proc. Natl. Acad. Sci. USA 108 (2011) 8114–8121 MR

[17] **F Bourgeois**, **T Ekholm**, **Y Eliashberg**, *Effect of Legendrian surgery*, Geom. Topol. 16 (2012) 301–389 MR

[18] **S Brenner**, **M C R Butler**, **A D King**, *Periodic algebras which are almost Koszul*, Algebr. Represent. Theory 5 (2002) 331–367 MR

[19] **Y Chekanov**, *Differential algebra of Legendrian links*, Invent. Math. 150 (2002) 441–483 MR

[20] **R L Cohen**, **J D S Jones**, *A homotopy theoretic realization of string topology*, Math. Ann. 324 (2002) 773–798 MR

[21] **R L Cohen**, **J D S Jones**, **J Yan**, *The loop homology algebra of spheres and projective spaces*, from "Categorical decomposition techniques in algebraic topology" (G Arone, J Hubbuck, R Levi, M Weiss, editors), Progr. Math. 215, Birkhäuser, Basel (2004) 77–92 MR

[22] **P M Cohn**, *Free ideal rings and localization in general rings*, New Mathematical Monographs 3, Cambridge Univ. Press (2006) MR

[23] **W Crawley-Boevey**, **P Etingof**, **V Ginzburg**, *Noncommutative geometry and quiver algebras*, Adv. Math. 209 (2007) 274–336 MR

[24] **W Crawley-Boevey**, **P Shaw**, *Multiplicative preprojective algebras, middle convolution and the Deligne–Simpson problem*, Adv. Math. 201 (2006) 180–208 MR

[25] **P Deligne**, **P Griffiths**, **J Morgan**, **D Sullivan**, *Real homotopy theory of Kähler manifolds*, Invent. Math. 29 (1975) 245–274 MR

[26] **G Dimitroglou Rizell**, *Nontriviality results for the characteristic algebra of a DGA*, Math. Proc. Cambridge Philos. Soc. 162 (2017) 419–433 MR

[27] **T Ekholm**, **Y Lekili**, *Duality between Lagrangian and Legendrian invariants*, preprint (2017) arXiv

[28] **T Ekholm**, **L Ng**, *Legendrian contact homology in the boundary of a subcritical Weinstein 4–manifold*, J. Differential Geom. 101 (2015) 67–157 MR

[29] **Y Eliashberg**, *Topological characterization of Stein manifolds of dimension > 2*, Internat. J. Math. 1 (1990) 29–46 MR

[30] **Y Eliashberg**, **A Givental**, **H Hofer**, *Introduction to symplectic field theory*, from "Visions in mathematics: GAFA special volume" (N Alon, J Bourgain, A Connes, M Gromov, V Milman, editors), volume II, Birkhäuser (2000) 560–673 MR

[31] **Y Eliashberg**, **M Gromov**, *Convex symplectic manifolds*, from "Several complex variables and complex geometry, 2" (E Bedford, J P D'Angelo, R E Greene, S G Krantz, editors), Proc. Sympos. Pure Math. 52, Amer. Math. Soc., Providence, RI (1991) 135–162 MR

[32] **K Erdmann**, **T Holm**, *Twisted bimodules and Hochschild cohomology for self-injective algebras of class* A_n, Forum Math. 11 (1999) 177–201 MR

[33] **T Etgü**, **Y Lekili**, *Fukaya categories of plumbings and multiplicative preprojective algebras*, preprint (2017) arXiv

[34] **P Etingof**, **V Ginzburg**, *Noncommutative complete intersections and matrix integrals*, Pure Appl. Math. Q. 3 (2007) 107–151 MR

[35] **P Etingof**, **V Ginzburg**, *Noncommutative del Pezzo surfaces and Calabi–Yau algebras*, J. Eur. Math. Soc. 12 (2010) 1371–1416 MR

[36] **Y Félix**, **L Menichi**, **J-C Thomas**, *Gerstenhaber duality in Hochschild cohomology*, J. Pure Appl. Algebra 199 (2005) 43–59 MR

[37] **H Geiges**, *An introduction to contact topology*, Cambridge Studies in Advanced Mathematics 109, Cambridge Univ. Press (2008) MR

[38] **M Gerstenhaber**, **C W Wilkerson**, *On the deformation of rings and algebras, V: Deformation of differential graded algebras*, from "Higher homotopy structures in topology and mathematical physics" (J McCleary, editor), Contemp. Math. 227, Amer. Math. Soc., Providence, RI (1999) 89–101 MR

[39] **V Ginzburg**, *Calabi–Yau algebras*, unpublished manuscript (2006) arXiv

[40] **R E Gompf**, *Handlebody construction of Stein surfaces*, Ann. of Math. 148 (1998) 619–693 MR

[41] **S Hermes**, *Minimal model of Ginzburg algebras*, J. Algebra 459 (2016) 389–436 MR

[42] **T Holm**, *Hochschild cohomology of Brauer tree algebras*, Comm. Algebra 26 (1998) 3625–3646 MR

[43] **R S Huerfano**, **M Khovanov**, *A category for the adjoint representation*, J. Algebra 246 (2001) 514–542 MR

[44] **J D S Jones**, **J McCleary**, *Hochschild homology, cyclic homology, and the cobar construction*, from "Adams Memorial Symposium on Algebraic Topology, 1" (N Ray, G Walker, editors), London Math. Soc. Lecture Note Ser. 175, Cambridge Univ. Press (1992) 53–65 MR

[45] **T V Kadeishvili**, *The structure of the* $A(\infty)$*–algebra, and the Hochschild and Harrison cohomologies*, Trudy Tbiliss. Mat. Inst. Razmadze Akad. Nauk Gruzin. SSR 91 (1988) 19–27 MR In Russian

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

[47] **B Keller**, *Derived invariance of higher structures on the Hochschild complex*, unpublished manuscript (2003) Available at `https://webusers.imj-prg.fr/~bernhard.keller/publ/dih.pdf`

[48] **M Kwon**, **O van Koert**, *Brieskorn manifolds in contact topology*, Bull. Lond. Math. Soc. 48 (2016) 173–241 MR

[49] **Y Lekili**, **M Maydanskiy**, *The symplectic topology of some rational homology balls*, Comment. Math. Helv. 89 (2014) 571–596 MR

[50] **J-L Loday**, *Cyclic homology*, Grundl. Math. Wissen. 301, Springer (1992) MR

[51] **D M Lu**, **J H Palmieri**, **Q S Wu**, **J J Zhang**, *Koszul equivalences in A_∞–algebras*, New York J. Math. 14 (2008) 325–378 MR

[52] **R Martínez-Villa**, *Applications of Koszul algebras: the preprojective algebra*, from "Representation theory of algebras" (R Bautista, R Martínez-Villa, J A de la Peña, editors), CMS Conf. Proc. 18, Amer. Math. Soc., Providence, RI (1996) 487–504 MR

[53] **L Menichi**, *String topology for spheres*, Comment. Math. Helv. 84 (2009) 135–157 MR

[54] **S B Priddy**, *Koszul resolutions*, Trans. Amer. Math. Soc. 152 (1970) 39–60 MR

[55] **J Rickard**, *Derived categories and stable equivalence*, J. Pure Appl. Algebra 61 (1989) 303–317 MR

[56] **T Schedler**, *Zeroth Hochschild homology of preprojective algebras over the integers, v1* (2007) arXiv

[57] **T Schedler**, *Hochschild homology of preprojective algebras over the integers*, Adv. Math. 299 (2016) 451–542 MR

[58] **E Segal**, *The A_∞ deformation theory of a point and the derived categories of local Calabi–Yaus*, J. Algebra 320 (2008) 3232–3268 MR

[59] **P Seidel**, *Graded Lagrangian submanifolds*, Bull. Soc. Math. France 128 (2000) 103–149 MR

[60] **P Seidel**, *A biased view of symplectic cohomology*, from "Current developments in mathematics" (B Mazur, T Mrowka, W Schmid, R Stanley, S-T Yau, editors), Int. Press, Somerville, MA (2008) 211–253 MR

[61] **P Seidel**, *Fukaya categories and Picard–Lefschetz theory*, Eur. Math. Soc., Zürich (2008) MR

[62] **P Seidel**, *Symplectic homology as Hochschild homology*, from "Algebraic geometry, 1" (D Abramovich, A Bertram, L Katzarkov, R Pandharipande, M Thaddeus, editors), Proc. Sympos. Pure Math. 80, Amer. Math. Soc., Providence, RI (2009) 415–434 MR

[63] **P Seidel**, *Lectures on categorical dynamics and symplectic topology*, MIT (2013) Available at `http://www-math.mit.edu/~seidel/937/lecture-notes.pdf`

[64] **P Seidel**, *Abstract analogues of flux as symplectic invariants*, Mém. Soc. Math. Fr. 137, Soc. Math. Fr., Paris (2014) MR

[65] **P Seidel**, *Fukaya A_∞–structures associated to Lefschetz fibrations,* II 1/2, preprint (2015) arXiv

[66] **P Seidel**, *Picard–Lefschetz theory and dilating $\mathbb{C}^*$–actions*, J. Topol. 8 (2015) 1167–1201 MR

[67] **P Seidel**, **R Thomas**, *Braid group actions on derived categories of coherent sheaves*, Duke Math. J. 108 (2001) 37–108 MR

[68] **S F Siegel**, **S J Witherspoon**, *The Hochschild cohomology ring of a group algebra*, Proc. London Math. Soc. 79 (1999) 131–157 MR

[69] **I Smith**, *Quiver algebras as Fukaya categories*, Geom. Topol. 19 (2015) 2557–2617 MR

[70] **C Viterbo**, *Functors and computations in Floer homology with applications, I*, Geom. Funct. Anal. 9 (1999) 985–1033 MR

[71] **Yu V Volkov**, *The Hochschild cohomology algebra for a series of self-injective algebras of tree type D_n*, Algebra i Analiz 23 (2011) 99–139 MR In Russian; translated in St. Petersburg Math. J. 23 (2012) 851–879

[72] **Yu V Volkov**, **A I Generalov**, *Hochschild cohomology for self-injective algebras of tree class D_n, I*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. 343 (2007) 121–182, 273–274 MR In Russian; translated in J. Math. Sci. (N. Y.) 147 (2007) 7042–7073

[73] **A Weinstein**, *Contact surgery and symplectic handlebodies*, Hokkaido Math. J. 20 (1991) 241–251 MR

Department of Mathematics, Koç University
34450 Istanbul, Turkey

Department of Mathematics, King's College London
London, WC2R 2LS, United Kingdom

tetgu@ku.edu.tr, yanki.lekili@kcl.ac.uk

Proposed: Yasha Eliashberg Received: 10 April 2015
Seconded: Ciprian Manolescu, András I. Stipsicz Revised: 13 September 2016

Geometry & Topology 21 (2017) 3391–3451

A complex hyperbolic Riley slice

John R Parker
Pierre Will

We study subgroups of $\mathrm{PU}(2,1)$ generated by two noncommuting unipotent maps A and B whose product AB is also unipotent. We call $\mathcal{U}$ the set of conjugacy classes of such groups. We provide a set of coordinates on $\mathcal{U}$ that make it homeomorphic to $\mathbb{R}^2$. By considering the action on complex hyperbolic space $\boldsymbol{H}^2_{\mathbb{C}}$ of groups in $\mathcal{U}$, we describe a two-dimensional disc $\mathcal{Z}$ in $\mathcal{U}$ that parametrises a family of discrete groups. As a corollary, we give a proof of a conjecture of Schwartz for $(3,3,\infty)$–triangle groups. We also consider a particular group on the boundary of the disc $\mathcal{Z}$ where the commutator $[A,B]$ is also unipotent. We show that the boundary of the quotient orbifold associated to the latter group gives a spherical CR uniformisation of the Whitehead link complement.

20H10, 22E40, 51M10; 57M50

1 Introduction

1.1 Context and motivation

The framework of this article is the study of the deformations of a discrete subgroup Γ of a Lie group H in a Lie group G containing H. This question has been addressed in many different contexts. A classical example is the one where Γ is a Fuchsian group, $H=\mathrm{PSL}(2,\mathbb{R})$ and $G=\mathrm{PSL}(2,\mathbb{C})$. When Γ is discrete, such deformations are called quasi-Fuchsian. We will be interested in the case where Γ is a discrete subgroup of $H=\mathrm{SO}(2,1)$ and G is the group $\mathrm{SU}(2,1)$ (or their natural projectivisations over $\mathbb{R}$ and $\mathbb{C}$, respectively). The geometrical motivation is very similar: In the classical case mentioned above, $\mathrm{PSL}(2,\mathbb{C})$ is the orientation-preserving isometry group of hyperbolic 3–space $\boldsymbol{H}^3$ and a Fuchsian group preserves a totally geodesic hyperbolic plane $\boldsymbol{H}^2$ in $\boldsymbol{H}^3$. In our case $G=\mathrm{SU}(2,1)$ is (a triple cover of) the holomorphic isometry group of complex hyperbolic 2–space $\boldsymbol{H}^2_{\mathbb{C}}$, and the subgroup $H=\mathrm{SO}(2,1)$ preserves a totally geodesic Lagrangian plane isometric to $\boldsymbol{H}^2$. A discrete subgroup Γ of $\mathrm{SO}(2,1)$ is called $\mathbb{R}$–Fuchsian. A second example of this construction is where G is again $\mathrm{SU}(2,1)$ but now $H=S(U(1)\times U(1,1))$. In this case H preserves a totally geodesic complex line in $\boldsymbol{H}^2_{\mathbb{C}}$. A discrete subgroup of H is called $\mathbb{C}$–Fuchsian. Deformations of

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3391

either $\mathbb{R}$–Fuchsian or $\mathbb{C}$–Fuchsian groups in $\mathrm{SU}(2,1)$ are called complex hyperbolic quasi-Fuchsian. See Parker and Platis [24] for a survey of this topic.

The title of this article refers to the so-called *Riley slice of Schottky space* (see [19; 1]). Riley considered the space of conjugacy classes of subgroups of $\mathrm{PSL}(2,\mathbb{C})$ generated by two noncommuting parabolic maps. This space may be identified with $\mathbb{C}-\{0\}$ under the map that associates the parameter $\rho\in\mathbb{C}-\{0\}$ with the conjugacy class of the group Γ_ρ, where

$$\Gamma_\rho=\left\langle\begin{bmatrix}1&1\\0&1\end{bmatrix},\begin{bmatrix}1&0\\\rho&1\end{bmatrix}\right\rangle.$$

Riley was interested in the set of those parameters ρ for which Γ_ρ is discrete. He was particularly interested in the (closed) set where Γ_ρ is discrete and free, which is now called the Riley slice of Schottky space; see Keen and Series [19]. This work has been taken up more recently by Akiyoshi, Sakuma, Wada and Yamashita. In their book [1] they illustrate one of Riley's original computer pictures,[1] Figure 0.2a, and their version of this picture, Figure 0.2b. Riley's main method was to construct the Ford domain for Γ_ρ. The different combinatorial patterns that arise in this Ford domain correspond to the differently coloured regions in these figures from [1]. Riley was also interested in groups Γ_ρ that are discrete but not free. In particular, he showed that when ρ is a complex sixth root of unity then the quotient of hyperbolic 3–space by Γ_ρ is the figure-eight knot complement.

1.2 Main definitions and discreteness result

The direct analogue of the Riley slice in complex hyperbolic plane would be the set of conjugacy classes of groups generated by two noncommuting, unipotent parabolic elements A and B of $\mathrm{SU}(2,1)$. (Note that in contrast to $\mathrm{PSL}(2,\mathbb{C})$, there exist parabolic elements in $\mathrm{SU}(2,1)$ that are not unipotent. In fact, there is a 1–parameter family of parabolic conjugacy classes; see for instance Goldman [15, Chapter 6].) This choice would give a four-dimensional parameter space, and we require additionally that AB be unipotent, making the dimension drop to 2. Specifically, we define

(1) $$\mathcal{U}=\{(A,B)\in\mathrm{SU}(2,1)^2 : A,B,AB \text{ all unipotent and } AB\neq BA\}/\mathrm{SU}(2,1).$$

Following Riley, we are interested in the (closed) subset of $\mathcal{U}$ where the group $\langle A,B\rangle$ is discrete and free and our main method for studying this set is to construct the Ford domain for its action on complex hyperbolic space $\boldsymbol{H}^2_{\mathbb{C}}$. We shall also indicate various other interesting discrete groups in $\mathcal{U}$ but these will not be our main focus.

[1] Parker has one of Riley's printouts of this picture dated 26th March 1979.

In Section 3.1, we will parametrise $\mathcal{U}$ so that it becomes the open square $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)^2$. The parameters we use will be the Cartan angular invariants α_1 and α_2 of the triples of (parabolic) fixed points of (A, AB, B) and (A, AB, BA), respectively (see Section 2.6 for the definitions). Note that the invariants α_1 and α_2 are defined to lie in the closed interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Our assumption that A and B don't commute implies that neither α_1 nor α_2 can equal $\pm\frac{\pi}{2}$ (see Section 3.1).

When α_1 and α_2 are both zero, that is, at the origin of the square, the group $\langle A, B\rangle$ is $\mathbb{R}$–Fuchsian. The quotient of the Lagrangian plane preserved by $\langle A, B\rangle$ is a hyperbolic thrice-punctured sphere where the three (homotopy classes of) peripheral elements are represented by (the conjugacy classes of) A, B and AB. The space $\mathcal{U}$ can thus be thought of as the slice of the $\mathrm{SU}(2,1)$–representation variety of the thrice-punctured sphere group defined by the conditions that the peripheral loops are mapped to unipotent isometries.

We can now state our main discreteness result.

Theorem 1.1 *Suppose that $\Gamma = \langle A, B\rangle$ is the group associated to parameters (α_1, α_2) satisfying $\mathcal{D}(4\cos^2\alpha_1, 4\cos^2\alpha_2) > 0$, where $\mathcal{D}$ is the polynomial given by*

$$\mathcal{D}(x, y) = x^3y^3 - 9x^2y^2 - 27xy^2 + 81xy - 27x - 27.$$

Then Γ is discrete and isomorphic to the free group F_2. This region is $\mathcal{Z}$ in Figure 1.

Note that at the centre of the square, we have $\mathcal{D}(4,4) = 1225$ for the $\mathbb{R}$–Fuchsian representation. The region $\mathcal{Z}$ where $\mathcal{D} > 0$ consists of groups Γ whose Ford domain has the simplest possible combinatorial structure. It is the analogue of the outermost region in the two figures from Akiyoshi, Sakuma, Wada and Yamashita [1] mentioned above.

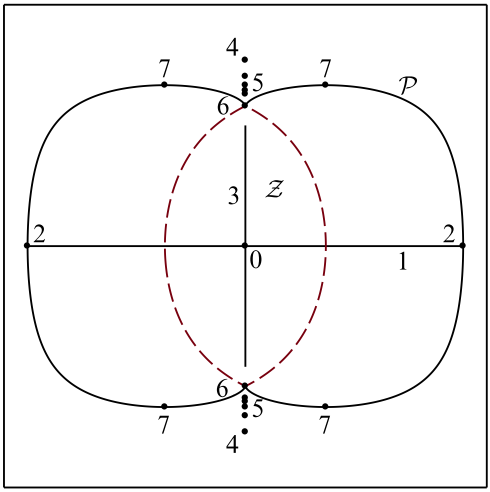
1.3 Decompositions and triangle groups

We will prove in Proposition 3.3 that all pairs (A, B) in $\mathcal{U}$ admit a (unique) decomposition of the form

$$A = ST \quad \text{and} \quad B = TS, \tag{2}$$

where S and T are order-three regular elliptic elements (see Section 2.2). In turn, the group generated by A and B has index three in the one generated by S and T. When either $\alpha_1 = 0$ or $\alpha_2 = 0$ there is a further decomposition making $\langle A, B\rangle$ a subgroup of a triangle group.

Deformations of triangle groups in $\mathrm{PU}(2,1)$ have been considered in many places, including Goldman and Parker [16], Parker, Wang and Xie [25], Pratoussevitch [28] and Schwartz [32]. A complex hyperbolic (p,q,r)–triangle is one generated by three complex involutions about (complex) lines with pairwise angles $\frac{\pi}{p}$, $\frac{\pi}{q}$, and $\frac{\pi}{r}$, where



0 $\mathbb{R}$–Fuchsian representation of the 3–punctured sphere group.

1 Horizontal segment corresponding to even word subgroups of ideal triangle groups; see Goldman and Parker [16] and Schwartz [30; 31; 33].

2 Last ideal triangle group, contained with index three in a group uniformising the Whitehead link complement obtained by Schwartz [30; 31; 33].

3 Vertical segment corresponding to bending groups that have been proved to be discrete by Will [37].

4 $(3,3,4)$–group uniformising the figure-eight knot complement. Obtained by Deraux and Falbel [8].

5 $(3,3,n)$–groups, proved to be discrete by Parker, Wang and Xie [25]. On this picture, $4 \leqslant n \leqslant 8$.

6 Uniformisation of the Whitehead link complement we obtain in this work.

7 Subgroup of the Eisenstein–Picard lattice; see Falbel and Parker [14].

Figure 1: The parameter space for $\mathcal{U}$. The exterior curve $\mathcal{P}$ corresponds to classes of groups for which $[A,B]$ is parabolic. The central dashed curve bounds the region $\mathcal{Z}$ where we prove discreteness. The labels correspond to various special values of the parameters. Points with the same labels are obtained from one another by symmetries about the coordinate axes. The results of Section 3.3 imply that they correspond to groups conjugate in $\mathrm{Isom}(\boldsymbol{H}^2_{\mathbb{C}})$.

p, q and r are integers or ∞ (when one of them is ∞ the corresponding angle is 0). Groups generated by complex reflections of higher order are also interesting; see Mostow [22] for example, but we do not consider them here. For a given triple (p,q,r) with $\min\{p,q,r\} \geqslant 3$, the deformation space of the (p,q,r)–triangle group is one-dimensional, and can be thought of as the deformation space of the $\mathbb{R}$–Fuchsian triangle group. Schwartz [32] develops a series of conjectures about which points in this space yield discrete and faithful representations of the triangle group. For a given triple (p,q,r), Conjecture 5.1 of [32] states that a complex hyperbolic (p,q,r)–triangle group is a discrete and faithful representation of the Fuchsian one if and only if the words $I_iI_jI_k$ and $I_iI_jI_kI_j$ (with i, j and k pairwise distinct) are nonelliptic. Moreover, depending on p, q and r, he predicts which of these words one should choose.

We now explain the relationship between triangle groups and groups on the axes of our parameter space $\mathcal{U}$. First consider groups with $\alpha_2 = 0$. Let I_1, I_2 and I_3 be the involutions fixing the complex lines spanned by the fixed points of $(A,\ B)$, of $(A,\ AB)$ and of $(B,\ AB)$, respectively. If $\alpha_2 = 0$ then A and B may be decomposed as $A = I_2I_1$ and $B = I_1I_3$, and also $\langle A,\ B\rangle$ has index 2 in $\langle I_1, I_2, I_3\rangle$ (Proposition 3.8). Since $I_2I_1 = A$, $I_1I_3 = B$ and $I_2I_3 = AB$ are all unipotent, we see that $\langle I_1, I_2, I_3\rangle$ is a complex hyperbolic ideal triangle group, as studied by Goldman and Parker [16] and Schwartz [30; 31; 33]. Their results gave a complete characterisation of when such a group is discrete. (Our Cartan invariant α_1 is the same as the Cartan invariant $\mathbb{A}$ used in these papers.)

Theorem 1.2 [16; 31; 33] *Let I_1, I_2 and I_3 be complex involutions fixing distinct, pairwise asymptotic complex lines. Let $\mathbb{A}$ be the Cartan invariant of the fixed points of I_1I_2, I_2I_3 and I_3I_1.*

(1) *The group $\langle I_1, I_2, I_3\rangle$ is a discrete and faithful representation of an (∞,∞,∞)–triangle group if and only if $I_1I_2I_3$ is nonelliptic. This happens when*
$$|\mathbb{A}| \le \arccos \sqrt{3/128}.$$

(2) *When $I_1I_2I_3$ is elliptic the group is not discrete. In this case,*
$$\arccos \sqrt{3/128} < |\mathbb{A}| < \tfrac{\pi}{2}.$$

When $\alpha_1 = 0$ we get an analogous result. In this case, it is the order-three maps S and T from (2) which decompose into products of complex involutions. Namely, if $\alpha_1 = 0$, there exist three involutions I_1, I_2 and I_3, each fixing a complex line, such that $S = I_2I_1$ and $T = I_1I_3$ have order 3 and $ST = A = I_2I_3$ is unipotent (Proposition 3.8). Furthermore, writing $B = TS = I_1I_3I_2I_1$ we have $[A,B] = (ST^{-1})^3 = (I_2I_1I_3I_1)^3$. A corollary of Theorem 1.1 is a statement analogous to Theorem 1.2 for $(3,3,\infty)$–triangle groups, proving a special case of Schwartz [32, Conjecture 5.1]. Compare with the proof of this conjecture for $(3,3,n)$–triangle groups given by Parker, Wang and Xie [25].

Theorem 1.3 *Let I_1, I_2 and I_3 be complex involutions fixing distinct complex lines and such that $S = I_2 I_1$ and $T = I_1 I_3$ have order three and $A = ST = I_2 I_3$ is unipotent. Let $\mathbb{A}$ be the Cartan invariant of the fixed points of A, SAS^{-1} and $S^{-1}AS$. The group $\langle I_1, I_2, I_3\rangle$ is a discrete and faithful representation of the $(3,3,\infty)$–triangle group if and only if $I_2 I_1 I_3 I_1 = ST^{-1}$ is nonelliptic. This happens when*

$$|\mathbb{A}| \leq \arccos\sqrt{3/8}.$$

Theorem 1.3 follows directly from Theorem 1.1 by restricting it to the case where $(\alpha_1, \alpha_2) = (0, \mathbb{A})$. These groups are a special case of those studied by Will [37] from a different point of view. There, using bending, he proved that these groups are discrete as long as $|\mathbb{A}| = |\alpha_2| \leq \frac{\pi}{4}$. The gap between the vertical segment in Figure 1 and the curve where $[A, B]$ is parabolic illustrates the nonoptimality of the result of [37].

1.4 Spherical CR uniformisations of the Whitehead link complement

The quotient of $\boldsymbol{H}^2_{\mathbb{C}}$ by an $\mathbb{R}$– or $\mathbb{C}$–Fuchsian punctured surface group is a disc bundle over the surface. If the surface is noncompact, this bundle is trivial. Its boundary at infinity is a circle bundle over the surface. Such three-manifolds appearing on the boundary at infinity of quotients of $\boldsymbol{H}^2_{\mathbb{C}}$ are naturally equipped with a *spherical CR structure*, which is the analogue of the flat conformal structure in the real hyperbolic case. These structures are examples of (X, G)–structure, with $X = S^3 = \partial\boldsymbol{H}^2_{\mathbb{C}}$ and $G = \mathrm{PU}(2,1)$. To any such structure on a three-manifold M are associated a holonomy representation $\rho\colon \pi_1(M) \to \mathrm{PU}(2,1)$ and a developing map $D = \widetilde{M} \to X$. This motivates the study of representations of fundamental groups of hyperbolic three-manifolds in $\mathrm{PU}(2,1)$ and $\mathrm{PGL}(3,\mathbb{C})$ initiated by Falbel [11], and continued by Falbel, Guilloux, Koseleff, Rouillier and Thistlethwaite [12; 13] (see also Heusener, Munoz and Porti [18]). Among $\mathrm{PU}(2,1)$ representations, *uniformisations* (see Deraux [6, Definition 1.3]) are of special interest. There, the manifold at infinity is the quotient of the discontinuity region by the group action.

For parameter values in the open region $\mathcal{Z}$, the manifold at infinity of $\boldsymbol{H}^2_{\mathbb{C}}/\langle S, T\rangle$ is a Seifert fibre space over a $(3,3,\infty)$–orbifold. This is obviously true in the case where $\alpha_1 = \alpha_2 = 0$ (the central point on Figure 1). Indeed, for these values the group $\langle S, T\rangle$ preserves $\boldsymbol{H}^2_{\mathbb{R}}$ (it is $\mathbb{R}$–Fuchsian) and the fibres correspond to boundaries of real planes orthogonal to $\boldsymbol{H}^2_{\mathbb{R}}$. As the combinatorics of our fundamental domain remains unchanged in $\mathcal{Z}$, the topology of the quotient is constant in $\mathcal{Z}$.

Things become interesting if we deform the group in such a way that a loop on the surface is represented by a parabolic map: the topology of the manifold at infinity can change. A hyperbolic manifold arising in this way was first constructed by Schwartz:

Theorem 1.4 [30] *Let I_1, I_2 and I_3 be as in Theorem 1.2. Let $\mathbb{A}$ be the Cartan invariant of the fixed points of $I_1 I_2$, $I_2 I_3$ and $I_3 I_1$ and let S be the regular elliptic map cyclically permuting these points. When $I_1 I_2 I_3$ is parabolic, the quotient of $\boldsymbol{H}^2_{\mathbb{C}}$ by the group $\langle I_1 I_2, S\rangle$ is a complex hyperbolic orbifold with isolated singularities whose boundary at infinity is a spherical CR uniformisation of the Whitehead link complement. These groups have Cartan invariant $\mathbb{A} = \pm \arccos \sqrt{3/128}$.*

Schwartz's example provides a uniformisation of the Whitehead link complement. More recently, Deraux and Falbel [8] described a uniformisation of the complement of the figure-eight knot. Deraux [7] proved that this uniformisation was flexible: he described a one-parameter deformation of the uniformisation described by Deraux and Falbel [8], each group in the deformation being a uniformisation of the figure-eight knot complement.

Our second main result concerns the $(3,3,\infty)$–triangles group from Theorem 1.3, and it states that when $I_2 I_1 I_3 I_1$ is parabolic the associated groups give a uniformisation of the Whitehead link complement which is different from Schwartz's one. Indeed in our case the cusps of the Whitehead link complement both have unipotent holonomy. In Schwartz's case, one of them is unipotent whereas the other is screw-parabolic. The representation of the Whitehead link group we consider here was identified from a different point of view by Falbel, Koseleff and Rouillier [13, page 254] in their census of $\mathrm{PGL}(3,\mathbb{C})$ representations of knot and link complement groups.

Theorem 1.5 *Let I_1, I_2 and I_3 be as in Theorem 1.3 and define $S = I_2 I_1$ and $A = I_2 I_3$. Let $\mathbb{A}$ be the Cartan invariant of the fixed points of A, SAS^{-1} and $S^{-1}AS$. When $I_2 I_1 I_3 I_1$ is parabolic, the quotient of $\boldsymbol{H}^2_{\mathbb{C}}$ by $\langle A, S\rangle$ is a complex hyperbolic orbifold with isolated singularities whose boundary is a spherical CR uniformisation of the Whitehead link complement. These groups have Cartan invariant $\mathbb{A} = \pm \arccos \sqrt{3/8}$.*

Schwartz's uniformisation of the Whitehead link complement corresponds to each of the endpoints of the horizontal segment, marked 2 in Figure 1, and our uniformisation corresponds to each of the points on the vertical axis, marked 6 in that figure.

It should be noted that the image of the holonomy representation of our uniformisation of the Whitehead link complement is the group generated by S and T, which is isomorphic to $\mathbb{Z}_3 * \mathbb{Z}_3$. We note in Proposition 3.4 that the fundamental group of the Whitehead link complement surjects onto $\mathbb{Z}_3 * \mathbb{Z}_3$. Furthermore, the group $\mathbb{Z}_3 * \mathbb{Z}_3$ is the fundamental group of the (double) Dehn filling of the Whitehead link complement with slope -3 at each cusp in the standard marking (the same as in SnapPy). This Dehn filling is nonhyperbolic, as can be easily verified using the software SnapPy [5] (it also follows from Martelli and Petronio [20, Theorem 1.3]). This fact should be compared with Deraux's remark in [6] that all known examples of noncompact finite volume hyperbolic manifold admitting a spherical CR uniformisation also admit an exceptional Dehn filling which is a Seifert fibre space over a (p,q,r)–orbifold with $p,q,r \geqslant 3$.

1.5 Ideas of the proofs

Proof of Theorem 1.1 The rough idea of this proof is to construct fundamental domains for the groups corresponding to parameters in the region $\mathcal{Z}$. To this end, we construct their *Ford domains*, which can be thought of as a fundamental domain for a coset decomposition of the group with respect to a parabolic element (here, this element is $A = ST$). The Ford domain is invariant by the subgroup generated by A and we obtain a fundamental domain for the group by intersecting the Ford domain with a fundamental domain for the subgroup generated by A. The sides of the Ford domain are built out of pieces of *isometric spheres* of various group elements (see Sections 2.4 and 4) This method is classical, and is described in the case of the Poincaré disc in of Beardon [2, Section 9.6].

We thus have to consider a 2–parameter family of such polyhedra, and the polynomial $\mathcal{D}$ controls the combinatorial complexity of the Ford domain within our parameter space for $\mathcal{U}$ in the following sense. The null-locus of $\mathcal{D}$ is depicted on Figure 1 as a dashed curve, which bounds the region $\mathcal{Z}$. In the interior of this curve, the combinatorics of our domain is constant, and stays the same as it is for the $\mathbb{R}$–Fuchsian group. On the boundary of $\mathcal{Z}$ the isometric spheres of the elements S, S^{-1} and T have a common point. More precisely, the isometric spheres of S^{-1} and T intersect for all values of α_1 and α_2, but inside $\mathcal{Z}$ their intersection is contained in one of the two connected components of the complement of the isometric sphere of S in $\boldsymbol{H}^2_{\mathbb{C}}$. When one reaches the boundary curve of $\mathcal{Z}$, one of their intersection points lies on the isometric sphere of S.

We believe that it should be possible to mimic Riley's approach and to construct regions in our parameter space where the Ford domain is more complicated. However, as with Riley's work, this may only be accessible via computer experiments.

Proof of Theorem 1.5 The groups where $[A,B] = (I_2 I_1 I_3 I_1)^3$ is parabolic are the focus of Section 6 and Theorem 1.5 will follow from Theorem 6.4. In order to prove this result, we analyse in detail our fundamental domain, and show that it gives the classical description of the Whitehead link complement from an ideal octahedron equipped with face identifications. The Whitehead link is depicted in Figure 2. We refer to Ratcliffe [29, Section 10.3] and Thurston [35, Section 3.3] for classical information about the topology of the Whitehead link complement and its hyperbolic structure.

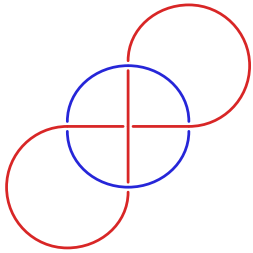


Figure 2: The Whitehead link

1.6 Further remarks

Other discrete groups appearing in $\mathcal{U}$ As well as the ideal triangle groups and bending groups discussed above, there are some other previously studied discrete groups in this family. We give them in (α_1, α_2) coordinates and illustrate them in Figure 1.

(1) The groups corresponding to $\alpha_1 = 0$ and $\alpha_2 = \pm \arccos \sqrt{1/8}$ have been studied in great detail by Deraux and Falbel [8], who proved that they give a spherical CR uniformisation of the figure-eight knot complement. This illustrates the fact that there is no statement for Theorem 1.3 analogous to the second part of Theorem 1.2: the group from [8] is contained in a discrete (nonfaithful) $(3, 3, \infty)$–triangle group where $I_2 I_1 I_3 I_1$ is elliptic.

(2) The groups with parameters $\alpha_1 = 0$ and for which ST^{-1} has order n correspond to the $(3, 3, n)$–triangle groups studied by Parker, Wang and Xie [25]. The corresponding value of α_2 is given by

$$\alpha_2 = \pm \arccos \sqrt{\tfrac{1}{8}\bigl(4\cos^2\bigl(\tfrac{\pi}{n}\bigr) - 1\bigr)}.$$

(3) The groups where $\alpha_1 = \pm\frac{\pi}{6}$ and $\alpha_2 = \pm\frac{\pi}{3}$ are discrete, since they are subgroups of the Eisenstein–Picard lattice $\mathrm{PU}(2, 1; \mathbb{Z}[\omega])$, where ω is a cube root of unity. That lattice has been studied by Falbel and Parker [14].

Comparison with the classical Riley slice There is, conjecturally, one extremely significant difference between the classical Riley slice and our complex hyperbolic version. The boundary of the classical Riley slice is not a smooth curve and has a dense set of points where particular group elements are parabolic (see for instance the beautiful picture in the introduction of Keen and Series [19]). On the other hand, we believe that in the complex hyperbolic case, discreteness is completely controlled by the commutator $[A, B]$, or equivalently ST^{-1}, as is true for the two cases where

$\alpha_1 = 0$ or $\alpha_2 = 0$ described above. If this is true, then the boundary of the set of (classes of) discrete and faithful representations in $\mathrm{SU}(2,1)$ of the three punctured sphere group with unipotent peripheral holonomy is piecewise smooth, and it is given by the simple closed curve $\mathcal{P}$ in Figure 1. This curve provides a one-parameter family of (conjecturally discrete) representations that connects Schwartz's uniformisation of the Whitehead link complement to ours. We believe that all these representations give uniformisations of the Whitehead link complement as well, but we are not able to prove this with our techniques. What seems to happen is that if one deforms our uniformisation by following the curve $\mathcal{P}$, the number of isometric spheres contributing to the boundary at infinity of the Ford domain becomes too large to be understood using our techniques. Possibly, this is because deformations of fundamental domains with tangencies between bisectors are complicated. This should be compared to Deraux's construction [7] of deformations of the figure-eight knot complement mentioned above. There, he had to use a different domain to the one by Deraux and Falbel [8], which also has tangencies between the bisectors.

1.7 Organisation of the article

This article is organised as follows. In Section 2 we present the necessary background facts on complex hyperbolic space and its isometries. In Section 3, we describe coordinates on the space of (conjugacy classes of) group generated by two unipotent isometries with unipotent product. Section 4 is devoted to the description of the isometric spheres that bound our fundamental domains. We state and apply the Poincaré polyhedron theorem in Section 5. In Section 6, we focus on the specific case where the commutator becomes parabolic, and prove that the corresponding manifold at infinity is homeomorphic to the complement of the Whitehead link. In Section 7, we give the technical proofs which we have omitted for readability in the earlier sections.

Acknowledgements The authors would like to thank Miguel Acosta, Martin Deraux, Elisha Falbel and Antonin Guilloux for numerous interesting discussions. Will thanks Craig Hodgson, Neil Hoffman and Chris Leininger for kindly answering his naïve questions. This research was financially supported by ANR SGT and an LMS Scheme 2 grant. The research took place during visits of both authors to Les Diablerets, Durham, Grenoble, Hunan University, ICTP and Luminy, and we would like to thank all these institutions for their hospitality. We are very grateful to the referee for their careful report and many helpful suggestions for improvement of the paper. Will had the pleasure to share a very useful discussion with Lucien Guillou on a sunny June afternoon in Grenoble. Lucien has since passed away, and we remember him with affection.

2 Preliminary material

Throughout we will work in the complex hyperbolic plane using a projective model and will therefore pass from projective objects to lifts of them. Our convention is that the same letter will be used to denote a point in $\mathbb{C}P^2$ and a lift of it to $\mathbb{C}^3$ with a bold font for the lift. As an example, each time p is a point of $\boldsymbol{H}^2_{\mathbb{C}}$, $\boldsymbol{p}$ will be a lift of p to $\mathbb{C}^3$.

2.1 The complex hyperbolic plane

The standard reference for complex hyperbolic space is Goldman's book [15]. A lot of information can also be found in Chen and Greenberg's paper [3]; see also the survey articles [24; 38].

Let H be the matrix

$$H = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

The Hermitian product on $\mathbb{C}^3$ associated to H is given by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^* H \boldsymbol{x}$. The corresponding Hermitian form has signature $(2,1)$, and we denote by V_- (resp. V_0 and V_+) the associated negative (resp. null and positive) cones in $\mathbb{C}^3$.

Definition 2.1 The *complex hyperbolic plane* $\boldsymbol{H}^2_{\mathbb{C}}$ is the image of V_- in $\mathbb{C}P^2$ by projectivisation and its boundary $\partial \boldsymbol{H}^2_{\mathbb{C}}$ is the image of V_0 in $\mathbb{C}P^2$. The complex hyperbolic plane is endowed with the *Bergman metric*

$$ds^2 = \frac{-4}{\langle \boldsymbol{z}, \boldsymbol{z} \rangle^2} \det \begin{pmatrix} \langle \boldsymbol{z}, \boldsymbol{z} \rangle & \langle d\boldsymbol{z}, \boldsymbol{z} \rangle \\ \langle \boldsymbol{z}, d\boldsymbol{z} \rangle & \langle d\boldsymbol{z}, d\boldsymbol{z} \rangle \end{pmatrix}.$$

The Bergman metric is equivalent to the *Bergman distance function* ρ defined by

$$\cosh^2\left(\frac{\rho(m,n)}{2}\right) = \frac{\langle \boldsymbol{m}, \boldsymbol{n} \rangle \langle \boldsymbol{n}, \boldsymbol{m} \rangle}{\langle \boldsymbol{m}, \boldsymbol{m} \rangle \langle \boldsymbol{n}, \boldsymbol{n} \rangle},$$

where $\boldsymbol{m}$ and $\boldsymbol{n}$ are lifts of m and n to $\mathbb{C}^3$.

Let $\boldsymbol{z} = [z_1, z_2, z_3]^T$ be a (column) vector in $\mathbb{C}^3 - \{\boldsymbol{0}\}$. Then $\boldsymbol{z} \in V_-$ (resp. V_0) if and only if $2\,\mathrm{Re}(z_1 \overline{z}_3) + |z_2|^2 < 0$ (resp. $= 0$). Vectors in V_0 with $z_3 = 0$ must have $z_2 = 0$ as well. Such a vector is unique up to scalar multiplication. We call its projectivisation the *point at infinity* $q_\infty \in \partial \boldsymbol{H}^2_{\mathbb{C}}$. If $z_3 \neq 0$ then we can use inhomogeneous coordinates with $z_3 = 1$. Writing $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = -2u$, we give $\boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$ *horospherical coordinates* $(z, t, u) \in \mathbb{C} \times \mathbb{R} \times \mathbb{R}_{\geq 0}$, defined as follows: a point $q \in \boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}}$ with horospherical coordinates (z,t,u) is represented by the following vector, which we call its *standard lift*:

$$\text{(3)} \qquad \boldsymbol{q}=\begin{bmatrix}-|z|^2-u+it\\ z\sqrt{2}\\ 1\end{bmatrix} \quad \text{if } q\neq q_\infty, \qquad \boldsymbol{q}_\infty=\begin{bmatrix}1\\0\\0\end{bmatrix} \quad \text{if } q=q_\infty.$$

Points of $\partial\boldsymbol{H}^2_{\mathbb{C}}-\{q_\infty\}$ have $u=0$ and we will abbreviate $(z,t,0)$ to $[z,t]$.

Horospherical coordinates give a model of complex hyperbolic space analogous to the upper half-plane model of the hyperbolic plane. The *Cygan metric* d_{Cyg} on $\partial\boldsymbol{H}^2_{\mathbb{C}}-\{q_\infty\}$ plays the role of the Euclidean metric on the upper half-plane. It is defined by the distance function

$$\text{(4)} \qquad d_{\mathrm{Cyg}}(p,q)=|\langle\boldsymbol{p},\boldsymbol{q}\rangle|^{1/2}=\big||z-w|^2+i(t-s+\mathrm{Im}(z\overline{w}))\big|^{1/2},$$

where p and q have horospherical coordinates $[z,t]$ and $[w,s]$. We may extend this metric to points p and q in $\boldsymbol{H}^2_{\mathbb{C}}$ with horospherical coordinates (z,t,u) and (w,s,v) by writing

$$d_{\mathrm{Cyg}}(p,q)=\big||z-w|^2+|u-v|+i(t-s+\mathrm{Im}(z\overline{w}))\big|^{1/2}.$$

If (at least) one of p and q lies in $\partial\boldsymbol{H}^2_{\mathbb{C}}$ then the formula $d_{\mathrm{Cyg}}(p,q)=|\langle\boldsymbol{p},\boldsymbol{q}\rangle|^{1/2}$ is still valid.

2.2 Isometries

Since the Bergman metric and distance function are both given solely in terms of the Hermitian form, any unitary matrix preserving this form is an isometry. Similarly, complex conjugation of points in $\mathbb{C}^3$ leaves both the metric and the distance function unchanged. Hence, complex conjugation is also an isometry.

Define $U(2,1)$ to be the group of unitary matrices preserving the Hermitian form and $\mathrm{PU}(2,1)$ to be the projective unitary group obtained by identifying nonzero scalar multiples of matrices in $U(2,1)$. We also consider the subgroup $\mathrm{SU}(2,1)$ of matrices in $U(2,1)$ with determinant 1.

Proposition 2.2 *Every Bergman isometry of $\boldsymbol{H}^2_{\mathbb{C}}$ is either holomorphic or antiholomorphic. The group of holomorphic isometries is $\mathrm{PU}(2,1)$, acting by projective transformations. Every antiholomorphic isometry is complex conjugation followed by an element of $\mathrm{PU}(2,1)$.*

Elements of $\mathrm{SU}(2,1)$ fall into three types, according to the number and type of the fixed points of the corresponding isometry. Namely, an isometry is *loxodromic* (resp. *parabolic*) if it has exactly two fixed points (resp. one fixed point) on $\partial\boldsymbol{H}^2_{\mathbb{C}}$. It is called *elliptic* when it has (at least) one fixed point inside $\boldsymbol{H}^2_{\mathbb{C}}$. An elliptic element $A \in \mathrm{SU}(2,1)$ is called *regular elliptic* whenever it has three distinct eigenvalues, and *special elliptic* if it has a repeated eigenvalue. The following criterion distinguishes the different isometry types:

Proposition 2.3 [15, Theorem 6.2.4] *Let $\mathcal{F}$ be the polynomial given by $\mathcal{F}(z) = |z|^4 - 8\,\mathrm{Re}(z^3) + 18|z|^2 - 27$, and A be a nonidentity matrix in $\mathrm{SU}(2,1)$. Then:*

1. *A is loxodromic if and only if $\mathcal{F}(\mathrm{tr}A) > 0$.*
2. *A is regular elliptic if and only if $\mathcal{F}(\mathrm{tr}A) < 0$.*
3. *If $\mathcal{F}(\mathrm{tr}A) = 0$, then A is either parabolic or special elliptic.*

We will be especially interested in elements of $\mathrm{SU}(2,1)$ with trace 0 or trace 3.

Lemma 2.4 [15, Section 7.1.3] (1) *A matrix A in $\mathrm{SU}(2,1)$ is regular elliptic of order three if and only if its trace is equal to zero.*

(2) *Let (p,q,r) be three pairwise distinct points in $\partial\boldsymbol{H}^2_{\mathbb{C}}$, not contained in a common complex line. Then there exists a unique order-three regular elliptic isometry E such that $E(p) = q$ and $E(q) = r$.*

Suppose that $T \in \mathrm{SU}(2,1)$ has trace equal to 3. Then all eigenvalues of T equal 1, that is, T is *unipotent*. If T is diagonalisable then it must be the identity; if it is nondiagonalisable then it must fix a point of $\partial\boldsymbol{H}^2_{\mathbb{C}}$. Conjugating within $\mathrm{SU}(2,1)$ if necessary, we may assume that T fixes q_∞. This implies that T is upper triangular with each diagonal element equal to 1.

Lemma 2.5 [15, Section 4.2] *Suppose that $[w,s] \in \partial\boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$. Then there is a unique $T_{[w,s]} \in \mathrm{SU}(2,1)$ taking the point $[0,0] \in \partial\boldsymbol{H}^2_{\mathbb{C}}$ to $[w,s]$. As a matrix this map is*

$$T_{[w,s]} = \begin{bmatrix} 1 & -\overline{w}\sqrt{2} & -|w|^2 + is \\ 0 & 1 & w\sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}. \tag{5}$$

Moreover, composition of such elements gives $\partial\boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$ the structure of the Heisenberg group

$$[w,s]\cdot[z,t] = [w+z, s+t-2\mathrm{Im}(z\overline{w})]$$

and $T_{[w,s]}$ acts as left Heisenberg translation on $\partial\boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$.

The action of $T_{[w,s]}$ on horospherical coordinates is

$$T_{[w,s]}\colon (z,t,u) \mapsto (w+z, s+t-2\mathrm{Im}(z\overline{w}), u).$$

An important observation is that this is an affine map, namely a translation and shear.

We can restate Lemma 2.5 in an invariant way. This result is actually true for any parabolic conjugacy class, as a special case of [26, Proposition 3.1].

Proposition 2.6 *Let (p_1, p_2, p_3) be a triple of pairwise distinct points in $\partial \boldsymbol{H}^2_{\mathbb{C}}$. Then there is a unique unipotent element of* $\mathrm{PU}(2,1)$ *fixing p_1 and taking p_2 to p_3.*

Proof We can choose $A \in \mathrm{SU}(2,1)$ taking p_1 to q_∞ and p_2 to $[0,0]$. The result then follows from Lemma 2.5. □

2.3 Totally geodesic subspaces

Maximal totally geodesic subspaces of $\boldsymbol{H}^2_{\mathbb{C}}$ have real dimension 2, and they fall into two types. Complex lines are intersections with $\boldsymbol{H}^2_{\mathbb{C}}$ of projective lines in $\mathbb{C}P^2$. By Hermitian duality, any complex line L is *polar* to a point in $\mathbb{C}P^2$ that is outside the closure of $\boldsymbol{H}^2_{\mathbb{C}}$. Any lift of this point is called a *polar vector* to L. Any two distinct points p and q in the closure of $\boldsymbol{H}^2_{\mathbb{C}}$ belong to a unique complex line, and a vector polar to this line is given by $\boldsymbol{p} \boxtimes \boldsymbol{q} = H\,\overline{\boldsymbol{p} \wedge \boldsymbol{q}}$. This can be verified directly using $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{y}^* H \boldsymbol{x}$ and the fact that, here, $H^2 = 1$. A more general description of cross-products in Hermitian vector spaces can be found in [15, Section 2.2.7].

The other type of maximal totally geodesic subspace is a Lagrangian plane. Lagrangian planes are $\mathrm{PU}(2,1)$ images of the set of real points $\boldsymbol{H}^2_{\mathbb{R}} \subset \boldsymbol{H}^2_{\mathbb{C}}$. In particular, real planes are fixed points sets of antiholomorphic isometric involutions (sometimes called *real symmetries*). The symmetry fixing $\boldsymbol{H}^2_{\mathbb{R}}$ is complex conjugation. In turn, the symmetry about any other Lagrangian plane $M \cdot \boldsymbol{H}^2_{\mathbb{R}}$, where $M \in \mathrm{SU}(2,1)$, is given by $\boldsymbol{z} \mapsto M\,\overline{M^{-1}}\,\overline{\boldsymbol{z}} = M\big(\overline{M^{-1}\boldsymbol{z}}\big)$. Note that the matrix $N = M\,\overline{M^{-1}}$ satisfies $N\overline{N} = \mathrm{id}$: this reflects the fact that real symmetries are involutions. We refer the reader to [15, Chapters 3 and 4].

2.4 Isometric spheres

Definition 2.7 For any $B \in \mathrm{SU}(2,1)$ that does not fix q_∞, the *isometric sphere* of B (denoted by $\mathcal{I}(B)$) is defined to be

$$\mathcal{I}(B) = \big\{ p \in \boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}} : |\langle \boldsymbol{p}, \boldsymbol{q}_\infty \rangle| = |\langle \boldsymbol{p}, B^{-1}(\boldsymbol{q}_\infty) \rangle| = |\langle B(\boldsymbol{p}), \boldsymbol{q}_\infty \rangle| \big\}, \tag{6}$$

where $\boldsymbol{p}$ is the standard lift of $p \in \boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}}$ given in (3).

The *interior* of $\mathcal{I}(B)$ is the component of its complement in $\boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}}$ that does not contain q_∞, namely,

$$\big\{ p \in \boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}} : |\langle \boldsymbol{p}, \boldsymbol{q}_\infty \rangle| > |\langle \boldsymbol{p}, B^{-1}(\boldsymbol{q}_\infty) \rangle| \big\}.$$

The *exterior* of $\mathcal{I}(B)$ is the component that contains the point at infinity q_∞

Suppose B is written as a matrix as

$$B = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & j \end{bmatrix}. \tag{7}$$

Then $B^{-1}(\boldsymbol{q}_\infty) = [\bar{j}, \bar{h}, \bar{g}]^T$. Thus B fixes q_∞ if and only if $g = 0$. If B does not fix q_∞ (that is, $g \neq 0$) the horospherical coordinates of $B^{-1}(q_\infty)$ are

$$B^{-1}(q_\infty) = \left[\frac{\bar{h}}{\bar{g}\sqrt{2}}, \operatorname{Im}\left(\frac{\bar{j}}{\bar{g}}\right) \right].$$

Lemma 2.8 [15, Section 5.4.5] *Let $B \in \mathrm{PU}(2,1)$ be an isometry of $\boldsymbol{H}^2_{\mathbb{C}}$ not fixing q_∞.*

(1) *The transformation B maps $\mathcal{I}(B)$ to $\mathcal{I}(B^{-1})$, and the interior of $\mathcal{I}(B)$ to the exterior of $\mathcal{I}(B^{-1})$.*

(2) *For any $A \in \mathrm{PU}(2,1)$ fixing q_∞ and such that the corresponding eigenvalue has unit modulus, we have $\mathcal{I}(B) = \mathcal{I}(AB)$.*

Using the characterisation (4) of the Cygan metric in terms of the Hermitian form, the following lemma is obvious:

Lemma 2.9 *Suppose that $B \in \mathrm{SU}(2,1)$ written in the form (7) does not fix q_∞. Then the isometric sphere $\mathcal{I}(B)$ is the Cygan sphere in $\boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}}$ with centre $B^{-1}(q_\infty)$ and radius $r_A = 1/|g|^{1/2}$.*

The importance of isometric spheres is that they form the boundary of the *Ford polyhedron*. This is the limit of Dirichlet polyhedra as the centre point approaches $\partial \boldsymbol{H}^2_{\mathbb{C}}$; see [15, Section 9.3]. The Ford polyhedron D for a discrete group Γ is the intersection of the (closures of the) exteriors of all isometric spheres for elements of Γ not fixing q_∞. That is,

$$D_\Gamma = \left\{ p \in \boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}} : |\langle \boldsymbol{p}, \boldsymbol{q}_\infty \rangle| \geq |\langle \boldsymbol{p}, B^{-1}\boldsymbol{q}_\infty \rangle| \text{ for all } B \in \Gamma \text{ with } B(q_\infty) \neq q_\infty \right\}.$$

Of course, just as for Dirichlet polyhedra, to construct the Ford polyhedron one must check infinitely many equalities. Therefore our method will be to guess the Ford polyhedron and check this using the Poincaré polyhedron theorem. When q_∞ is either in the domain of discontinuity or is a parabolic fixed point, the Ford polyhedron is preserved by Γ_∞, the stabiliser of q_∞ in Γ. It is a fundamental polyhedron for the partition of Γ into Γ_∞–cosets. In order to obtain a fundamental domain for Γ, one must intersect the Ford domain with a fundamental domain for Γ_∞.

2.5 Cygan spheres and geographical coordinates

We now give some geometrical results about Cygan spheres. They are, in particular, applicable to isometric spheres. The Cygan sphere $\mathcal{S}_{[0,0]}(r)$ of radius $r>0$ with centre the origin $[0,0]$ is the (real) hypersurface of $\boldsymbol{H}^2_{\mathbb{C}} \cup \partial \boldsymbol{H}^2_{\mathbb{C}}$ described in horospherical coordinates by

$$\mathcal{S}_{[0,0]}(r) = \{(z,t,u) : (|z|^2+u)^2+t^2 = r^4\}. \tag{8}$$

From (8) we immediately see that when written in horospherical coordinates the interior of $\mathcal{S}_{[0,0]}(r)$ is convex. The Cygan sphere $\mathcal{S}_{[w,s]}(r)$ of radius r with centre $[w,s]$ is the image of $\mathcal{S}_{[0,0]}(r)$ under the Heisenberg translation $T_{[w,s]}$. Since Heisenberg translations are affine maps in horospherical coordinates, we see that the interior of any Cygan sphere is convex. This immediately gives:

Proposition 2.10 *The intersection of two Cygan spheres is connected.*

Cygan spheres are examples of bisectors (otherwise called spinal hypersurfaces) and their intersection is an example of what Goldman calls an intersection of covertical bisectors. Thus Proposition 2.10 is a restatement of [15, Theorem 9.2.6]. There is a natural system of coordinates on bisectors in terms of totally geodesic subspaces; see [15, Section 5.1]. In particular for Cygan spheres, these are defined as follows:

Definition 2.11 Let $\mathcal{S}_{[0,0]}(r)$ be the Cygan sphere with centre the origin $[0,0]$ and radius $r>0$. The point $g(\alpha,\beta,w)$ of $\mathcal{S}_{[0,0]}(r)$ with *geographical coordinates* (α,β,w) is the point whose lift to $\mathbb{C}^3$ is

$$\boldsymbol{g}(\alpha,\beta,w) = \begin{bmatrix} -r^2e^{-i\alpha} \\ rwe^{i(-\alpha/2+\beta)} \\ 1 \end{bmatrix}, \tag{9}$$

where $\beta \in [0,\pi)$, $\alpha \in \left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ and $w \in [-\sqrt{2\cos\alpha}, \sqrt{2\cos\alpha}]$,

Let $\mathcal{S}_{[z,t]}(r)$ be the Cygan sphere with centre $[z,t]$ and radius r. Then geographical coordinates on $\mathcal{S}_{[z,t]}(r)$ are obtained from the ones on $\mathcal{S}_{[0,0]}(r)$ by applying the Heisenberg translation $T_{[z,t]}$ to the vector (9).

We will only be interested in geographical coordinates on $\mathcal{S}_{[0,0]}(1)$, the unit Cygan sphere centred at the origin. Note that for the point $g(\alpha,\beta,w)$ of this sphere, $\langle g(\alpha,\beta,u), g(\alpha,\beta,u)\rangle = w^2-2\cos\alpha$. Therefore the horospherical coordinates of $g(\alpha,\beta,w)$ are

$$\left(\tfrac{1}{\sqrt{2}}we^{i(-\alpha/2+\beta)},\ \sin\alpha,\ \cos\alpha-\tfrac{1}{2}w^2\right)$$

In particular, the points of $\mathcal{S}_{[0,0]}(1)$ on $\partial \boldsymbol{H}^2_{\mathbb{C}}$ are those with $w = \pm\sqrt{2\cos\alpha}$.

The level sets of α and β are totally geodesic subspaces of $\boldsymbol{H}^2_{\mathbb{C}}$; see [15, Example 5.1.8].

Proposition 2.12 *Let $\mathcal{S}_{[w,s]}(r)$ be a Cygan sphere with geographical coordinates (α,β,w).*

(1) *For each $\alpha_0 \in \left(-\frac{\pi}{2},\frac{\pi}{2}\right)$, the set of points $L_{\alpha_0} = \{g(\alpha,\beta,w) \in \mathcal{S}_{[w,s]}(r) : \alpha = \alpha_0\}$ is a complex line, called a* slice *of $\mathcal{S}_{[w,s]}(r)$.*

(2) *For each $\beta_0 \in [0,\pi)$, the set of points $R_{\beta_0} = \{g(\alpha,\beta,w) \in \mathcal{S}_{[w,s]}(r) : \beta = \beta_0\}$ is a Lagrangian plane, called a* meridian *of $\mathcal{S}_{[w,s]}(r)$.*

(3) *The set of points with $w = 0$ is the* spine *of $\mathcal{S}_{[w,s]}(r)$. It is a geodesic contained in every meridian.*

Remark 2.13 From (8), it is easy to see that projections of boundaries of Cygan spheres onto the z–factor are closed Euclidean discs in $\mathbb{C}$. This corresponds to the vertical projection onto $\mathbb{C}$ in the Heisenberg group. This fact is often useful to prove that two Cygan spheres are disjoint.

2.6 Cartan's angular invariant

Élie Cartan defined an invariant of triples of pairwise distinct points p_1, p_2, p_3 in $\partial \boldsymbol{H}^2_{\mathbb{C}}$; see [15, Section 7.1]. For any lifts $\boldsymbol{p}_j$ of p_j to $\mathbb{C}^3$, this invariant is defined by $\arg(-\langle \boldsymbol{p}_1, \boldsymbol{p}_2\rangle\langle \boldsymbol{p}_2 \boldsymbol{p}_3\rangle\langle \boldsymbol{p}_3, \boldsymbol{p}_1\rangle)$, where the argument is chosen to lie in $(-\pi,\pi]$. We state here some important properties of $\mathbb{A}$.

Proposition 2.14 [15, Sections 7.1.1 and 7.1.2] (1) $-\frac{\pi}{2} \le \mathbb{A}(p_1,p_2,p_3) \le \frac{\pi}{2}$ *for any triple of pairwise distinct points p_1, p_2, p_3.*

(2) $\mathbb{A}(p_1,p_2,p_3) = \pm\frac{\pi}{2}$ *if and only if p_1, p_2, p_3 lie on the same complex line.*

(3) $\mathbb{A}(p_1,p_2,p_3) = 0$ *if and only if p_1, p_2, p_3 lie on the same Lagrangian plane.*

(4) *Two triples p_1, p_2, p_3 and q_1, q_2, q_3 have $\mathbb{A}(p_1,p_2,p_3) = \mathbb{A}(q_1,q_2,q_3)$ if and only if there exists $A \in \mathrm{SU}(2,1)$ such that $A(p_j) = q_j$ for $j = 1, 2, 3$.*

(5) *Two triples p_1, p_2, p_3 and q_1, q_2, q_3 have $\mathbb{A}(p_1,p_2,p_3) = -\mathbb{A}(q_1,q_2,q_3)$ if and only if there exists an antiholomorphic isometry A such that $A(p_j) = q_j$ for $j = 1, 2, 3$.*

The following proposition will be useful to us when we parametrise the family of classes of groups Γ.

Proposition 2.15 *Let* $(\alpha_1,\alpha_2)\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2$. *Then there exists a unique* $\mathrm{PU}(2,1)$*–class of quadruples* (p_1,p_2,p_3,p_4) *of pairwise distinct boundary points of* $\boldsymbol{H}^2_{\mathbb{C}}$ *such that:*

(1) *The complex lines* L_{12} *and* L_{34} *respectively spanned by* (p_1,p_2) *and* (p_3,p_4) *are orthogonal.*

(2) $\mathbb{A}(p_1,p_3,p_2)=\alpha_1$ *and* $\mathbb{A}(p_1,p_3,p_4)=\alpha_2$.

Proof Since $\mathrm{PU}(2,1)$ acts transitively on pairs of distinct points of $\partial\boldsymbol{H}^2_{\mathbb{C}}$, we may assume using the Siegel model, that the points p_i are given in Heisenberg coordinates by

$$p_1=q_\infty,\qquad p_2=[0,0],\qquad p_3=[1,t],\qquad p_4=[z,s]. \tag{10}$$

Using the standard lifts given in Section 2.1 (denoted by $\boldsymbol{p}_i$), we see by a direct computation using the Hermitian cross-product that

$$\langle \boldsymbol{p}_1\boxtimes\boldsymbol{p}_2,\boldsymbol{p}_3\boxtimes\boldsymbol{p}_4\rangle=|z|^2-1+i(t-s).$$

Thus the condition $L_{12}\perp L_{34}$ gives $|z|=1$ and $t=s$. We thus write $z=e^{i\theta}$ with $\theta\in[0,2\pi)$. Now computing the triple products we see that

$$\mathbb{A}(p_1,p_3,p_2)=\arg(1-it),$$
$$\mathbb{A}(p_1,p_3,p_4)=\arg(1-\overline{z})=\arg\left(2ie^{i\theta/2}\sin\left(\tfrac{1}{2}\theta\right)\right).$$

In particular, α_1 and α_2 determine the values of t and θ. □

3 The parameter space

3.1 Coordinates

Our space of interest is the following:

Definition 3.1 Let $\mathcal{U}$ be the set of $\mathrm{PU}(2,1)$–conjugacy classes of nonelementary pairs (A,B) such that A, B and AB are unipotent.

Here, by nonelementary, we mean that the two isometries A and B have no common fixed point in $\partial\boldsymbol{H}^2_{\mathbb{C}}$. In fact, a slightly stronger statement will follow from Theorem 3.2 below. Namely A and B do not preserve a common complex line and so they have no common fixed point in $\mathbb{C}P^2$ (see Section 2.3). Another way to see this is that if A in $\mathrm{PU}(2,1)$ is unipotent and preserves a complex line, then its action on that complex line is via a unipotent element of $\mathrm{SL}(2,\mathbb{R})$ (that is, parabolic with trace $+2$). It is well known that if A and B are unipotent elements of $\mathrm{SL}(2,\mathbb{R})$ whose product is also unipotent then A and B must share a fixed point (if A, B and AB are all parabolic with distinct fixed points, at least one of them should have trace -2).

Note that $BA = A^{-1}(AB)A = B(AB)B^{-1}$ and so if AB is unipotent then so is BA. If p_{AB} and p_{BA} in $\partial \boldsymbol{H}^2_{\mathbb{C}}$ are the fixed points of AB and BA then we have $A(p_{BA}) = p_{AB}$ and $B(p_{AB}) = p_{BA}$. From Proposition 2.6 this means that A and B are uniquely determined by the fixed points of A, B, AB and BA. We describe a set of coordinates on $\mathcal{U}$ expressed in terms of the Cartan invariants of triples of these fixed points.

Theorem 3.2 *There is a bijection between $\mathcal{U}$ and the open square $(\alpha_1,\alpha_2)\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2$, which is given by the map*

$$\Lambda\colon (A,B)\mapsto (\mathbb{A}(p_A,p_{AB},p_B),\mathbb{A}(p_A,p_{AB},p_{BA})),$$

where p_A, p_B, p_{AB} and p_{BA} are the parabolic fixed points of the corresponding isometries.

This result can be see as a special case of the main result of [26]. For completeness, we include here a direct proof.

Proof First, the two quantities $\alpha_1 = \mathbb{A}(p_A,p_{AB},p_B)$ and $\alpha_2 = \mathbb{A}(p_A,p_{AB},p_{BA})$ are invariant under $\mathrm{PU}(2,1)$–conjugation and thus the map Λ is well-defined. Let us first prove that the image of Λ is contained in $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2$. In other words, we must show $\alpha_1\neq\pm\frac{\pi}{2}$ and $\alpha_2\neq\pm\frac{\pi}{2}$.

Fix a choice of lifts $\boldsymbol{p}_A$, $\boldsymbol{p}_B$, $\boldsymbol{p}_{AB}$ and $\boldsymbol{p}_{BA}$ for the fixed points of A, B, AB and BA. Since the fixed points are assumed to be distinct, we see that the Hermitian product of each pair of these vectors does not vanish. The conditions $A(p_{BA}) = p_{AB}$ and $B(p_{AB}) = p_{BA}$ imply that there exist two nonzero complex numbers λ and μ satisfying

$$A\boldsymbol{p}_{BA} = \lambda\boldsymbol{p}_{AB} \quad\text{and}\quad B\boldsymbol{p}_{AB} = \mu\boldsymbol{p}_{BA}.$$

As AB is unipotent, its eigenvalue associated to $\boldsymbol{p}_{AB}$ is 1, and therefore $\lambda\mu=1$. Moreover, using the fact that $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ are eigenvectors of A and B with eigenvalue 1, we have

$$\begin{aligned}\langle\boldsymbol{p}_{BA},\boldsymbol{p}_A\rangle &= \langle A\boldsymbol{p}_{BA},A\boldsymbol{p}_A\rangle = \lambda\langle\boldsymbol{p}_{AB},\boldsymbol{p}_A\rangle,\\ \langle\boldsymbol{p}_{AB},\boldsymbol{p}_B\rangle &= \langle B\boldsymbol{p}_{AB},B\boldsymbol{p}_B\rangle = \mu\langle\boldsymbol{p}_{BA},\boldsymbol{p}_B\rangle.\end{aligned}\tag{11}$$

Using $\lambda\mu = 1$ and (11), it is not hard to show that $\boldsymbol{n}_1 = \lambda\boldsymbol{p}_{AB} - \boldsymbol{p}_{BA}$ is a polar vector for the complex line L_1 spanned by $\boldsymbol{p}_A$ and $\boldsymbol{p}_B$ (see Section 2.3). Moreover, $\langle\boldsymbol{p}_{AB},\boldsymbol{n}_1\rangle = -\langle\boldsymbol{p}_{AB},\boldsymbol{p}_{BA}\rangle\neq 0$. Thus p_{AB} does not lie on L_1. That is, the three points p_A, p_B, p_{AB} do not lie on the same complex line and so $\alpha_1\neq\pm\frac{\pi}{2}$.

Likewise, again using $\lambda\mu = 1$ and (11) we find that $\boldsymbol{n}_2 = \langle \boldsymbol{p}_B, \boldsymbol{p}_{AB}\rangle \boldsymbol{p}_A - \langle \boldsymbol{p}_A, \boldsymbol{p}_{AB}\rangle \boldsymbol{p}_B$ is a polar vector for L_2 and $\langle \boldsymbol{p}_A, \boldsymbol{n}_2\rangle = -\langle \boldsymbol{p}_A, \boldsymbol{p}_{AB}\rangle\langle \boldsymbol{p}_A, \boldsymbol{p}_B\rangle \neq 0$. Hence p_A does not lie on L_2 and so $\alpha_2 \neq \pm\frac{\pi}{2}$. We remark that, by construction, we have $\langle \boldsymbol{n}_1, \boldsymbol{n}_2\rangle = 0$ and so in fact L_1 and L_2 are orthogonal.

To see that Λ is surjective, fix (α_1, α_2) in $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)^2$ and define

$$x_1 = \sqrt{2\cos\alpha_1} \quad \text{and} \quad x_2 = \sqrt{2\cos\alpha_2} \quad \text{for } \alpha_i \in \left(-\tfrac{\pi}{2}, \tfrac{\pi}{2}\right), \tag{12}$$

so $x_1, x_2 \in \mathbb{R}_+^*$. Now consider the elements of $\mathrm{SU}(2,1)$

$$A = \begin{bmatrix} 1 & -x_1x_2^2 & -x_1^2x_2^2e^{-i\alpha_2} \\ 0 & 1 & x_1x_2^2 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ x_1x_2^2e^{-i\alpha_1} & 1 & 0 \\ -x_1^2x_2^2e^{i\alpha_2} & -x_1x_2^2e^{i\alpha_1} & 1 \end{bmatrix}, \tag{13}$$

Clearly, A and B are unipotent, and AB is also unipotent since $\operatorname{tr}(AB) = 3$. The four fixed points can be lifted to the vectors

$$\boldsymbol{p}_A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{p}_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{p}_{AB} = \begin{bmatrix} -e^{i\alpha_1} \\ x_1e^{i\alpha_2} \\ 1 \end{bmatrix}, \quad \boldsymbol{p}_{BA} = \begin{bmatrix} -e^{i\alpha_1} \\ -x_1e^{-i\alpha_2} \\ 1 \end{bmatrix}. \tag{14}$$

They satisfy $\mathbb{A}(p_A, p_{AB}, p_B) = \alpha_1$ and $\mathbb{A}(p_A, p_{AB}, p_{BA}) = \alpha_2$. Note that when either α_1 or α_2 tends to $\pm\frac{\pi}{2}$ (that is, x_1 or x_2, respectively, tends to 0), A and B both tend to the identity matrix.

To see that Λ is injective, it suffices to prove that the quadruple $(p_A, p_B, p_{AB}, p_{BA})$ is uniquely determined by (α_1, α_2) up to isometry. Indeed, once this quadruple is fixed, A and B are uniquely determined by Proposition 2.6. The above discussion has proved that for any pair (A, B) in $\mathcal{U}$ the two complex lines spanned by (p_A, p_B) and (p_{AB}, p_{BA}), respectively, are orthogonal. The result then follows straightforwardly from Proposition 2.15. □

From now on, we will identify any conjugacy class of pair in $\mathcal{U}$ with its representative given by (13). We will repeatedly use the notation $x_i = \sqrt{2\cos\alpha_i}$ from (12) and, when necessary, we will freely combine x_i with trigonometric notation. It should be noted that the unipotent isometry A given by (13) is equal to the Heisenberg translation $T_{[\ell_A, t_A]}$ (see Lemma 2.5), where

$$\begin{aligned} \ell_A &= \tfrac{1}{\sqrt{2}}x_1x_2^2 = 2\cos\alpha_1\cos^2\alpha_2, \\ t_A &= x_1^2x_2^2\sin\alpha_2 = 4\cos\alpha_1\cos\alpha_2\sin\alpha_2. \end{aligned} \tag{15}$$

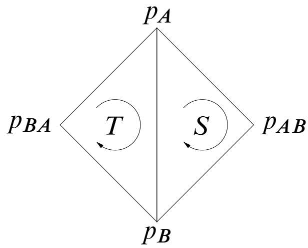


Figure 3: Action of S and T on the tetrahedron $(p_A, p_B, p_{AB}, p_{BA})$

3.2 Products of order-3 elliptics

The following proposition gives a decomposition of pairs in $\mathcal{U}$ that we will use in the rest of this work.

Proposition 3.3 *For any pair $(A, B) \in \mathcal{U}$, there exists a unique pair of isometries (S, T) such that:*

(1) *Both S and T have order three, and they cyclically permute (p_A, p_{AB}, p_B) and (p_A, p_B, p_{BA}), respectively.*

(2) *$A = ST$ and $B = TS$.*

Proof The first item is a direct consequence of Lemma 2.4 (note that neither of the triples (p_A, p_{AB}, p_B) or (p_A, p_B, p_{BA}) is contained in a complex line by Theorem 3.2). The action of S and T is summed up on Figure 3. From this, we see that ST (resp. TS) fixes p_A (resp. p_B) and maps p_{BA} to p_{AB} (resp. p_{AB} to p_{BA}). Provided that ST and TS are unipotent, this suffices to prove the second item by Proposition 2.6. To see that ST and TS are indeed unipotent, we can use the lifts of p_A, p_B, p_{AB} and p_{BA} given by (14). In this case we have

$$S = e^{-i\alpha_1/3} \begin{bmatrix} e^{i\alpha_1} & x_1 e^{i\alpha_1 - i\alpha_2} & -1 \\ -x_1 e^{i\alpha_2} & -e^{i\alpha_1} & 0 \\ -1 & 0 & 0 \end{bmatrix}, \qquad (16)$$
$$T = e^{i\alpha_1/3} \begin{bmatrix} 0 & 0 & -1 \\ 0 & -e^{-i\alpha_1} & -x_1 e^{-i\alpha_1 - i\alpha_2} \\ -1 & x_1 e^{i\alpha_2} & e^{-i\alpha_1} \end{bmatrix},$$

where, as usual, $x_i = \sqrt{2\cos\alpha_i}$; see (12). Computing the products ST and TS gives the result. □

We will use the notation S and T for these order-three symmetries throughout the paper.

A more geometric proof of the existence of order-three elliptic isometries decomposing pairs of parabolics as above can be found in a slightly more general context in [26].

One consequence of the existence of this decomposition as a product of order-three elliptics is that any group generated by a pair (A, B) in $\mathcal{U}$ is the image of the fundamental group of the Whitehead link complement by a morphism to $\mathrm{PU}(2,1)$. This follows directly from the following:

Proposition 3.4 *The free product $\mathbb{Z}_3 * \mathbb{Z}_3$ is a quotient of the fundamental group of the Whitehead link complement.*

Proof The fundamental group of the Whitehead link complement is presented by $\pi = \langle u, v \mid \mathrm{rel}(u, v)\rangle$, where

$$\mathrm{rel}(u,v) = [u,v]\cdot[u,v^{-1}]\cdot[u^{-1},v^{-1}]\cdot[u^{-1},v].$$

Making the substitution $u = st$ and $v = tst$, we observe $\mathrm{rel}(st, tst) = [st, s^{-1}t^{-3}s^{-2}]$. This relation is trivial whenever $s^3 = t^3 = 1$. Therefore, one defines a morphism $\mu\colon \pi \to \mathbb{Z}_3 * \mathbb{Z}_3$ by setting $\mu(u) = st$ and $\mu(v) = tst$. The morphism μ is surjective: t is the image of vu^{-1} and s the image of u^2v^{-1}. □

3.3 Symmetries of the moduli space

The parameters (α_1, α_2) determine Γ up to $\mathrm{PU}(2,1)$ conjugation. We now show that there is an antiholomorphic conjugation that changes the sign of both α_1 and α_2.

Proposition 3.5 *There is an antiholomorphic involution ι with the properties:*

(1) *ι interchanges p_A and p_B and interchanges p_{AB} and p_{BA}.*

(2) *ι conjugates S to T and A to B (and vice versa).*

(3) *ι conjugates the group Γ with parameters (α_1, α_2) to the group with parameters $(-\alpha_1, -\alpha_2)$.*

Proof The action on $\mathbb{C}^3$ of ι is

$$\iota : \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \mapsto \begin{bmatrix} \overline{z}_3 \\ e^{-i\alpha_1}\overline{z}_2 \\ \overline{z}_1 \end{bmatrix}.$$

It is easy to see that ι^2 is the identity and that ι sends $\boldsymbol{p}_A$ to $\boldsymbol{p}_B$ and sends $\boldsymbol{p}_{AB}$ to $(-e^{-i\alpha_1})\boldsymbol{p}_{BA}$. Projectivising gives the first part.

Since A is the unique unipotent map fixing p_A and sending p_{BA} to p_{AB}, we see $\iota A \iota$ is the unique unipotent map fixing $\iota(p_A) = p_B$ and sending $\iota(p_{BA}) = p_{AB}$ to $\iota(p_{AB}) = p_{BA}$. Thus $\iota A \iota = B$ and so $\iota B \iota = A$. Applying Proposition 3.3 we see that $\iota S \iota = T$ and $\iota T \iota = S$, proving the second part.

The parameters for the group $\iota\Gamma\iota$ are $\mathbb{A}(\iota p_A, \iota p_{AB}, \iota p_B) = \mathbb{A}(p_B, p_{BA}, p_A) = -\alpha_1$ and $\mathbb{A}(\iota p_A, \iota p_{AB}, \iota p_{BA}) = \mathbb{A}(p_B, p_{BA}, p_{AB}) = -\alpha_2$. This completes the proof. □

There are other symmetries of the parameter space $\mathcal{U}$ that, in general, do not arise from conjugation by isometries.

Proposition 3.6 *Let $\phi_h\colon (\alpha_1,\alpha_2) \mapsto (\alpha_1, -\alpha_2)$ and $\phi_v\colon (\alpha_1,\alpha_2) \mapsto (-\alpha_1, \alpha_2)$ denote the symmetries about the horizontal and vertical axes of the (α_1,α_2)–square. Then $\phi_h \circ \phi_v$ induces the conjugation by ι given in Proposition 3.5. Moreover:*

(1) *The symmetry ϕ_h induces the changes of generators $(S,T) \mapsto (T^{-1}, S^{-1})$ and $(A,B) \mapsto (A^{-1}, B^{-1})$.*

(2) *The symmetry ϕ_v induces the changes of generators $(S,T) \mapsto (S^{-1}, T^{-1})$ and $(A,B) \mapsto (B^{-1}, A^{-1})$,*

Proof Applying the change ϕ_h to the points in (14) and multiplying by the diagonal element $\mathrm{diag}(1, -1, 1) \in \mathrm{PU}(2,1)$ fixes p_A and p_B and swaps p_{AB} and p_{BA}. Therefore it sends S to the map cyclically permuting (p_A, p_{BA}, p_B), which is T^{-1}. Similarly it sends T to S^{-1}.

It is clear that the change of generators $(S,T) \mapsto (T^{-1}, S^{-1})$ sends $A = ST$ to $T^{-1}S^{-1} = A^{-1}$ and $B = TS$ to $S^{-1}T^{-1} = B^{-1}$.

The change of generators $(A,B) \mapsto (A^{-1}, B^{-1})$ fixes p_A and p_B. Since it sends AB to $A^{-1}B^{-1} = (BA)^{-1}$, it sends p_{AB} to p_{BA}, and similarly sends p_{BA} to p_{AB}. From this we can calculate the new Cartan invariants and we obtain the symmetry ϕ_h.

Hence all three conditions in the first part are equivalent. The second part then follows the first part and Proposition 3.5 by first applying ϕ_h and then conjugating by ι. □

The fixed-point sets of these automorphisms are related to $\mathbb{R}$–decomposability and $\mathbb{C}$–decomposability of Γ.

Definition 3.7 (compare Will [36]) A pair (S,T) of elements in $\mathrm{PU}(2,1)$ is $\mathbb{R}$–decomposable if there exist three antiholomorphic involutions $(\iota_1, \iota_2, \iota_3)$ such that $S = \iota_2\iota_1$ and $T = \iota_1\iota_3$.

A pair (S,T) of elements in $\mathrm{PU}(2,1)$ is $\mathbb{C}$–decomposable if there exists three involutions (I_1, I_2, I_3) in $\mathrm{PU}(2,1)$ such that $S = I_2 I_1$ and $T = I_1 I_3$.

The properties of $\mathbb{R}$– and $\mathbb{C}$–decomposability have also been studied (in the special case of pairs of loxodromic isometries) from the point of view of traces in $\mathrm{SU}(2,1)$ in [36], and (in the general case) using cross-ratios in [27]. We could take either point of view here, but instead we choose to argue directly with fixed points.

Proposition 3.8 *Let (A,B) be in $\mathcal{U}$ and (S,T) be the corresponding elliptic isometries.*

(1) *If $\alpha_1=0$, then the pair (S,T) is $\mathbb{C}$–decomposable and the pair (A,B) is $\mathbb{R}$–decomposable. In particular, $\langle S,T\rangle$ has index 2 in a $(3,3,\infty)$–triangle group.*

(2) *If $\alpha_2=0$, then the pair (S,T) is $\mathbb{R}$–decomposable and the pair (A,B) is $\mathbb{C}$–decomposable. In particular, $\langle A,B\rangle$ has index two in a complex hyperbolic ideal triangle group.*

Proof Consider the antiholomorphic involution $\iota_1\colon [z_1,\,z_2,\,z_3]\mapsto[\overline{z}_1,\,-\overline{z}_2,\,\overline{z}_3]$. Applying ι_1 to the points in (14) with $\alpha_1=0$, we see that ι_1 fixes p_A and p_B and interchanges p_{AB} and p_{BA}. Therefore ι_1 conjugates A to A^{-1} and B to B^{-1}. Hence $A\iota_1A\iota_1$ and $\iota_1B\iota_1B$ are the identity. That is, $\iota_2=A\iota_1$ and $\iota_3=\iota_1B$ are involutions. Hence (A,B) is $\mathbb{R}$–decomposable.

Again assuming $\alpha_1=0$, consider the holomorphic involution defined by $I_1=\iota_1\iota$ (where ι is the involution defined in Proposition 3.5). Then I_1 fixes p_{AB} and p_{BA} and interchanges p_A and p_B. Therefore, it conjugates S to S^{-1} and T to T^{-1}. This means $I_2=SI_1$ and $I_3=I_1T$ are involutions. Hence (S,T) is $\mathbb{C}$–decomposable.

Now consider the holomorphic involution $I_1'\colon [z_1,\,z_2,\,z_3]\mapsto[z_1,\,-z_2,\,z_3]$. This fixes p_A and p_B and when $\alpha_2=0$ it interchanges p_{AB} and p_{BA}. As above this means $I_2'=AI_1'$ and $I_3'=I_1'B$ are involutions and (A,B) is $\mathbb{C}$–decomposable. Finally, define $\iota_1'=I_1'\iota$. Arguing as above, again with $\alpha_2=0$, we see that $\iota_2'=S\iota_1'$ and $\iota_3'=\iota_1'T$ are involutions. Hence (S,T) is $\mathbb{R}$–decomposable. □

As indicated above, when $\alpha_1=0$ the group generated by (I_1,I_2,I_3) is a $(3,3,\infty)$ reflection triangle group. This group can be thought of as a limit as n tends to infinity of the $(3,3,n)$–triangle groups which have been studied by Parker, Wang and Xie [25]. The special case $(3,3,4)$ has been studied by Falbel and Deraux [8]. Both [8] and [25] constructed Dirichlet domains, and the Ford domain we construct can be seen as a limit of these. Moreover, $\mathbb{R}$–decomposability of the pair (A,B) when $\alpha_1=0$ can be used to show that these groups correspond to the bending representations of the fundamental group of a 3–punctured sphere that have been studied in [37]. Ideal triangle groups have been studied in great detail in [16; 31; 30; 33; 34].

3.4 Isometry type of the commutator

The isometry type of the commutator will play an important role in the rest of this paper. It is easily described using the order-three elliptic maps given by Proposition 3.3.

Proposition 3.9 *The commutator $[A,B]$ has the same isometry type as ST^{-1}. More precisely, consider $\mathcal{G}(x_1^4,x_2^4)=\mathcal{G}(4\cos^2\alpha_1,4\cos^2\alpha_2)$, where*

$$\mathcal{G}(x,y)=x^2y^4-4x^2y^3+18xy^2-27.$$

Then $[A,B]$ is loxodromic (resp. parabolic, elliptic) if and only if $\mathcal{G}(x_1^4,x_2^4)$ is positive (resp. zero, negative).

Proof First, from $A=ST$, $B=TS$ and the fact that S and T have order 3, we see that

$$[A,B]=ABA^{-1}B^{-1}=STTST^{-1}S^{-1}S^{-1}T^{-1}=(ST^{-1})^3.$$

This implies that $[A,B]$ has the same isometry type as ST^{-1} unless ST^{-1} is elliptic of order three, in which case $[A,B]$ is the identity. This would mean that A and B commute, which cannot be because their fixed point sets are disjoint.

Representatives of S and T in $\mathrm{SU}(2,1)$ are given in (16). A direct calculation using these matrices shows that $\mathrm{tr}(ST^{-1})=x_1^2x_2^4e^{i\alpha_1/3}$. The function $\mathcal{G}(x_1^4,x_2^4)$ above is obtained by plugging this value in the function $\mathcal{F}$ given in Proposition 2.3. □

The null locus of $\mathcal{G}(4\cos^2\alpha_1,4\cos^2\alpha_2)$ in the square $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2$ is a curve, which we will refer to as the *parabolicity curve* and denote by $\mathcal{P}$. It is depicted on Figure 4. Similarly, the region where $\mathcal{G}$ is positive (thus $[A,B]$ loxodromic) will be denoted by $\mathcal{L}$. It is a topological disc, which is the connected component of the complement of the curve $\mathcal{P}$ that contains the origin. The region where $[A,B]$ is elliptic will be denoted by $\mathcal{E}$.

4 Isometric spheres and their intersections

4.1 Isometric spheres for S, S^{-1} and their A–translates

In this section we give details of the isometric spheres that will contain the sides of our polyhedron D. The polyhedron D is our guess for the Ford polyhedron of Γ, subject to the combinatorial restriction discussed in Section 4.2.

We start with the isometric spheres $\mathcal{I}(S)$ and $\mathcal{I}(S^{-1})$ for S and its inverse. From the matrix for S given in (16), using Lemma 2.9 we see that $\mathcal{I}(S)$ and $\mathcal{I}(S^{-1})$ have radius $1/|-e^{-i\alpha_1/3}|^{1/2}=1$ and centres $S^{-1}(q_\infty)=p_B$ and $S(q_\infty)=p_{AB}$, respectively; see (14). In particular, $\mathcal{I}(S)$ is the Cygan sphere $\mathcal{S}_{[0,0]}(1)$ of radius 1 centred at the origin; see (8). In our computations we will use geographical coordinates in $\mathcal{I}(S)$ as in Definition 2.11. The polyhedron D will be the intersection of the exteriors of $\mathcal{I}(S^{\pm 1})$ and all their translates by powers of A. We now fix some notation:

Definition 4.1 For $k \in \mathbb{Z}$ let $\mathcal{I}_k^+$ be the isometric sphere $\mathcal{I}(A^k S A^{-k}) = A^k \mathcal{I}(S)$ and let $\mathcal{I}_k^-$ be the isometric sphere $\mathcal{I}(A^k S^{-1} A^{-k}) = A^k \mathcal{I}(S^{-1})$.

With this notation, we have:

Proposition 4.2 *For any integer $k \in \mathbb{Z}$, the isometric sphere $\mathcal{I}_k^+$ has radius 1 and is centred at the point with Heisenberg coordinates $[k\ell_A, kt_A]$, where ℓ_A and t_A are as in (15). Similarly, the isometric sphere $\mathcal{I}_k^-$ has radius 1 and centre the point with Heisenberg coordinates $[k\ell_A + \sqrt{\cos\alpha_1}e^{i\alpha_2}, -\sin\alpha_1]$.*

Proof As A is unipotent and fixes q_∞, it is a Cygan isometry, and thus preserves the radius of isometric spheres. This gives the part about radius. Moreover, it follows directly from (13) that A^k acts on the boundary of $\boldsymbol{H}^2_{\mathbb{C}}$ by left Heisenberg multiplication by $[k\ell_A, kt_A]$. This gives the part about centres by a straightforward verification. □

The following proposition describes a symmetry of the family $\{\mathcal{I}_k^\pm : k \in \mathbb{Z}\}$ which will be useful in the study of intersections of the isometric spheres $\mathcal{I}_k^\pm$.

Proposition 4.3 *Let φ be the antiholomorphic isometry $S\iota = \iota T$, where ι is as in Proposition 3.5. Then $\varphi^2 = A$, and φ acts on the Heisenberg group as a screw motion preserving the affine line parametrised by*

$$\Delta_\varphi = \left\{\delta_\varphi(x) = \left[x + \tfrac{i}{2}\sqrt{\cos\alpha_1}\sin\alpha_2, x\sqrt{\cos\alpha_1}\sin\alpha_2 - \tfrac{1}{2}\sin\alpha_1\right] : x \in \mathbb{R}\right\}. \tag{17}$$

Moreover, φ acts on isometric spheres as $\varphi(\mathcal{I}_k^+) = \mathcal{I}_k^-$ and $\varphi(\mathcal{I}_k^-) = \mathcal{I}_{k+1}^+$ for all $k \in \mathbb{Z}$.

Proof Using the fact that $T = \iota S \iota$ we see that $A = ST = S\iota S\iota = \varphi^2$. Moreover, $\varphi(p_A) = S\iota(p_A) = S(p_B) = p_A$. Hence φ is a Cygan isometry. It follows by direct calculation that φ sends $\delta_\varphi(x)$ to $\delta_\varphi\left(x + \frac{1}{2}\ell_A\right)$, and so preserves Δ_φ. Moreover,

$$\varphi(p_{BA}) = S\iota(p_{BA}) = S(p_{AB}) = p_B, \quad \varphi(p_B) = S\iota(p_B) = S(p_A) = p_{AB}.$$

Hence, φ sends $\mathcal{I}_{-1}^-$ to $\mathcal{I}_0^+$ since it is a Cygan isometry mapping the centre of $\mathcal{I}_{-1}^-$ to the centre of $\mathcal{I}_0^+$. Similarly, φ sends $\mathcal{I}_0^+$ to $\mathcal{I}_0^-$. The action on other isometric spheres follows since $\varphi^2 = A$. □

4.2 A combinatorial restriction

The following section is the crucial technical part of our work. As most of the proofs are computational, we will omit many of them here; they will be provided in Section 7. We are now going to restrict our attention to those parameters in the region $\mathcal{L}$ such that the three isometric spheres $\mathcal{I}_0^+ = \mathcal{I}(S)$, $\mathcal{I}_0^- = \mathcal{I}(S^{-1})$ and $\mathcal{I}_{-1}^- = \mathcal{I}(T)$ have no

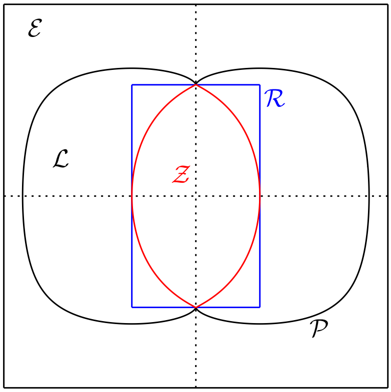


Figure 4: The parameter space, with the parabolicity curve $\mathcal{P}$ and the regions $\mathcal{E}$ and $\mathcal{L}$. The region $\mathcal{Z}$ is the central region, which is contained in the rectangle $\mathcal{R}$.

triple intersection. We will describe the region we are interested in by an inequality on α_1 and α_2. Prior to stating it, let us fix a little notation.

We let $\alpha_2^{\lim} = \arccos \sqrt{3/8}$. The two points $(0, \pm\alpha_2^{\lim})$ are the cusps of the curve $\mathcal{P}$; they satisfy $\mathcal{G}(4\cos^2 0, 4\cos^2 \alpha_2^{\lim}) = \mathcal{G}\big(4, \frac{3}{2}\big) = 0$ (see Figure 4). Now, let $\mathcal{R}$ be the rectangle (depicted in Figure 4) defined by

$$\mathcal{R} = \big\{(\alpha_1, \alpha_2) : |\alpha_1| \leqslant \tfrac{\pi}{6},\ |\alpha_2| \leqslant \alpha_2^{\lim}\big\}. \tag{18}$$

We remark that in Lemma 7.3 we will prove that when $(\alpha_1, \alpha_2) \in \mathcal{R}$, the commutator $[A, B]$ is nonelliptic. This means that $\mathcal{R}$ is contained in the closure of $\mathcal{L}$.

Definition 4.4 Let $\mathcal{Z}$ be the subset of $\mathcal{R}$ where the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^-$ is empty.

The following proposition characterises those points (α_1, α_2) that lie in $\mathcal{Z}$:

Proposition 4.5 *A parameter $(\alpha_1, \alpha_2) \in \mathcal{R}$ is in $\mathcal{Z}$ if and only if it satisfies*

$$\mathcal{D}(x_1^4, x_2^4) = \mathcal{D}(4\cos^2\alpha_1, 4\cos^2\alpha_2) > 0,$$

where $\mathcal{D}$ is the polynomial given by

$$\mathcal{D}(x, y) = x^3y^3 - 9x^2y^2 - 27xy^2 + 81xy - 27x - 27.$$

The region $\mathcal{Z}$ is depicted in Figure 4; it is the interior of the central region of the figure. In fact, $\mathcal{Z}$ is the region in all of $\mathcal{L}$ where $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^-$ is empty, but, as proving this is more involved, we restrict ourselves to the rectangle $\mathcal{R}$. This provides a priori bounds on the parameters α_1 and α_2 that will make our computations easier. We will prove Proposition 4.5 in Section 7.3. It relies on Proposition 4.6, describing the set of points where $\mathcal{D}(x_1^4, x_2^4) > 0$, and on Proposition 4.7, which gives geometric properties of the triple intersection. Proofs of Propositions 4.6 and 4.7 will be given in Sections 7.2 and 7.1, respectively.

Proposition 4.6 *The region $\mathcal{Z}$ is an open topological disc in $\mathcal{R}$, symmetric about the axes and intersecting them in the intervals*

$$\left\{\alpha_2 = 0,\ -\tfrac{\pi}{6} < \alpha_1 < \tfrac{\pi}{6}\right\} \quad \text{and} \quad \{\alpha_1 = 0,\ -\alpha_2^{\lim} < \alpha_2 < \alpha_2^{\lim}\}.$$

Moreover, the intersection of the closure of $\mathcal{Z}$ with the parabolicity curve $\mathcal{P}$ consists of the two points $(0, \pm\alpha_2^{\lim})$.

Proposition 4.7 (1) *The triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ is contained in the meridian $\mathfrak{m}$ of $\mathcal{I}_0^+$ defined in geographical coordinates by $\beta = \frac{1}{2}(\pi - \alpha_1)$.*

(2) *If the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ is nonempty, it contains a point in $\partial \boldsymbol{H}_{\mathbb{C}}^2$.*

The second part of Proposition 4.7 is not true for general triples of bisectors. It will allow us to restrict ourselves to the boundary of $\boldsymbol{H}_{\mathbb{C}}^2$ to prove Proposition 4.5. Restricting ourselves to the region $\mathcal{Z}$ will considerably simplify the combinatorics of the family of isometric spheres $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$. The following fact will be crucial in our study; compare Figure 5.

Proposition 4.8 *Fix a point (α_1, α_2) in $\mathcal{Z}$. Then the isometric sphere $\mathcal{I}_0^+$ is contained in the exterior of the isometric spheres $\mathcal{I}_k^{\pm}$ for all k, except for $\mathcal{I}_1^+$, $\mathcal{I}_{-1}^+$, $\mathcal{I}_0^-$ and $\mathcal{I}_{-1}^-$.*

The proof of Proposition 4.8 will be detailed in Section 7.4. We can give more information about the intersections $\mathcal{I}_0^{\pm}$ with these four other isometric spheres; compare Figure 5.

Proposition 4.9 *If $(\alpha_1, \alpha_2) \in \mathcal{Z}$, then the intersection $\mathcal{I}_{-1}^- \cap \mathcal{I}_0^-$ is contained in the interior of $\mathcal{I}_0^+$.*

Proof Since the point p_B is the centre of $\mathcal{I}_0^+$, it lies in its interior. Moreover, p_B lies on both $\mathcal{I}_{-1}^-$ and $\mathcal{I}_0^-$; indeed, $\langle \boldsymbol{p}_{AB}, \boldsymbol{p}_B \rangle = \langle \boldsymbol{p}_{BA}, \boldsymbol{p}_B \rangle = 1$. By convexity of Cygan spheres (see Proposition 2.10), the intersection of the latter two isometric spheres is connected. This implies that $\mathcal{I}_{-1}^- \cap \mathcal{I}_0^-$ is contained in the interior of $\mathcal{I}_0^+$, for otherwise $\mathcal{I}_0^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^-$ would not be empty. □

Using Proposition 4.3, applying powers of φ to Propositions 4.8 and 4.9 gives the following results describing all pairwise intersections:

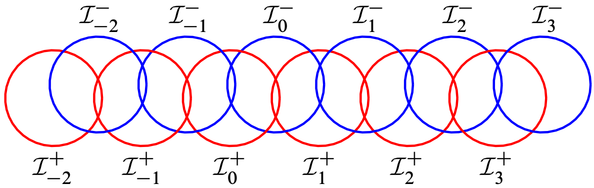


Figure 5: Vertical projections of the isometric spheres $\mathcal{I}_k^{\pm}$ for small values of k at the point $(\alpha_1, \alpha_2) = (0.4, 0.3)$

Corollary 4.10 *Fix $(\alpha_1, \alpha_2) \in \mathcal{Z}$. Then for all $k \in \mathbb{Z}$:*

(1) *$\mathcal{I}_k^+$ is contained in the exterior of all isometric spheres in $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$ except $\mathcal{I}_{k-1}^+$, $\mathcal{I}_{k-1}^-$, $\mathcal{I}_k^-$ and $\mathcal{I}_{k+1}^+$. Moreover, $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^- \cap \mathcal{I}_k^- = \varnothing$ and $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^+$ (resp. $\mathcal{I}_k^+ \cap \mathcal{I}_{k+1}^+$) is contained in the interior of $\mathcal{I}_{k-1}^-$ (resp. $\mathcal{I}_k^-$).*

(2) *$\mathcal{I}_k^-$ is contained in the exterior of all isometric spheres in $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$ except $\mathcal{I}_{k-1}^-$, $\mathcal{I}_k^+$, $\mathcal{I}_{k+1}^+$, and $\mathcal{I}_{k+1}^-$. Moreover, $\mathcal{I}_k^- \cap \mathcal{I}_k^- \cap \mathcal{I}_{k+1}^- = \varnothing$ and $\mathcal{I}_k^- \cap \mathcal{I}_{k-1}^-$ (resp. $\mathcal{I}_k^- \cap \mathcal{I}_{k+1}^-$) is contained in the interior of $\mathcal{I}_k^+$ (resp. $\mathcal{I}_{k+1}^+$).*

Proposition 4.8 and Corollary 4.10 are illustrated in Figure 5.

5 Applying the Poincaré polyhedron theorem inside $\mathcal{Z}$

5.1 The Poincaré polyhedron theorem

For the proof of our main result we need to use the Poincaré polyhedron theorem for coset decompositions. The general principle of this result is described in [2, Section 9.6] in the context of the Poincaré disc. A generalisation to the case of $\boldsymbol{H}_{\mathbb{C}}^2$ has already appeared in Mostow [22] and Deraux, Parker and Paupert [9]. In these cases the stabiliser of the polyhedron was assumed to be finite. In our case the stabiliser is the infinite cyclic group generated by the unipotent parabolic map A. There are two main differences from the version given in [9]. First, we allow the polyhedron D to have infinitely many facets; the stabiliser group Υ is also infinite, but we require that there are only finitely many Υ–orbits of facets. Secondly, we allow the boundary D to intersect $\partial \boldsymbol{H}_{\mathbb{C}}^2$ in an open set, which we refer to as the ideal boundary of D. In fact, the version we need has many things in common with the version given by Parker, Wang and Xie [25]. A more general statement will appear in Parker's book [23]. In what follows we will adapt our statement of the Poincaré theorem to the case we have in mind.

The polyhedron and its cell structure Let D be an open polyhedron in $\boldsymbol{H}_{\mathbb{C}}^2$ and let $\overline{D}$ denote its closure in $\overline{\boldsymbol{H}_{\mathbb{C}}^2} = \boldsymbol{H}_{\mathbb{C}}^2 \cup \partial \boldsymbol{H}_{\mathbb{C}}^2$. We define the ideal boundary $\partial_\infty D$ of D to be the intersection of $\overline{D}$ with $\partial \boldsymbol{H}_{\mathbb{C}}^2$. This polyhedron has a natural cell structure which we suppose is locally finite inside $\boldsymbol{H}^2_{\mathbb{C}}$. We suppose that the facets of D of all dimensions are piecewise smooth submanifolds of $\overline{\boldsymbol{H}^2_{\mathbb{C}}}$. Let $\mathcal{F}_k(D)$ be the collection of facets of codimension k having nontrivial intersection with $\boldsymbol{H}^2_{\mathbb{C}}$. We suppose that facets are closed subsets of $\overline{\boldsymbol{H}^2_{\mathbb{C}}}$. We write f° to denote the interior of a facet f, that is, the collection of points of f that are not contained in $\partial\boldsymbol{H}^2_{\mathbb{C}}$ or any facet of a lower dimension (higher codimension). Elements of $\mathcal{F}_1(D)$ and $\mathcal{F}_2(D)$ are called *sides* and *ridges* of D, respectively. Since D is a polyhedron, $\mathcal{F}_0(D)=\overline{D}$ and each ridge in $\mathcal{F}_2(D)$ lies in exactly two sides in $\mathcal{F}_1(D)$. Similarly, the intersection of facets of D with $\partial\boldsymbol{H}^2_{\mathbb{C}}$ gives rise to a polyhedral structure on a subset of $\partial_\infty D$. We let $\mathcal{IF}_k(D)$ denote the ideal facets of $\partial_\infty D$ of codimension k such that each facet in $\mathcal{IF}_k(D)$ is contained in some facet of $\mathcal{F}_\ell(D)$ with $\ell<k$. In particular, we will also need to consider *ideal vertices* in $\mathcal{IF}_4(D)$. These are either the endpoints of facets in $\mathcal{F}_3(D)$ or else they are points of $\partial\boldsymbol{H}^2_{\mathbb{C}}$ contained in (at least) two facets of D that do not intersect inside $\boldsymbol{H}^2_{\mathbb{C}}$. Note that, since we have defined ideal facets to be subsets of facets, it may be that $\partial\boldsymbol{H}^2_{\mathbb{C}}$ contains points of $\partial_\infty D$ not contained in any ideal facet. In the case we consider, there will be one such point, namely the point at ∞ fixed by A.

The side pairing We suppose that there is a *side pairing* $\sigma\colon \mathcal{F}_1(D)\to \mathrm{PU}(2,1)$ satisfying the following conditions:

(1) For each side $s\in\mathcal{F}_1(D)$ with $\sigma(s)=S$ there is another side $s^-\in\mathcal{F}_1(D)$ such that S maps s homeomorphically onto s^- preserving the cell structure. Moreover, $\sigma(s^-)=S^{-1}$. Furthermore, if $s=s^-$ then $S=S^{-1}$ and S is an involution. In this case, we call $S^2=\mathrm{id}$ a *reflection relation*.

(2) For each $s\in\mathcal{F}_1(D)$ with $\sigma(s)=S$ we have
$$\overline{D}\cap S^{-1}(\overline{D})=s \quad\text{and}\quad D\cap S^{-1}(D)=\varnothing.$$

(3) For each w in the interior s° of s there is an open neighbourhood $U(w)\subset\boldsymbol{H}^2_{\mathbb{C}}$ of w contained in $\overline{D}\cup S^{-1}(\overline{D})$.

In the example we consider, D will be the Ford domain of a group. In particular, each side s will be contained in the isometric sphere $\mathcal{I}(S)$ of $S=\sigma(s)$. Indeed, $s=\mathcal{I}(S)\cap\overline{D}$. By construction we have $S\colon\mathcal{I}(S)\mapsto\mathcal{I}(S^{-1})$ and in this case $s^-=\mathcal{I}(S^{-1})\cap\overline{D}$. The polyhedron D will be the (open) infinite-sided polyhedron formed by the intersection of the exteriors of all the $\mathcal{I}(S)$ where $S=\sigma(s)$ and s varies over $\mathcal{F}_1(D)$. By construction, the sides of D are smooth hypersurfaces (with boundary) in $\boldsymbol{H}^2_{\mathbb{C}}$.

Suppose that D is invariant under a group Υ that is *compatible* with the side pairing map in the sense that for all $P\in\Upsilon$ and $s\in\mathcal{F}_1(D)$ we have $P(s)\in\mathcal{F}_1(D)$ and $\sigma(Ps)=P\sigma(s)P^{-1}$. We call the latter a *compatibility relation*. We suppose that there are finitely many Υ–orbits of facets in each $\mathcal{F}_k(D)$. Since $P \in \Upsilon$ cannot fix a side $s \in \mathcal{F}_1(D)$ pointwise, subdividing sides if necessary, we suppose that if $P \in \Upsilon$ maps a side in $\mathcal{F}_1(D)$ to itself then P is the identity. In particular, given sides s_1 and s_2 in $\mathcal{F}_1(D)$, there is at most one $P \in \Upsilon$ sending s_1 to s_2. In the example of a Ford domain, Υ will be Γ_∞, the stabiliser of the point ∞ in the group Γ.

Ridges and cycle relations Consider a ridge $r_1 \in \mathcal{F}_2(D)$. Then r_1 is contained in precisely two sides of D, say s_0^- and s_1. Consider the ordered triple (r_1, s_0^-, s_1). The side pairing map $\sigma(s_1) = S_1$ sends s_1 to the side s_1^- preserving its cell structure. In particular, $S_1(r_1)$ is a ridge of s_1^-, say r_2. Let s_2 be the other side, containing r_2. Then we obtain a new ordered triple (r_2, s_1^-, s_2). Now apply $\sigma(s_2) = S_2$ to r_2 and repeat. Because there are only finitely many Υ–orbits of ridges, we eventually find an m such that the ordered triple $(r_{m+1}, s_m^-, s_{m+1}) = (P^{-1}r_1, P^{-1}s_0^-, P^{-1}s_1)$ for some $P \in \Upsilon$ (note that, by hypothesis, P is unique). We define a map $\rho\colon \mathcal{F}_2(D) \to \mathrm{PU}(2,1)$ called the *cycle transformation* by $\rho(r_1) = P \circ S_m \circ \cdots \circ S_1$. (Note that for any ridge $r_1 = s_0^- \cap s_1$, the cycle transformation map $\rho(r_1) = R$ depends on a choice of one of the sides s_0^- and s_1. If we choose the other one then the ridge cycle becomes R^{-1}. This follows from the fact that then $\sigma(s_j^-) = \sigma(s_j)^{-1}$ and from the compatibility relations.) By construction, the cycle transformation $R = \rho(r_1)$ maps the ridge r_1 to itself setwise. However, R may not be the identity on r_1, nor on $\boldsymbol{H}^2_{\mathbb{C}}$. Nevertheless, we suppose that R has order n. The relation $R^n = \mathrm{id}$ is called the *cycle relation* associated to r_1.

Writing the cycle transformation $\rho(r_1) = R$ in terms of P and the S_j, we let $\mathcal{C}(r_1)$ be the collection of suffix subwords of R^n. That is,

$$\mathcal{C}(r_1) = \{S_j \circ \cdots \circ S_1 \circ R^k : 0 \le j \le m-1,\ 0 \le k \le n-1\}.$$

We say that *the cycle condition* is satisfied at r_1 provided:

(1) $r_1 = \bigcap_{C \in \mathcal{C}(r_1)} C^{-1}(\bar{D})$.

(2) If $C_1, C_2 \in \mathcal{C}(r_1)$ with $C_1 \neq C_2$, then $C_1^{-1}(D) \cap C_2^{-1}(D) = \varnothing$.

(3) For each $w \in r_1^\circ$ there is an open neighbourhood $U(w)$ of w such that

$$U(w) \subset \bigcup_{C \in \mathcal{C}(r_1)} C^{-1}(\bar{D}).$$

Ideal vertices and consistent horoballs Suppose that the set $\mathcal{IF}_4(D)$ of ideal vertices of D is nonempty. In our applications, there are no edges (that is, $\mathcal{F}_3(D)$ is empty) and the only ideal vertices arise as points of tangency between the ideal boundaries of ridges in $\mathcal{F}_2(D)$. In order to simplify our discussion below, we will only treat this case. We require that there is a system of *consistent horoballs* based at the ideal vertices and their images under the side pairing maps (see [10, page 152] for the definition). For each ideal vertex $\xi \in \mathcal{IF}_4(D)$, the consistent horoball H_ξ is a horoball based at ξ with the following property: Let $\xi \in \mathcal{IF}_4(D)$ and let $s \in \mathcal{F}_1(D)$ be a side with $\xi \in s$. Then the side pairing $S = \sigma(s)$ maps ξ to a point ξ^- in s^-. Note that ξ^- is not necessarily an ideal vertex (since it could be that ξ is a point of tangency between two sides whose closures in $\overline{\boldsymbol{H}_{\mathbb{C}}^2}$ are otherwise disjoint and ξ^- may be a point of tangency between two nested bisectors only one of which contributes a side of D). In our case this does not happen and so we may assume ξ^- also lies in $\mathcal{IF}_4(D)$ and so has a consistent horoball H_{ξ^-}. In order for these horoballs to form a system of consistent horoballs we require that for each ideal vertex ξ and each side s with $\xi \in s$ the side pairing map $\sigma(s)$ should map the horoball H_ξ onto the horoball H_{ξ^-}. In particular, any cycle of side pairing maps sending ξ to itself must also send H_ξ to itself.

Statement of the Poincaré polyhedron theorem We can now state the version of the Poincaré polyhedron theorem that we need (compare [22] or [9]).

Theorem 5.1 *Let D be a smoothly embedded polyhedron D in $\boldsymbol{H}_{\mathbb{C}}^2$ together with a side pairing $\sigma\colon \mathcal{F}_1(D) \to \mathrm{PU}(2,1)$. Let $\Upsilon < \mathrm{PU}(2,1)$ be a group of automorphisms of D compatible with the side pairing and suppose that each $\mathcal{F}_k(D)$ contains finitely many Υ–orbits. Fix a presentation for Υ with generating set $\mathcal{P}_\Upsilon$ and relations $\mathcal{R}_\Upsilon$. Let Γ be the group generated by $\mathcal{P}_\Upsilon$ and the side pairing maps $\{\sigma(s)\}$. Suppose that the cycle condition is satisfied for each ridge in $\mathcal{F}_2(D)$ and that there is a system of consistent horoballs at all the ideal vertices of D (if any). Then:*

(1) *The images of D under the cosets of Υ in Γ tessellate $\boldsymbol{H}_{\mathbb{C}}^2$. That is, $\boldsymbol{H}_{\mathbb{C}}^2 \subset \bigcup_{A\in\Gamma} A(\overline{D})$ and $D \cap A(D) = \varnothing$ for all $A \in \Gamma - \Upsilon$.*

(2) *The group Γ is discrete and a fundamental domain for its action on $\boldsymbol{H}_{\mathbb{C}}^2$ is obtained from the intersection of D with a fundamental domain for Υ.*

(3) *A presentation for Γ (with respect to the generating set $\mathcal{P}_\Upsilon \cup \{\sigma(s)\}$) has the following set of relations: the relations $\mathcal{R}_\Upsilon$ in Υ, the compatibility relations between σ and Υ, the reflection relations and the cycle relations.*

5.2 Application to our examples

We are now going to apply Theorem 5.1 to the group generated by S and A. Explicit matrices for these transformations are provided in (13) and (16). Our aim is to prove:

Theorem 5.2 *Suppose that (α_1,α_2) is in $\mathcal{Z}$. That is, $\mathcal{D}(4\cos^2\alpha_1, 4\cos^2\alpha_1) > 0$, where $\mathcal{D}(x,y)$ is the polynomial defined in Proposition 4.5. Then the group $\Gamma = \langle S, A\rangle$ associated to the parameters (α_1,α_2) is discrete and has the presentation*

(19) $$\langle S, A : S^3 = (A^{-1}S)^3 = \mathrm{id}\rangle.$$

We obtain the presentation $\langle S, T : S^3 = T^3 = \mathrm{id}\rangle$ by changing generators to S and $T = A^{-1}S$.

Definition of the polyhedron and its cell structure The infinite polyhedron we consider is the intersection of the exteriors of all the isometric spheres in $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$.

Definition 5.3 We call D the intersection of the exteriors of all isometric spheres $\mathcal{I}_k^+$ and $\mathcal{I}_k^-$ with centres $A^k S^{-1}(q_\infty)$ and $A^k S(q_\infty)$, respectively:

$$D = \{q \in \boldsymbol{H}_{\mathbb{C}}^2 : d_{\mathrm{Cyg}}(q, A^k S^{\pm 1}(q_\infty)) > 1 \text{ for all } k \in \mathbb{Z}\}. \tag{20}$$

The set of sides of D is $\mathcal{F}_1(D) = \{s_k^+, s_k^- : k \in \mathbb{Z}\}$, where $s_k^+ = \mathcal{I}_k^+ \cap \overline{D}$ and $s_k^- = \mathcal{I}_k^- \cap \overline{D}$.

Using Corollary 4.10 we can completely describe s_k^+ and s_k^-.

Proposition 5.4 *The side $s_k^{\pm}$ is topologically a solid cylinder in $\boldsymbol{H}_{\mathbb{C}}^2 \cup \partial \boldsymbol{H}_{\mathbb{C}}^2$. More precisely, $s_k^{\pm}$ is a product $D \times [0,1]$, where for each $t \in [0,1]$, the fibre $D \times \{t\}$ is homeomorphic to a closed disc in $\overline{\boldsymbol{H}_{\mathbb{C}}^2}$ whose boundary is contained in $\partial \boldsymbol{H}_{\mathbb{C}}^2$. The intersection of ∂s_k^+ (resp. ∂s_k^-) with $\boldsymbol{H}_{\mathbb{C}}^2$ is the disjoint union of the topological discs $s_k^+ \cap s_{k-1}^-$ and $s_k^+ \cap s_k^-$ (resp. $s_k^- \cap s_k^+$ and $s_k^- \cap s_{k+1}^-$).*

Proof Since s_k^+ is contained in $\mathcal{I}_k^+$, its only possible intersections with other sides are contained in $\mathcal{I}_{k-1}^+$, $\mathcal{I}_{k-1}^-$, $\mathcal{I}_{k+1}^+$ and $\mathcal{I}_{k+1}^-$ by Corollary 4.10. Since $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^+$ and $\mathcal{I}_k^+ \cap \mathcal{I}_{k+1}^+$ are contained in the interiors of other isometric spheres, the intersections $s_k^+ \cap s_{k-1}^+$ and $s_k^+ \cap s_{k+1}^+$ are empty. Also, $\mathcal{I}_k^+ \cap \mathcal{I}_{k-1}^- \cap \mathcal{I}_k^- = \varnothing$ and so $s_k^+ \cap s_{k-1}^-$ and $s_k^+ \cap s_k^-$ are disjoint. Since isometric spheres are topological balls and their pairwise intersections are connected, the description of s_k^+ follows. A similar argument describes s_k^-. □

The side pairing $\sigma\colon \mathcal{F}_1(D) \to \mathrm{PU}(2,1)$ is defined by

$$\sigma(s_k^+) = A^k S A^{-k}, \quad \sigma(s_k^-) = A^k S^{-1} A^{-k}. \tag{21}$$

Let $\Upsilon = \langle A \rangle$ be the infinite cyclic group generated by A. By construction the side pairing σ is compatible with Υ. Furthermore, using Proposition 5.4 the set of ridges is $\mathcal{F}_2(D) = \{r_k^+, r_k^- : k \in \mathbb{Z}\}$, where $r_k^+ = s_k^+ \cap s_k^-$ and $r_k^- = s_k^+ \cap s_{k-1}^-$. We can now verify that σ satisfies the first condition of being a side pairing.

Proposition 5.5 *The side pairing map $\sigma(s_k^+) = A^k S A^{-k}$ is a homeomorphism from s_k^+ to s_k^-. Moreover, $\sigma(s_k^+)$ sends $r_k^+ = s_k^+ \cap s_k^-$ to itself and sends $r_k^- = s_k^+ \cap s_{k-1}^-$ to $r_{k+1}^- = s_k^- \cap s_{k+1}^+$.*

Proof By applying powers of A we need only need to consider the case where $k=0$. First, the ridge $r_0^+ = s_0^+ \cap s_0^- = \mathcal{I}(S) \cap \mathcal{I}(S^{-1})$ is defined by the triple equality

$$|\langle \boldsymbol{z}, \boldsymbol{q}_\infty \rangle| = |\langle \boldsymbol{z}, S^{-1}\boldsymbol{q}_\infty \rangle| = |\langle \boldsymbol{z}, S\boldsymbol{q}_\infty \rangle|. \tag{22}$$

The map S cyclically permutes $p_B = S^{-1}(q_\infty)$, $p_A = q_\infty$ and $p_{AB} = S(q_\infty)$, and so maps r_0^+ to itself. Similarly, consider $r_0^- = s_0^+ \cap s_{-1}^-$. The side pairing map S sends $A^{-1}S(q_\infty)$, the centre of $\mathcal{I}_{-1}^-$, to

$$\begin{aligned} S(A^{-1}S)(q_\infty) = S(T^{-1}S^{-1})S(q_\infty) &= ST^2(q_\infty) \\ &= (ST)S^{-1}(ST)(q_\infty) = AS^{-1}(q_\infty), \end{aligned}$$

which is the centre of $\mathcal{I}_1^+$, where we have used $A^{-1} = T^{-1}S^{-1}$, $T^{-1} = T^2$ and $ST(q_\infty) = q_\infty$. Therefore, $r_0^- = s_0^+ \cap s_{-1}^-$ is sent to $r_1^- = s_0^- \cap s_1^+$, as claimed. The rest of the result follows from our description of $s_k^\pm$ in Proposition 5.4. □

Local tessellation We now prove local tessellation around the sides and ridges of D.

- $\boldsymbol{s_k^\pm}$ Since $\sigma(s_k^\pm) = A^k S^{\pm 1} A^{-1}$ sends the exterior of $\mathcal{I}_k^\pm$ to the interior of $\mathcal{I}_k^\mp$ we see that D and $A^k S^{\pm 1} A^{-k}(D)$ have disjoint interiors and cover a neighbourhood of each point in $s_k^\mp$. Together with Proposition 5.5 this means σ satisfies the three conditions of being a side pairing.

- $\boldsymbol{r_0^+}$ Consider the case of $r_0^+ = s_0^+ \cap s_0^- = \mathcal{I}(S) \cap \mathcal{I}(S^{-1})$, which is given by (22). Observe that r_0^+ is mapped to itself by S. Using Proposition 5.5, we see that when constructing the cycle transformation for r_0^+ we have one ordered triple (r_0^+, s_0^-, s_0^+) and the cycle transformation $\rho(r_0^+) = S$. The cycle relation is $S^3 = \mathrm{id}$ and $\mathcal{C}(r_0^+) = \{\mathrm{id},\ S,\ S^2\}$. Consider an open neighbourhood U_0^+ of r_0^+ that intersects no other ridge. The intersection of D with U_0^+ is the same as the intersection of U_0^+ with the Ford domain D_S for the order-three group $\langle S \rangle$. Since S has order three, this Ford domain is the intersection of the exteriors of $\mathcal{I}(S)$ and $\mathcal{I}(S^{-1})$. For z in D_S, $|\langle \boldsymbol{z}, \boldsymbol{q}_\infty \rangle|$ is the smallest of the three quantities in (22). Applying $S = \sigma(s_0^+)$ and $S^{-1} = \sigma(s_0^-)$ gives regions $S(D_S)$ and $S^{-1}(D_S)$ where one of the other two quantities is the smallest. Therefore $U_0^+ \cap S(U_0^+) \cap S(U_0^-)$ is an open neighbourhood of r_0^+ contained in $D \cup S(D) \cup S^{-1}(D)$. This proves the cycle condition at r_0^+.

- $\boldsymbol{r_0^-}$ Now consider $r_0^- = s_0^+ \cap s_{-1}^-$. When constructing the cycle transformation for r_0^- we start with the ordered triple (r_0^-, s_{-1}^-, s_0^+). Applying $S = \sigma(s_0^+)$ to r_0^- gives the ordered triple (r_1^-, s_0^-, s_1^+), which is simply $(Ar_0^-, As_{-1}^-, As_0^+)$. Thus the cycle transformation of r_0^- is $\rho(r_0^-) = A^{-1}S = T^{-1}$, which has order 3. Therefore the cycle relation is $(A^{-1}S)^3 = \mathrm{id}$, and $\mathcal{C}(r_0^-) = \{\mathrm{id},\ A^{-1}S,\ (A^{-1}S)^2\}$. Noting that $\mathcal{I}_0^+$ has centre $S^{-1}(q_\infty) = S^{-1}A(q_\infty) = T(q_\infty)$ and $\mathcal{I}_{-1}^-$ has centre $A^{-1}S(q_\infty) =$

$T^{-1}(q-\infty)$, we see $\mathcal{I}_0^+ = \mathcal{I}(T^{-1})$ and $\mathcal{I}_0^- = \mathcal{I}(T)$. Therefore a similar argument involving the Ford domain for $\langle T \rangle$ shows that the cycle condition is satisfied at r_0^-.

- $\boldsymbol{r_k^{\pm}}$ Using compatibility of the side pairings with the cyclic group $\Upsilon = \langle A \rangle$, we see that $\rho(r_k^+) = A^k S A^{-k}$ with cycle relation $(A^k S A^{-k})^3 = A^k S^3 A^{-k} = \mathrm{id}$ and that the cycle condition is satisfied at r_k^+. Likewise, r_k^- is mapped by ρ to $A^k(A^{-1}S)A^{-k} = A^{k-1}SA^{-k}$ and $(A^{k-1}SA^{-k})^3 = A^k(A^{-1}S)A^{-k} = \mathrm{id}$, so the cycle condition is satisfied at r_k^-.

This is sufficient to prove Theorem 5.2 by applying the Poincaré polyhedron theorem when D has no ideal vertices, that is, to all groups Γ in the interior of $\mathcal{Z}$. In particular, Γ is generated by the generator A of Υ and the side pairing maps. Using the compatibility relations, there is only one side pairing map up to the action of Υ, namely S. There are no reflection relations, and (again up to the action of Υ) the only cycle relations are $S^3 = \mathrm{id}$ and $(A^{-1}S)^3 = \mathrm{id}$. Thus the Poincaré polyhedron theorem gives the presentation (19). This completes the proof of Theorem 5.2.

For groups on the boundary of $\mathcal{Z}$ the same result is also true. This follows from the fact (Chuckrow's theorem) that the algebraic limit of a sequence of discrete and faithful representations of a nonvirtually nilpotent group in $\mathrm{Isom}(\boldsymbol{H}^n_{\mathbb{C}})$ is discrete and faithful (see for instance [4, Theorem 2.7] or [21] for a more general result in the framework of negatively curved groups).

We do not need to apply the Poincaré polyhedron theorem for these groups. However, to describe the manifold at infinity for the limit groups, we will need to know a fundamental domain, and we will have to go through a similar analysis in the next section.

6 The limit group

In this section, we consider the group Γ^{lim}, and unless otherwise stated, the parameters α_1 and α_2 will always be assumed to be equal to 0 and α_2^{lim}, respectively. We know already that Γ^{lim} is discrete and isomorphic to $\mathbb{Z}_3 * \mathbb{Z}_3$. Our goal is to prove that its manifold at infinity is homeomorphic to the complement of the Whitehead link. For these values of the parameters, the maps $S^{-1}T$ and ST^{-1} are unipotent parabolic (see the results of Section 3.4), and we denote by $V_{S^{-1}T}$ and $V_{ST^{-1}}$, respectively, the sets of (parabolic) fixed points of conjugates of $S^{-1}T$ and ST^{-1} by powers of A.

(1) As in the previous section, we apply the Poincaré polyhedron theorem, this time to the group Γ^{lim}. We obtain an infinite A–invariant polyhedron, still denoted by D, which is a fundamental domain for A–cosets. This polyhedron is slightly more complicated than the one in the previous section due to the appearance of ideal vertices that are the points in $V_{S^{-1}T}$ and $V_{ST^{-1}}$.

(2) We analyse the combinatorics of the ideal boundary $\partial_\infty D$ of this polyhedron. More precisely, we will see that the quotient of $\partial_\infty D \setminus (\{p_A\} \cup V_{S^{-1}T} \cup V_{ST^{-1}})$ by the action of the group $\langle S, T\rangle$ is homeomorphic the complement of the Whitehead link, as stated in Theorem 6.4.

6.1 Matrices and fixed points

Before going any further, we provide specific expressions for the various objects we consider at the limit point. When $\alpha_1 = 0$ and $\alpha_2 = \alpha_2^{\lim}$, the map φ described in Proposition 4.3 is given in Heisenberg coordinates by

$$\varphi\colon [z,t] \mapsto \left[\overline{z} + \sqrt{3/8} + i\sqrt{5/8}, -t + x\sqrt{5/2} + y\sqrt{3/2}\right]. \tag{23}$$

In particular its invariant line Δ_φ is parametrised by

$$\Delta_\varphi = \left\{\delta_\varphi(x) = \left[x + i\sqrt{5/32},\ x\sqrt{5/8}\right] : x \in \mathbb{R}\right\}. \tag{24}$$

The parabolic map $A = \varphi^2$ acts on Δ_φ as $A\colon \delta_\varphi(x) \mapsto \delta_\varphi\left(x + \sqrt{3/2}\right)$. As a matrix it is given by

$$A = \begin{bmatrix} 1 & -\sqrt{3} & -\frac{3}{2} + \frac{i\sqrt{15}}{2} \\ 0 & 1 & \sqrt{3} \\ 0 & 0 & 1 \end{bmatrix}. \tag{25}$$

We can decompose A into the product of regular elliptic maps S and T, where

$$S = \begin{bmatrix} 1 & \frac{\sqrt{3}}{2} - \frac{i\sqrt{5}}{2} & -1 \\ -\frac{\sqrt{3}}{2} - \frac{i\sqrt{5}}{2} & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad T = \begin{bmatrix} 0 & 0 & -1 \\ 0 & -1 & -\frac{\sqrt{3}}{2} + \frac{i\sqrt{5}}{2} \\ -1 & \frac{\sqrt{3}}{2} + \frac{i\sqrt{5}}{2} & 1 \end{bmatrix}.$$

These maps cyclically permute (p_A, p_{AB}, p_B) and (p_A, p_B, p_{BA}), where

$$\boldsymbol{p}_A = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \boldsymbol{p}_B = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \boldsymbol{p}_{AB} = \begin{bmatrix} -1 \\ \frac{\sqrt{3}}{2} + \frac{i\sqrt{5}}{2} \\ 1 \end{bmatrix}, \quad \boldsymbol{p}_{BA} = \begin{bmatrix} -1 \\ -\frac{\sqrt{3}}{2} + \frac{i\sqrt{5}}{2} \\ 1 \end{bmatrix}. \tag{26}$$

Using $\alpha_1 = 0$, we will occasionally use the facts from Proposition 3.8 that (S, T) is $\mathbb{C}$–decomposable and (A, B) is $\mathbb{R}$–decomposable.

As mentioned above, in the group $\Gamma^{\lim}$ the elements ST^{-1}, $S^{-1}T$, TST, STS and the commutator $[A, B] = (ST^{-1})^3$ are unipotent parabolic. For future reference, we provide here lifts of their fixed points, both as vectors in $\mathbb{C}^3$ and in terms of geographical coordinates $g(\alpha,\beta)$ (we omit the w coordinate — since we are on the boundary at infinity, it is equal to $\sqrt{2\cos\alpha}$):

$$
\begin{aligned}
\boldsymbol{p}_{ST^{-1}} &= \begin{bmatrix} -\frac{1}{4}+\frac{i\sqrt{15}}{4} \\ \frac{\sqrt{3}}{4}+\frac{i\sqrt{5}}{4} \\ 1 \end{bmatrix} = \boldsymbol{g}\left(\arccos\tfrac{1}{4}, \tfrac{\pi}{2}\right), \\
\boldsymbol{p}_{S^{-1}T} &= \begin{bmatrix} -\frac{1}{4}-\frac{i\sqrt{15}}{4} \\ -\frac{\sqrt{3}}{4}+\frac{i\sqrt{5}}{4} \\ 1 \end{bmatrix} = \boldsymbol{g}\left(-\arccos\tfrac{1}{4}, \tfrac{\pi}{2}\right), \\
\boldsymbol{p}_{TST} &= \begin{bmatrix} -1 \\ -\frac{3\sqrt{3}}{4}+\frac{i\sqrt{5}}{4} \\ 1 \end{bmatrix} = \boldsymbol{g}\left(0, -\arccos\sqrt{27/32}\right), \\
\boldsymbol{p}_{STS} &= \begin{bmatrix} -1 \\ \frac{3\sqrt{3}}{4}+\frac{i\sqrt{5}}{4} \\ 1 \end{bmatrix} = \boldsymbol{g}\left(0, \arccos\sqrt{27/32}\right).
\end{aligned} \tag{27}
$$

It follows from (23) that φ acts on these parabolic fixed points as follows:

$$\cdots \to p_{T^{-1}STST} \xrightarrow{\varphi} p_{TST} \xrightarrow{\varphi} p_{S^{-1}T} \xrightarrow{\varphi} p_{ST^{-1}} \xrightarrow{\varphi} p_{STS} \xrightarrow{\varphi} p_{STSTS^{-1}} \to \cdots. \tag{28}$$

6.2 The Poincaré theorem for the limit group

The limit group has extra parabolic elements. Therefore, in order to apply the Poincaré theorem, we must construct a system of consistent horoballs at these parabolic fixed points (see Section 5.1).

Lemma 6.1 *The isometric spheres $\mathcal{I}_1^+$ and $\mathcal{I}_{-1}^-$ are tangent at $p_{ST^{-1}}$. The isometric spheres $\mathcal{I}_{-1}^+$ and $\mathcal{I}_0^-$ are tangent at $p_{S^{-1}T}$.*

Proof It is straightforward to verify that $|\langle \boldsymbol{p}_{ST^{-1}}, \boldsymbol{p}_{BA}\rangle| = |\langle \boldsymbol{p}_{ST^{-1}}, A(\boldsymbol{p}_B)\rangle| = 1$, and therefore $p_{ST^{-1}}$ belongs to both $\mathcal{I}_{-1}^-$ and $\mathcal{I}_1^+$. Projecting vertically — see Remark 2.13 — we see that the projections of $\mathcal{I}_{-1}^-$ and $\mathcal{I}_1^+$ are tangent discs and, as they are strictly convex, their intersection contains at most one point. This gives the result. The other tangency is along the same lines. □

A consequence of Lemma 6.1 is that the parabolic fixed points are tangency points of isometric spheres. The following lemma is proved in Section 7.1.

Lemma 6.2 *For the group Γ^{lim} the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ contains exactly two points, namely the parabolic fixed points $p_{ST^{-1}}$ and $p_{S^{-1}T}$.*

Applying powers of φ, we see that these triple intersections are actually quadruple intersections of sides and triple intersections of ridges.

Corollary 6.3 *The parabolic fixed point $A^k(p_{ST^{-1}})$ lies on $\mathcal{I}_{k-1}^- \cap \mathcal{I}_k^+ \cap \mathcal{I}_k^- \cap \mathcal{I}_k^+$. In particular, it is the triple ridge intersection $r_k^- \cap r_k^+ \cap r_{k+1}^-$. Similarly, $A^k(p_{S^{-1}T})$ lies on $\mathcal{I}_{-1}^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^-$. In particular it is $r_{k-1}^+ \cap r_k^- \cap r_k^+$.*

To construct a system of consistent horoballs at the parabolic fixed points we must investigate the action of the side pairing maps on them. First, $p_{S^{-1}T} \in \mathcal{I}_{-1}^+ \cap \mathcal{I}_{-1}^- \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^-$: we have

$$\begin{aligned}
\sigma(s_{-1}^+) = A^{-1}SA&\colon p_{S^{-1}T} \mapsto p_{T^{-1}STST},\\
\sigma(s_{-1}^-) = A^{-1}S^{-1}A&\colon p_{S^{-1}T} \mapsto p_{TST},\\
\sigma(s_0^+) = S&\colon p_{S^{-1}T} \mapsto p_{ST^{-1}},\\
\sigma(s_0^-) = S^{-1}&\colon p_{S^{-1}T} \mapsto p_{STS}.
\end{aligned}$$

Likewise, $p_{ST^{-1}} \in \mathcal{I}_{-1}^- \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_1^+$: we have

$$\begin{aligned}
\sigma(s_{-1}^-) = A^{-1}S^{-1}A&\colon p_{ST^{-1}} \mapsto A^{-2}(p_{ST^{-1}}),\\
\sigma(s_0^+) = S&\colon p_{ST^{-1}} \mapsto p_{STS},\\
\sigma(s_0^-) = S^{-1}&\colon p_{ST^{-1}} \mapsto p_{S^{-1}T},\\
\sigma(s_1^+) = ASA^{-1}&\colon p_{ST^{-1}} \mapsto A^2(p_{ST^{-1}}).
\end{aligned}$$

We can combine these maps to show how the points $A^k(p_{ST^{-1}})$ and $A^k(p_{S^{-1}T})$ are related by the side pairing maps. This leads to an infinite graph, a section of which is:

$$\begin{tikzcd}
{} \arrow[rr, "A^{-1}SA"] & & p_{ST^{-1}} \arrow[rr, "ASA^{-1}"] \arrow[dr, "S"] & & A^2(p_{ST^{-1}}) \arrow[r] & {} \\
{} & p_{T^{-1}STST} \arrow[l] \arrow[d, "A^{-1}SA"'] & p_{S^{-1}T} \arrow[l, "A^{-1}SA"'] \arrow[u, "S"] & p_{STS} \arrow[l, "S"] \arrow[d, "ASA^{-1}"'] & A^2(p_{S^{-1}T}) \arrow[l, "ASA^{-1}"'] \arrow[u, "A^2SA^{-2}"] & {} \arrow[l] \\
{} \arrow[r] & p_{TST} \arrow[rr, "S"'] \arrow[ur, "A^{-1}SA"] & & p_{STSTS^{-1}} \arrow[rr, "A^2SA^{-2}"'] \arrow[ur, "ASA^{-1}"] & & {}
\end{tikzcd} \tag{29}$$

From this it is clear that all the cycles in the graph (29) are generated by triangles and quadrilaterals. Up to powers of A, the triangles lead to the word S^3, which is the identity. Up to powers of A the quadrilaterals lead to words cyclically equivalent to the one coming from

$$p_{S^{-1}T} \xrightarrow{S^{-1}} p_{STS} \xrightarrow{ASA^{-1}} p_{STSTS^{-1}} \xrightarrow{S^{-1}} p_{TST} \xrightarrow{A^{-1}SA} p_{S^{-1}T}.$$

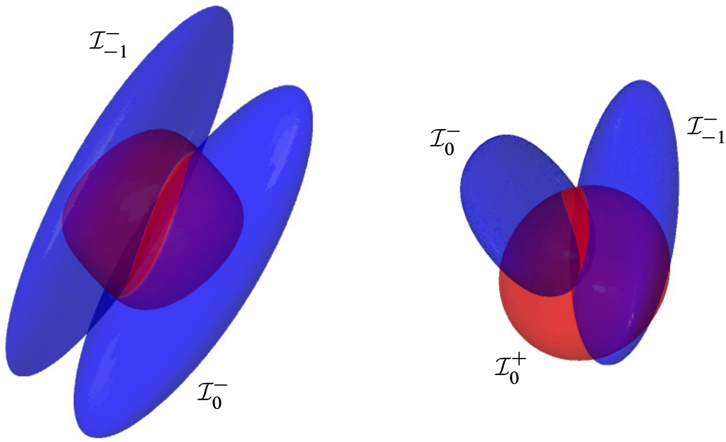


Figure 6: Two realistic views of the isometric spheres $\mathcal{I}_0^+$, $\mathcal{I}_1^+$ and $\mathcal{I}_0^-$ for the limit group Γ^{lim}. The thin bigon is $\mathcal{B}_0^+$ (defined in Proposition 6.5). Compare with Figures 7 and 12

In other words, $p_{S^{-1}T}$ is fixed by $(A^{-1}SA)(S^{-1})(ASA^{-1})(S^{-1}) = (T^{-1}S)^3$. This is parabolic and so preserves all horoballs based at $p_{S^{-1}T}$.

Therefore, we can define a system of horoballs as follows. Let U_0^+ be a horoball based at $p_{S^{-1}T}$, disjoint from the closure of any side not containing $p_{S^{-1}T}$ in its closure. Now define horoballs U_k^+ and U_k^- by applying the side pairing maps to U_0^+. Since every cycle in the graph (29) gives rise either to the identity map or to a parabolic map, this process is well-defined and gives rise to a consistent system of horoballs. Therefore we can apply the Poincaré polyhedron theorem for the two limit groups. Using the same arguments as we did for groups in the interior of $\mathcal{Z}$, we see that Γ has the presentation (19).

6.3 The boundary of the limit orbifold

Theorem 6.4 *The manifold at infinity of the group Γ^{lim} is homeomorphic to the Whitehead link complement.*

The ideal boundary of D is made up of those pieces of the isometric spheres $\mathcal{I}_k^{\pm}$ that are outside all other isometric spheres in $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$. Recall that the (ideal boundary of) the side $s_k^{\pm}$ is the part of $\partial\mathcal{I}_k^{\pm}$ which is outside (the ideal boundary of) all other isometric spheres. In this section, when we speak of sides and ridges we implicitly mean their intersection with $\partial \boldsymbol{H}_{\mathbb{C}}^2$.

We will see that each isometric sphere in $\{\mathcal{I}_k^{\pm} : k \in \mathbb{Z}\}$ contributes a side $s_k^{\pm}$ made up of one quadrilateral, denoted by $\mathcal{Q}_k^{\pm}$, and one bigon $\mathcal{B}_k^{\pm}$. A very similar configuration of isometric spheres has been observed by Deraux and Falbel [8]. We begin by analysing the contribution of $\mathcal{I}_0^+$.

Proposition 6.5 *The side $(s_0^+)^\circ$ of D has two connected components:*

(1) *One of them is a quadrilateral, denoted by $\mathcal{Q}_0^+$, whose vertices are points $p_{ST^{-1}}$, $p_{S^{-1}T}$, p_{STS} and p_{TST} (all of which are parabolic fixed points)*

(2) *The other is a bigon, denoted by $\mathcal{B}_0^+$, whose vertices are $p_{ST^{-1}}$ and $p_{S^{-1}T}$*

Proof Since isometric spheres are strictly convex, the ideal boundaries of the ridges $r_0^+ = \mathcal{I}_0^+ \cap \mathcal{I}_0^-$ and $r_0^- = \mathcal{I}_0^+ \cap \mathcal{I}_{-1}^-$ are Jordan curves on $\mathcal{I}_0^+$. We still denote them by $r_0^\pm$. The interiors of these curves are respectively the connected components containing p_{AB} and p_{BA}. By Lemma 6.2 in Section 7.1, r_0^+ and r_0^- have two intersection points, namely $p_{S^{-1}T}$ and $p_{ST^{-1}}$, and their interiors are disjoint. As a consequence the common exterior of the two curves has two connected components, and the points $p_{S^{-1}T}$ and $p_{ST^{-1}}$ lie on the boundary of both.

To finish the proof, consider the involution ι_1 defined in the proof of Proposition 3.8. (Note that since $\alpha_1 = 0$, this involution conjugates $\Gamma^{\lim}$ to itself.) In Heisenberg coordinates it is defined by $\iota_1\colon [z,t] \mapsto [-\overline{z}, -t]$ and is clearly a Cygan isometry. As in Proposition 3.8, ι_1 fixes p_A and p_B and it interchanges p_{AB} and p_{BA}. Thus it conjugates S to T^{-1}, and so it interchanges $p_{ST^{-1}}$ and $p_{S^{-1}T}$ and it interchanges p_{STS} and p_{TST}. Moreover, since it is a Cygan isometry, ι_1 preserves $\mathcal{I}_0^+$ and interchanges $\mathcal{I}_{-1}^-$ and $\mathcal{I}_0^-$ and thus it also exchanges the two curves r_0^+ and r_0^-. Again, since it is a Cygan isometry, it maps interior to interior and exterior to exterior for both curves. As a consequence, the two connected components of the common exterior are either exchanged or both preserved.

Now consider the point with Heisenberg coordinates $[i, 0]$. It is fixed by ι_1, and belongs to the common exterior of both r_0^+ and r_0^-. This implies that both connected components are preserved. Finally, since $p_{STS} \in \mathcal{I}_0^+ \cap \mathcal{I}_0^-$ and $p_{TST} \in \mathcal{I}_0^+ \cap \mathcal{I}_{-1}^-$ are exchanged by ι_1, these two points belong to the closure of the same connected component. As a consequence, one of the two connected components has $p_{ST^{-1}}$, $p_{S^{-1}T}$, p_{STS} and p_{TST} on its boundary. This is the quadrilateral. The other one has $p_{ST^{-1}}$ and $p_{S^{-1}T}$ on its boundary. This is the bigon. □

We now apply powers of A to get a result about all the isometric sphere intersections in the ideal boundary of D. Define $\mathcal{Q}_0^- = \varphi(\mathcal{Q}_0^+)$ and $\mathcal{B}_0^- = \varphi(\mathcal{B}_0^+)$. Then applying powers of A we define quadrilaterals $\mathcal{Q}_k^\pm = A^k(\mathcal{Q}_0^\pm)$ and bigons $\mathcal{B}_k^\pm = A^k(\mathcal{B}_0^\pm)$. The action of the Heisenberg translation A and the glide reflection φ are:

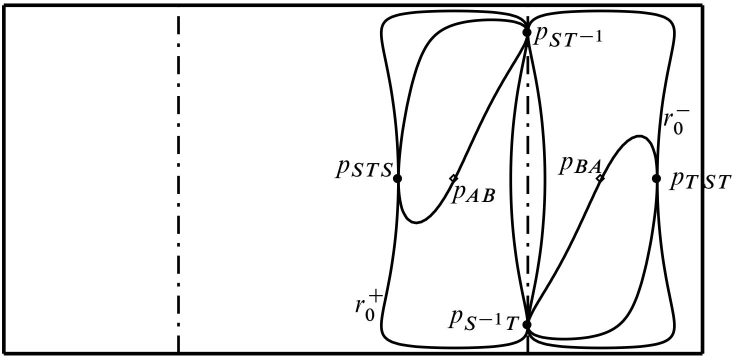


Figure 7: Intersections of the isometric spheres $\mathcal{I}_0^-$, $\mathcal{I}_{-1}^-$, $\mathcal{I}_1^+$ and $\mathcal{I}_{-1}^+$ with $\mathcal{I}_0^+$ in the boundary of $\boldsymbol{H}_{\mathbb{C}}^2$, viewed in geographical coordinates. Recall that $r_0^+ = \mathcal{I}_0^+ \cap \mathcal{I}_0^-$ and $r_0^- = \mathcal{I}_0^+ \cap \mathcal{I}_{-1}^-$. Here $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is the vertical coordinate, and $\beta \in [-\pi, \pi]$ the horizontal one. The vertical dash-dotted segments $\beta = \pm\frac{\pi}{2}$ are the two halves of the boundary of the meridian $\mathfrak{m}$. The bigon between the two curves r_0^+ and r_0^- is $\mathcal{B}_0^+$ (see Proposition 6.5). Compare to [8, Figure 2].

Corollary 6.6 *For the group Γ^{lim}, the (ideal boundary of) the side $s_k^{\pm}$ is the union of the quadrilateral $\mathcal{Q}_k^{\pm}$ and the bigon $\mathcal{B}_k^{\pm}$. The action of A and φ are as follows:*

(1) *A maps $\mathcal{Q}_k^{\pm}$ to $\mathcal{Q}_{k+1}^{\pm}$ and $\mathcal{B}_k^{\pm}$ to $\mathcal{B}_{k+1}^{\pm}$.*

(2) *φ maps $\mathcal{Q}_k^+$ to $\mathcal{Q}_k^-$, $\mathcal{Q}_k^-$ to $\mathcal{Q}_{k+1}^+$, $\mathcal{B}_k^+$ to $\mathcal{B}_k^-$ and $\mathcal{B}_k^-$ to $\mathcal{B}_{k+1}^+$.*

In order to understand the combinatorics of the sides of D, we describe the edges of the faces lying in $\mathcal{I}_0^+$. The three points $p_{S^{-1}T}$, $p_{ST^{-1}}$ and p_{STS} lie on the ridge $r_0^+ = \mathcal{I}_0^+ \cap \mathcal{I}_0^-$. Likewise, the points $p_{ST^{-1}}$, $p_{S^{-1}T}$ and p_{TST} lie on the ridge $r_0^- = \mathcal{I}_0^+ \cap \mathcal{I}_{-1}^-$. Indeed, these points divide (the ideal boundaries of) these ridges into three segments. We have listed the ideal vertices in positive cyclic order (see Figure 7). Using the graph (29), the action of the cycle transformations $\rho(s_0^+) = S$ and $\rho(r_0^-) = A^{-1}S = T^{-1}$ on these ideal vertices, and hence on the segments of the ridges, is

$$p_{S^{-1}T} \xrightarrow{S} p_{ST^{-1}} \xrightarrow{S} p_{STS} \xrightarrow{S} p_{S^{-1}T},$$
$$p_{ST^{-1}} \xrightarrow{A^{-1}S} p_{S^{-1}T} \xrightarrow{A^{-1}S} p_{TST} \xrightarrow{A^{-1}S} p_{ST^{-1}}.$$

Furthermore, S maps p_{TST} to $p_{STSTS^{-1}}$.

The quadrilateral $\mathcal{Q}_0^+$ has two edges $[p_{S^{-1}T}, p_{TST}] \cup [p_{TST}, p_{ST^{-1}}]$ in the ridge r_0^- and two edges $[p_{ST^{-1}}, p_{STS}] \cup [p_{STS}, p_{S^{-1}T}]$ in the ridge r_0^+. It is sent by S to the quadrilateral $\mathcal{Q}_0^-$ with two edges $[p_{ST^{-1}}, p_{STSTS^{-1}}] \cup [p_{STSTS^{-1}}, p_{STS}]$ in r_1^-

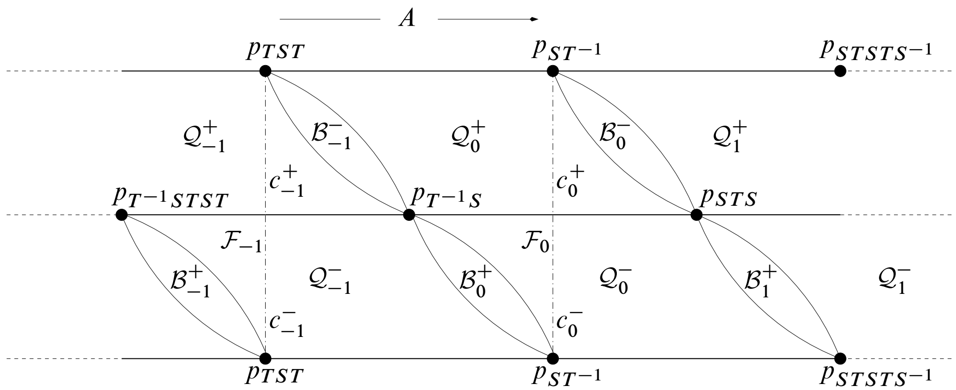


Figure 8: A combinatorial picture of ∂D. The top and bottom lines are identified.

and two edges $[p_{STS}, p_{S^{-1}T}] \cup [p_{S^{-1}T}, p_{ST^{-1}}]$ in r_0^+. Similarly, the edges of the bigon $\mathcal{B}_0^+$ are the remaining segments in r_0^- and r_0^+, both with endpoints $p_{S^{-1}T}$ and $p_{ST^{-1}}$. It is sent by S to the bigon $\mathcal{B}_0^-$ with vertices $p_{ST^{-1}}$ and p_{STS}.

Applying powers of A gives the other quadrilaterals and bigons. As usual, the image under A^k can be found by adding k to each subscript and conjugating each side pairing map and ridge cycle by A^k. The combinatorics of D is summarised on Figure 8.

Lemma 6.7 *The line Δ_φ given in (24) is contained in the complement of D.*

Proof As noted above, A acts on Δ_φ as a translation through $\sqrt{3/2}$. We claim that the segment of Δ_φ with parameter $x \in \left[-\sqrt{3/8}, \sqrt{3/8}\right]$ in contained in the interior of $\mathcal{I}_0^+$. Applying powers of A we see that each point of Δ_φ is contained in $\mathcal{I}_k^+$ for some k. Hence the line is in the complement of D.

Consider $\delta_\varphi(x) \in \Delta_\varphi$ with $x^2 \leq \frac{3}{8}$. The Cygan distance between p_B and $\delta_\varphi(x)$ satisfies

$$d_{\mathrm{Cyg}}(p_B, \delta_\varphi(x))^4 = \left|-x^2 - \tfrac{5}{32} + ix\sqrt{5/8}\right|^2 = x^4 + \tfrac{15}{16}x^2 + \tfrac{25}{1} - 24 \leq \tfrac{529}{1024}.$$

Since $d_{\mathrm{Cyg}}(p_B, \delta_\varphi(x)) < 1$ this means $\delta_\varphi(x)$ is in the interior of $\mathcal{I}_0^+$, as claimed. □

The following result, which will be proved in Section 7.5, is crucial for proving Theorem 6.4.

Proposition 6.8 *There exists a homeomorphism $\Psi\colon \mathbb{R}^3 \to \partial \boldsymbol{H}_{\mathbb{C}}^2 - \{q_\infty\}$ mapping the exterior of $S^1 \times \mathbb{R}$, that is, $\{(x,y,z) : x^2 + y^2 \geq 1\}$, homeomorphically onto D and such that $\Psi(x,y,z+1) = A\Psi(x,y,z)$, that is, Ψ is equivariant with respect to unit translation along the z axis and A.*

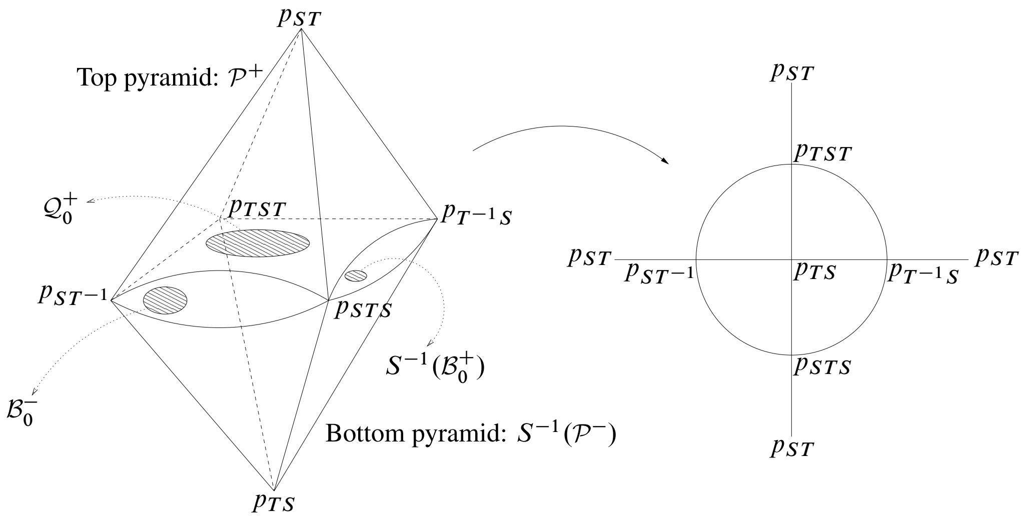


Figure 9: A combinatorial picture of the octahedron

As a consequence of Proposition 6.8, D admits an A–invariant 1–dimensional foliation, the leaves being the images of radial lines $\{(r\cos\theta_0, r\sin\theta_0, z_0) : r \geq 1\}$ that foliate the exterior of $S^1 \times \mathbb{R}$. Each of these leaves is a curve connecting a point of ∂D with q_∞. We can now prove Theorem 6.4.

Proof of Theorem 6.4. The union $\mathcal{Q}_0^+ \cup \mathcal{B}_0^+ \cup \mathcal{Q}_0^- \cup \mathcal{B}_0^-$ is a fundamental domain for the action of A on the boundary cylinder ∂D. As the foliation obtained above is A–invariant, the cone to the point q_∞ built over it via the foliation is a fundamental domain for the action of A over D, and thus it is a fundamental domain for the action of Γ^{lim} on the region of discontinuity $\Omega(\Gamma^{\mathrm{lim}})$.

This fundamental domain is the union of two pyramids $\mathcal{P}^+$ and $\mathcal{P}^-$, with respective bases $\mathcal{Q}_0^+ \cup \mathcal{B}_0^-$ and $\mathcal{Q}_0^- \cup \mathcal{B}_0^+$, and common vertex $q_\infty = p_{ST}$. The two pyramids share a common face, which is a triangle with vertices p_{STS}, $p_{T^{-1}S}$ and p_{ST}. Cutting and pasting, consider the union $\mathcal{P}^+ \cup S^{-1}(\mathcal{P}^-)$. It is again a fundamental domain for Γ^{lim}. The apex of $S^{-1}(\mathcal{P}^-)$ is $S^{-1}(q_\infty) = p_B = p_{TS}$. The image under S^{-1} of $\mathcal{Q}_0^-$ is $\mathcal{Q}_0^+$, and the bigon $\mathcal{B}_0^+$ is mapped by S^{-1} to another bigon connecting $p_{T^{-1}S}$ to p_{STS}. Since $\mathcal{B}_0^- = S(\mathcal{B}_0^+)$, this new bigon is the image of $\mathcal{B}_0^-$ under $S^{-2} = S$.

The resulting object is a is a polyhedron (a combinatorial picture is provided on Figure 9), whose faces are triangles and bigons. The faces of this octahedron are paired as follows:

$$\begin{aligned}
TS\colon (p_{TS}, p_{T^{-1}S}, p_{STS}) &\mapsto (p_{TS}, p_{TST}, p_{TS^{-1}}),\\
ST\colon (p_{ST}, p_{TST}, p_{T^{-1}S}) &\mapsto (p_{ST}, p_{ST^{-1}}, p_{STS}),\\
T\colon (p_{ST}, p_{TST}, p_{ST^{-1}}) &\mapsto (p_{TS}, p_{T^{-1}S}, p_{TST}),\\
S\colon (p_{TS}, p_{ST^{-1}}, p_{STS}) &\mapsto (p_{ST}, p_{STS}, p_{S^{-1}T}),\\
S\colon (p_{ST^{-1}}, p_{STS}) &\mapsto (p_{STS}, p_{S^{-1}T}).
\end{aligned}$$

The last line is the bigon identification between $\mathcal{B}_0^-$ and $S^{-1}(\mathcal{B}_0^+)$. As the triangle $(p_{TS}, p_{ST^{-1}}, p_{STS})$ and the bigon $\mathcal{B}_0^-$ share a common edge and have the same face pairing, they can be combined into a single triangle, as well as their images. Thus the last two lines may be combined into a single side with side pairing map S. We therefore obtain a true combinatorial octahedron. The face identifications given above make the quotient manifold homeomorphic to the complement of the Whitehead link (compare for instance [35, Section 3.3]). □

7 Technicalities

7.1 The triple intersections: proofs of Proposition 4.7 and Lemma 6.2

In this section we first prove Proposition 4.7, which states that the triple intersection must contain a point of $\partial \boldsymbol{H}^2_{\mathbb{C}}$, and then we analyse the case of the limit group $\Gamma^{\lim}$, giving a proof of Lemma 6.2. First recall that the isometric spheres $\mathcal{I}_0^-$ and $\mathcal{I}_{-1}^-$ are the unit Heisenberg spheres with centres given respectively in geographical coordinates by (see Section 2.5)

$$\begin{aligned}
\boldsymbol{p}_{AB} &= S(\infty) = g\big(-\alpha_1, -\tfrac{1}{2}\alpha_1 + \alpha_2, \sqrt{2\cos\alpha_1}\big),\\
\boldsymbol{p}_{BA} &= A^{-1}S(\infty) = g\big(-\alpha_1, -\tfrac{1}{2}\alpha_1 - \alpha_2 + \pi, \sqrt{2\cos\alpha_1}\big).
\end{aligned} \tag{30}$$

Consider the two functions of points $q = g(\alpha,\beta,w) \in \mathcal{I}_0^+$ defined by

$$\begin{aligned}
f^{[0]}_{\alpha_1,\alpha_2}(q) &= 2\cos^2\big(\tfrac12\alpha - \tfrac12\alpha_1\big) + \cos(\alpha-\alpha_1)\\
&\qquad - 4wx_1\cos\big(\tfrac12\alpha - \tfrac12\alpha_1\big)\cos\big(\beta + \tfrac12\alpha_1 - \alpha_2\big) + w^2x_1^2,\\
f^{[-1]}_{\alpha_1,\alpha_2}(q) &= 2\cos^2\big(\tfrac12\alpha - \tfrac12\alpha_1\big) + \cos(\alpha-\alpha_1)\\
&\qquad + 4wx_1\cos\big(\tfrac12\alpha - \tfrac12\alpha_1\big)\cos\big(\beta + \tfrac12\alpha_1 + \alpha_2\big) + w^2x_1^2.
\end{aligned}$$

These functions characterise those points on $\mathcal{I}_0^+$ that belong to $\mathcal{I}_0^-$ and $\mathcal{I}_{-1}^-$.

Lemma 7.1 *A point q on $\mathcal{I}_0^+$ lies on $\mathcal{I}_0^-$ (resp. in its interior or exterior) if and only if it satisfies $f^{[0]}_{\alpha_1,\alpha_2}(q) = 0$ (resp. is negative or is positive). Similarly, a point q on $\mathcal{I}_0^+$ lies on $\mathcal{I}_{-1}^-$ (resp. in its interior or exterior) if and only if it satisfies $f^{[-1]}_{\alpha_1,\alpha_2}(q) = 0$ (resp. is negative or is positive).*

Proof A point $q \in \mathcal{I}_0^+$ lies on $\mathcal{I}_0^-$ (resp. in its interior or exterior) if and only if its Cygan distance from the centre of $\mathcal{I}_0^-$, which is the point p_{AB}, equals 1 (resp. is less than 1 or greater than 1). Equivalently (see Section 2.4), the following quantity vanishes (resp. is positive or negative):

$$
\begin{aligned}
(31)\quad & |\langle \boldsymbol{q}, \boldsymbol{p}_{AB}\rangle|^2-1 \\
&= |-e^{-i\alpha}+wx_1e^{-i\alpha/2+i\beta-i\alpha_2}-e^{-i\alpha_1}|^2-1 \\
&= \left|-2\cos\left(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\right)+wx_1e^{i\beta+i\alpha_1/2-i\alpha_2}\right|^2-1 \\
&= 4\cos^2\left(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\right)-1+w^2x_1^2-4\cos\left(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\right)wx_1\cos\left(\beta+\tfrac{1}{2}\alpha_1-\tfrac{1}{2}\alpha\right) \\
&= f^{[0]}_{\alpha_1,\alpha_2}(q).
\end{aligned}
$$

On the last line we used $2\cos^2\left(\frac{1}{2}\alpha-\frac{1}{2}\alpha_1\right)=1+\cos(\alpha-\alpha_1)$. This proves the first part of the lemma and the second is obtained by a similar computation. □

Corollary 7.2 *For given (α_1,α_2), if the sum $f^{[0]}_{\alpha_1,\alpha_2}+f^{[-1]}_{\alpha_1,\alpha_2}$ is positive for all q, then the triple intersection $\mathcal{I}_0^+\cap\mathcal{I}_0^-\cap\mathcal{I}_{-1}^-$ is empty.*

See Figure 7. We can now prove Proposition 4.7.

Proof of Proposition 4.7 To prove the first part, note that a necessary condition for a point $q\in\mathcal{I}_0^+$ to be in the intersection $\mathcal{I}_0^-\cap\mathcal{I}_{-1}^-$ is that $f^{[0]}_{\alpha_1,\alpha_2}(q)-f^{[-1]}_{\alpha_1,\alpha_2}(q)=0$. By a simple computation, we see that this difference is

$$f^{[0]}_{\alpha_1,\alpha_2}(q)-f^{[-1]}_{\alpha_1,\alpha_2}(q)=-8wx_1\cos\left(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\right)\cos\left(\beta+\tfrac{1}{2}\alpha_1\right)\cos\alpha_2.$$

Since α_1 and α_2 lie in $\left(-\frac{\pi}{2},\frac{\pi}{2}\right)$ and $\alpha\in\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$, the only solutions are

$$\cos\left(\beta+\tfrac{1}{2}\alpha_1\right)=0 \quad\text{or}\quad w=0.$$

Thus either $p=g(\alpha,\beta,w)$ lies on the meridian $\mathfrak{m}$, or on the spine of $\mathcal{I}_0^+$, and hence on every meridian, in particular on $\mathfrak{m}$ (compare with Proposition 2.12).

To prove the second part of Proposition 4.7, assume that the triple intersection contains a point $q=g\left(\alpha,\frac{\pi}{2}-\frac{1}{2}\alpha_1,w\right)$ inside $\boldsymbol{H}^2_{\mathbb{C}}$, that is, such that $w^2<2\cos\alpha$, and

$$f^{[0]}_{\alpha_1,\alpha_2}(q)+f^{[-1]}_{\alpha_1,\alpha_2}(q)=0.$$

In view of Corollary 7.2, we only need to prove that there exists a point on $\partial\mathfrak{m}$ where the above sum is nonpositive, and use the intermediate value theorem. To do so, let $\widetilde{\alpha}$ be defined by the condition $2\cos\widetilde{\alpha}=w^2$ and such that $\widetilde{\alpha}$ and α_1 have opposite signs. Since $w^2<2\cos\alpha$, these conditions imply that $|\widetilde{\alpha}|>|\alpha|$. We claim that the point $\widetilde{q}=g\left(\widetilde{\alpha},\pi-\frac{1}{2}\alpha_1,w\right)$ is satisfactory. Indeed, the conditions on $\widetilde{\alpha}$ give

$$|\alpha-\alpha_1|\le|\alpha|+|\alpha_1|<|\widetilde{\alpha}|+|\alpha_1|=|\widetilde{\alpha}-\alpha_1|,$$

where the last inequality follows from the fact that $\widetilde{\alpha}$ and α_1 have opposite signs. Therefore,

$$\cos\big(\tfrac{1}{2}\widetilde{\alpha}-\tfrac{1}{2}\alpha_1\big)<\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big). \tag{32}$$

On the other hand, we have

$$\begin{aligned} &f^{[0]}_{\alpha_1,\alpha_2}(q)+f^{[-1]}_{\alpha_1,\alpha_2}(q) \\ &\quad=4\cos^2\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)+2\cos(\alpha-\alpha_1)-8wx_1\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)\sin\alpha_2+2w^2x_1^2 \\ &\quad=8\cos^2\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)-2-8wx_1\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)\sin\alpha_2+2w^2x_1^2. \end{aligned} \tag{33}$$

We claim this is an increasing function of $\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)$. In order to see this, observe that its derivative with respect to this variable is

$$16\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)-8wx_1\sin\alpha_2>16\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)-16\sqrt{\cos\alpha\cos\alpha_1}\geq 0,$$

where we used $x_1=\sqrt{2\cos\alpha_1}$, $w<\sqrt{2\cos\alpha}$ and $\sin\alpha_2\leq 1$. Therefore,

$$\begin{aligned} 0&=f^{[0]}_{\alpha_1,\alpha_2}(q)+f^{[-1]}_{\alpha_1,\alpha_2}(q) \\ &=8\cos^2\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)-2-8wx_1\cos\big(\tfrac{1}{2}\alpha-\tfrac{1}{2}\alpha_1\big)\sin\alpha_2+2w^2x_1^2 \\ &>8\cos^2\big(\tfrac{1}{2}\widetilde{\alpha}-\tfrac{1}{2}\alpha_1\big)-2-8wx_1\cos\big(\tfrac{1}{2}\widetilde{\alpha}-\tfrac{1}{2}\alpha_1\big)\sin\alpha_2+2w^2x_1^2 \\ &=f^{[0]}_{\alpha_1,\alpha_2}(\widetilde{q})+f^{[-1]}_{\alpha_1,\alpha_2}(\widetilde{q}). \end{aligned}$$

This proves our claim. □

We now prove Lemma 6.2, w,hich completely describes the triple intersection at the limit point.

Proof of Lemma 6.2 From the first part of Proposition 4.7 we see that any point $q=g(\alpha,\beta,w)$ in $\mathcal{I}_0^+\cap\mathcal{I}_0^-\cap\mathcal{I}_{-1}^-$ must lie on $\mathfrak{m}$, that is, $\beta=\frac{1}{2}(\pi-\alpha_1)$. For such points it is enough to show that $f^{[0]}_{0,\alpha_2^{\lim}}(q)+f^{[-1]}_{0,\alpha_2^{\lim}}(q)=0$. Substituting $\alpha_1=0$ and $\sin\alpha_2=\sqrt{5/8}$, this becomes

$$\begin{aligned} f^{[0]}_{0,\alpha_2^{\lim}}(q)+f^{[-1]}_{0,\alpha_2^{\lim}}(q)&=4\cos^2\big(\tfrac{1}{2}\alpha\big)+\cos\alpha-4\sqrt{5}w\cos\big(\tfrac{1}{2}\alpha\big)+4w^2 \\ &=\big(2\cos\big(\tfrac{1}{2}\alpha\big)-\sqrt{5}w\big)^2+(2\cos\alpha-w^2). \end{aligned}$$

In order to vanish, both terms must be zero. Hence $w^2=2\cos\alpha$ and $2\cos\big(\tfrac{1}{2}\alpha\big)=\sqrt{5}w=\sqrt{10\cos\alpha}$, noting w cannot be negative since $\alpha\in\big[-\frac{\pi}{2},\frac{\pi}{2}\big]$. This means $\alpha=\pm\arccos\frac{1}{4}$ and $w=\sqrt{2\cos\alpha}=\frac{1}{\sqrt{2}}$. Therefore, the only points in $\mathcal{I}_0^+\cap\mathcal{I}_0^-\cap\mathcal{I}_{-1}^-$ have geographical coordinates $g\big(\pm\arccos\frac{1}{4},\frac{\pi}{2},\frac{1}{\sqrt{2}}\big)$. Using (27), we see these points are $p_{ST^{-1}}$ and $p_{S^{-1}T}$. □

7.2 The region $\mathcal{Z}$ is an open disc in the region $\mathcal{L}$: proof of Proposition 4.6

Consider the group $\Gamma_{\alpha_1,\alpha_2}$ and, as before, write $x_1^4 = 4\cos^2\alpha_1$ and $x_2^4 = 4\cos^2\alpha_2$. Recall, from Proposition 3.9, that (α_1,α_2) is in $\mathcal{L}$ (resp. $\mathcal{P}$) if $\mathcal{G}(x_1^4,x_2^4)>0$ (resp. $=0$), where

$$\mathcal{G}(x,y) = x^2y^4 - 4x^2y^3 + 18xy^2 - 27. \tag{34}$$

Recall this means $[A,B]$ is loxodromic (resp. parabolic). Also (α_1,α_2) is in the rectangle $\mathcal{R}$ if and only if $(x_1^4,x_2^4)\in[3,4]\times\left[\frac{3}{2},4\right]$. From Proposition 4.5, the point $(\alpha_1,\alpha_2)\in\mathcal{R}$ is in $\mathcal{Z}$ (resp. $\partial\mathcal{Z}$) if $\mathcal{D}(x_1^4,x_2^4)>0$ (resp. $=0$), where

$$\mathcal{D}(x,y) = x^3y^3 - 9x^2y^2 - 27xy^2 + 81xy - 27x - 27. \tag{35}$$

Lemma 7.3 *Suppose $(\alpha_1,\alpha_2)\in\mathcal{R}$. Then $(\alpha_1,\alpha_2)\in\mathcal{L}\cup\mathcal{P}$, that is, the commutator $[A,B]$ is loxodromic or parabolic* (see Section 3.4). *Moreover, $(\alpha_1,\alpha_2)\in\mathcal{P}$ if and only if $(\alpha_1,\alpha_2)=(0,\pm\alpha_2^{\lim})$.*

Proof We first claim that the function $\mathcal{G}(x,y)$ has no critical points in $(0,\infty)\times(0,\infty)$. Indeed, the first partial derivatives of $\mathcal{G}(x,y)$ are

$$\mathcal{G}_x(x,y) = 2y^2(xy^2-4xy+9), \quad \mathcal{G}_y(x,y) = 4xy(xy^2-3xy+9).$$

These are not simultaneously zero for any positive values of x and y. As a consequence, the minimum of $\mathcal{G}$ on $[3,4]\times\left[\frac{3}{2},4\right]$ is attained on the boundary of this rectangle. We then have

$$\begin{aligned} \mathcal{G}\left(x,\tfrac{3}{2}\right) &= \tfrac{27}{16}(4-x)(5x-4), & \mathcal{G}(x,4) &= 9(32x-3), \\ \mathcal{G}(3,y) &= 9(y-1)(y^3-3y^2+3y+3), & \mathcal{G}(4,y) &= (2y+1)(2y-3)^3. \end{aligned}$$

It is a simple exercise to check that under the assumptions that $(x,y)\in[3,4]\times\left[\frac{3}{2},4\right]$, all four of these terms are positive, except for when $(x,y)=\left(4,\frac{3}{2}\right)$, in which case $\mathcal{G}\left(4,\frac{3}{2}\right)=0$. Then $(x_1^4,x_2^4)=\left(4,\frac{3}{2}\right)$ if and only if $(\alpha_1,\alpha_2)=(0,\pm\alpha_2^{\lim})$; compare to Figure 4. □

Lemma 7.4 *The region $\mathcal{Z}$ is an open topological disc in $\mathcal{R}$ symmetric about the axes and intersecting them in the intervals*

$$\left\{\alpha_2=0,\ \tfrac{\pi}{6}<\alpha_1<\tfrac{\pi}{6}\right\} \quad \textit{and} \quad \{\alpha_1=0,\ -\alpha_2^{\lim}<\alpha_2<\alpha_2^{\lim}\}.$$

Moreover, the only points of $\partial\mathcal{Z}$ that lie in the boundary of $\mathcal{R}$ are $(\alpha_1,\alpha_2)=(0,\pm\alpha_2^{\lim})$ and $(\alpha_1,\alpha_2)=\left(\pm\frac{\pi}{6},0\right)$.

Proof First we examine the values of $\mathcal{D}(x,y)$ on the boundary of $[3,4]\times\left[\frac{3}{2},4\right]$:

$$\mathcal{D}\left(x, \tfrac{3}{2}\right) = \tfrac{27}{8}(x-4)(x^2-2x+2), \quad \mathcal{D}(x,4) = (x-3)(3+8x)^2,$$
$$\mathcal{D}(3,y) = 27(y-4)(y-1)^2, \quad \mathcal{D}(4,y) = (16y-15)(2y-3)^2. \tag{36}$$

We claim that, for any $y_0 \in \left[\frac{3}{2}, 4\right]$, the polynomial $\mathcal{D}(x, y_0)$ has exactly one root in $[3,4]$. Indeed, we have $\mathcal{D}(3, y_0) \le 0 \le \mathcal{D}(4, y_0)$ and thus $\mathcal{D}(x, y_0)$ has at least one such root. The x–derivative of $\mathcal{D}$ is

$$\partial_x \mathcal{D}(x,y) = 3(x-3)y^2(xy+3y-6) + 27(y-1)^3,$$

which is positive when $x \in [3,4]$ and $y \in \left[\frac{3}{2}, 4\right]$. Thus $\mathcal{D}(x, y_0)$ is increasing, and the root is unique.

Similarly, we claim that, for any $x_0 \in [3,4]$, the polynomial $\mathcal{D}(x_0, y)$ has a unique root in $\left[\frac{3}{2}, 4\right]$. It is clear from (36) when $x_0 = 4$; there the root is $y = \frac{3}{2}$. Now suppose $3 \le x_0 < 4$. Arguing as before, we have $\mathcal{D}\left(x_0, \frac{3}{2}\right) < 0 \le \mathcal{D}(x_0, 4)$. However, it is not true that $\mathcal{D}(x_0, y)$ is a monotone function of y. The partial derivative of $\mathcal{D}(x,y)$ with respect to y is

$$\partial_y \mathcal{D}(x,y) = 3x(x^2y^2 - 6xy - 18y + 27).$$

Therefore, for a fixed $x_0 \in [3,4)$ we have $\partial_y \mathcal{D}\left(x_0, \frac{3}{2}\right) = \frac{27}{4}x_0^2(x_0-4) < 0$. Since $\mathcal{D}(x_0, y)$ is a cubic with leading coefficient $x_0^3 > 0$ and such that both $\mathcal{D}\left(x_0, \frac{3}{2}\right)$ and $\partial_y \mathcal{D}\left(x_0, \frac{3}{2}\right)$ are negative, we see that $\mathcal{D}(x_0, y)$ has exactly one zero in $\left(\frac{3}{2}, \infty\right)$. Since $\mathcal{D}(x_0, 4) \ge 0$, this zero must lie in $\left(\frac{3}{2}, 4\right]$, as claimed.

Thus the zero-locus of $\mathcal{D}(x,y)$ in $[3,4] \times \left[\frac{3}{2}, 4\right]$ is the graph of a continuous bijection connecting the two points $(3,4)$ and $\left(4, \frac{3}{2}\right)$. The polynomial $\mathcal{D}(x,y)$ is positive in the part of $[3,4] \times \left[\frac{3}{2}, 4\right]$ above the zero-locus, that is, containing the point $(x,y) = (4,4)$ (see Figure 10). Likewise, it is negative in the part below the zero locus, that is, containing the point $(x,y) = \left(3, \frac{3}{2}\right)$. Changing coordinates to (α_1, α_2), we see that the zero locus of $\mathcal{D}(4\cos^2\alpha_1, 4\cos^2\alpha_2)$ in the rectangle $\left[0, \frac{\pi}{6}\right] \times [0, \alpha_2^{\lim}]$ is the graph of a

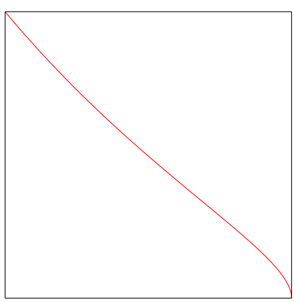


Figure 10: The null locus of $\mathcal{D}(x,y)$ in the rectangle $[3,4] \times \left[\frac{3}{2}, 4\right]$

continuous bijection connecting the points $(\alpha_1, \alpha_2) = \left(\frac{\pi}{6}, 0\right)$ and $(0, \alpha_2^{\lim})$. Moreover, $\mathcal{D}$ is positive on the part below this curve, in particular on the interval $\alpha_1 = 0$ and $0 \le \alpha_2 < \alpha_2^{\lim}$ and the interval $\alpha_2 = 0$ and $0 \le \alpha_1 < \frac{\pi}{6}$. The region $\mathcal{Z}$ is the union of the four copies of this region by the symmetries about the horizontal and vertical coordinate axes. It is clearly a disc and contains the relevant parts of the axes. This completes the proof. □

Combining Lemmas 7.3 and 7.4 proves Proposition 4.6.

7.3 Condition for no triple intersections: proof of Proposition 4.5

In this section we find a condition on (α_1, α_2) that characterises the set $\mathcal{Z}$ where the triple intersection of isometric spheres $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ is empty.

Lemma 7.5 *The triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ is empty if and only if $f_{\alpha_1,\alpha_2}(\alpha) > 0$ for all $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, where*

$$(37)\quad f_{\alpha_1,\alpha_2}(\alpha) = 4\cos^2\left(\tfrac{1}{2}\alpha - \tfrac{1}{2}\alpha_1\right) + 2\cos(\alpha - \alpha_1) + 8\cos\alpha\cos\alpha_1 - 16\sqrt{\cos\alpha\cos\alpha_1}\cos\left(\tfrac{1}{2}\alpha - \tfrac{1}{2}\alpha_1\right)|\sin\alpha_2|.$$

Proof By Corollary 7.2, it is enough to show that $f^{[0]}_{\alpha_1,\alpha_2} + f^{[-1]}_{\alpha_1,\alpha_2} > 0$. This sum is made explicit in (33). In view of the second part of Proposition 4.7, we can restrict our attention to showing that the triple intersection $\mathcal{I}_0^+ \cap \mathcal{I}_0^- \cap \mathcal{I}_{-1}^-$ contains no points of $\partial \boldsymbol{H}^2_{\mathbb{C}}$. That is, we may assume $w = \pm\sqrt{2\cos\alpha}$. Using the first part of Proposition 4.7 we restrict our attention to points m in the meridian $\mathfrak{m}$ where $\beta = \frac{1}{2}(\pi - \alpha_1)$. The triple intersection is empty if and only if the sum $f^{[0]}_{\alpha_1,\alpha_2}(q) + f^{[-1]}_{\alpha_1,\alpha_2}(q)$ is positive for any value of α, where $q = g\left(\alpha, \pi - \frac{1}{2}\alpha_1, \pm\sqrt{2\cos\alpha}\right)$. When $w\sin\alpha_2$ is negative, all terms in (33) are positive. Therefore we may suppose $w\sin\alpha_2 = \sqrt{2\cos\alpha_1}\,|\sin\alpha_2| \ge 0$. Substituting these values in the expression for $f^{[0]}_{\alpha_1,\alpha_2}(q) + f^{[-1]}_{\alpha_1,\alpha_2}(q)$ given in (33) gives the function $f_{\alpha_1,\alpha_2}(\alpha)$ in (37). □

We want to convert (37) into a polynomial expression in a function of α. The numerical condition given in the statement of Proposition 4.5 will follow from the next lemma.

Lemma 7.6 *If $\alpha \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ is a zero of f_{α_1,α_2} then $T_\alpha = \tan\left(\frac{1}{2}\alpha\right) \in [-1, 1]$ is a root of the quartic polynomial $L_{\alpha_1,\alpha_2}(T)$, where*

$$(38)\quad \begin{aligned} L_{\alpha_1,\alpha_2}(T) = {} & T^4(2x_1^4x_2^4 - 4x_1^2x_2^4 + x_1^4 + 10x_1^2 + 1) - 8T^3\sin\alpha_1(x_1^2x_2^4 - x_1^2 - 1) \\ & - 2T^2(2x_1^4x_2^4 + 3x_1^4 - 9) + 8T\sin\alpha_1(x_1^2x_2^4 - x_1^2 + 1) \\ & + (2x_1^4x_2^4 + 4x_1^2x_2^4 + x_1^4 - 10x_1^2 + 1). \end{aligned}$$

Proof Squaring the two lines of (37) and using $\sqrt{2\cos\alpha_1}\,|\sin\alpha_2| \geq 0$, we see that the condition $f_{\alpha_1,\alpha_2}(\alpha) = 0$ is equivalent to

(39) $$(1+2\cos(\alpha-\alpha_1)+4\cos\alpha\cos\alpha_1)^2 = 64\cos\alpha\cos\alpha_1\cos^2\bigl(\tfrac{1}{2}(\alpha-\alpha_1)\bigr)\sin^2\alpha_2.$$

After rearranging and expanding, we obtain the following polynomial equation in $\cos\alpha$ and $\sin\alpha$:

$$\begin{aligned} 0 = 4(8\cos^2\alpha_1\cos^2\alpha_2 + 2\cos^2\alpha_1 - 1)\cos^2\alpha \\ + 8\cos\alpha_1\sin\alpha_1(4\cos^2\alpha_2 - 1)\cos\alpha\sin\alpha \\ + 4\cos\alpha_1(8\cos^2\alpha_2 - 5)\cos\alpha + 4\sin\alpha_1\sin\alpha - 4\cos^2\alpha_1 + 5. \end{aligned}$$

Substituting $\tan\bigl(\tfrac{1}{2}\alpha\bigr) = T$, $2\cos\alpha_1 = x_1^2$ and $2\cos\alpha_2 = x_2^2$ into this equation gives $L_{\alpha_1,\alpha_2}(T)$. □

Before proving Proposition 4.5, we analyse the situation on the axes $\alpha_1 = 0$ and $\alpha_2 = 0$.

Lemma 7.7 *Let $L_{\alpha_1,\alpha_2}(T)$ be given by (38).*

(1) *When $\alpha_2 = 0$ and $-\frac{\pi}{6} < \alpha_1 < \frac{\pi}{6}$, the polynomial $L_{\alpha_1,0}(T)$ has two real double roots T_- and T_+, where $T_- < -1$ and $T_+ > 1$, and no other roots.*

(2) *When $\alpha_1 = 0$ and $0 < \alpha_2 < \alpha_2^{\lim}$ or $-\alpha_2^{\lim} < \alpha_2 < 0$, the polynomial $L_{0,\alpha_2}(T)$ has no real roots.*

Proof First, substituting $\alpha_2 = 0$ in (38) we find $L_{(\alpha_1,0)} = M_{\alpha_1}(T)^2$, where

$$M_{\alpha_1}(T) = T^2(3x_1^2 - 1) - 4T\sin\alpha_1 - (3x_1^2 + 1).$$

The condition on α_1 guarantees that $3x_1^2 - 1 > 0$ and so, as T tends to $\pm\infty$, also $M_{\alpha_1}(T)$ tends to $+\infty$. On the other hand,

$$M_{\alpha_1}(-1) = 4\sin\alpha_1 - 2 < 0, \quad M_{\alpha_1}(1) = -4\sin\alpha_1 - 2 < 0.$$

Therefore $M_{\alpha_1}(T)$ has two real roots $T_- < -1$ and $T_+ > 1$, as claimed. Since $M_{\alpha_1}(T)$ is quadratic, it cannot have any more roots. In particular, it is negative for $-1 \leq T \leq 1$.

Secondly, we substitute $\alpha_1 = 0$ in (38), giving

$$L_{0,\alpha_2}(T) = \bigl(5T^2 - \tfrac{1}{5}(8x_2^4 + 3)\bigr)^2 + \tfrac{32}{25}(2x_2^4 - 3)(4 - x_2^4).$$

When $\alpha_2 \in (-\alpha_2^{\lim}, \alpha_2^{\lim})$ and $\alpha_2 \neq 0$, we have $x_2^4 = 4\cos^2\alpha_2 \in \bigl(\tfrac{3}{2}, 4\bigr)$. In particular, this means that $(2x_2^4 - 3)(4 - x_2^4) > 0$ and so $L_{0,\alpha_2}(T)$ has no real roots, proving the second part. □

We note that if $\alpha_1 = \alpha_2 = 0$ then $L_{0,0}(T)$ has double roots at $T = \pm\sqrt{7/5}$ and, if $\alpha_1 = 0$ and $\alpha_2 = \pm\alpha_2^{\lim}$, then $L_{0,\pm\alpha_2^{\lim}}(T)$ has double roots at $T = \pm\sqrt{3/5}$.

Lemma 7.8 *If* $(\alpha_1,\alpha_2)\in\mathcal{Z}$ *then the polynomial* $L_{\alpha_1,\alpha_2}(T)$ *has no roots* T *in* $[-1,1]$.

Proof We analyse the number, type (real or nonreal) and location of roots of the polynomial $L_{\alpha_1,\alpha_2}(T)$ when $(\alpha_1,\alpha_2)\in\mathcal{R}$. As $L_{\alpha_1\alpha_2}(T)$ has real coefficients, whenever it has only simple roots, its root set is of one of the following types:

(a) two pairs of complex conjugate nonreal simple roots,

(b) a pair of nonreal complex conjugate simple roots and two simple real roots,

(c) four simple real roots.

But the set of roots of a polynomial is a continuous map (in bounded degree) for the Hausdorff distance on compact subsets of $\mathbb{C}$. In particular, the root set type of $L_{\alpha_1,\alpha_2}(T)$ is a continuous function of α_1 and α_2. This implies that it is not possible to pass from one of the above types to another without passing through a polynomial having a double root.

We compute the discriminant $\Delta_{\alpha_1,\alpha_2}$ of $L_{\alpha_1,\alpha_2}(T)$ (a computer may be useful to do so):

$$\Delta_{\alpha_1,\alpha_2}=2^{16}x_1^4(x_1^4+1)^2(2x_1^2(2-x_1^2)(4-x_2^4)+(3x_1^2-1)^2)(4-x_2^4)^2\cdot\mathcal{D}(x_1^4,x_2^4), \tag{40}$$

where $\mathcal{D}(x,y)$ is as in Proposition 4.5, and $x_i=\sqrt{2\cos\alpha_i}$. The polynomial $L_{\alpha_1,\alpha_2}(T)$ has a multiple root in $\mathbb{C}$ if and only if $\Delta_{\alpha_1\alpha_2}=0$. Let us examine the different factors.

- The first two factors x_1^4 and $(x_1^4+1)^2$ are positive when $(\alpha_1,\alpha_2)\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2$.
- Note that $(2-x_1^2)(4-x_2^4)\geq 0$ and $(3x_1^2-1)^2>0$ when $\sqrt{3}\leq x_1^2\leq 2$ and $x_2^4\leq 4$, and so the third factor is positive.

Thus, the only factors of $\Delta_{\alpha_1,\alpha_2}$ that can vanish on $\mathcal{R}$ are $(4-x_2^4)^2=16\sin^4\alpha_2$ and $\mathcal{D}(x_1^4,x_2^4)$. In particular $L_{\alpha_1,\alpha_2}(T)$ has a multiple root in $\mathbb{C}$ if and only if one of these two factors vanishes. We saw in Proposition 4.6 that the subset of $\mathcal{R}$ where $\mathcal{D}(x_1^4,x_2^4)>0$ is a topological disc $\mathcal{Z}$, symmetric about the α_1 and α_2 axes and intersecting them in the intervals $\left\{\alpha_2=0,\ -\frac{\pi}{6}<\alpha_1<\frac{\pi}{6}\right\}$ and $\{\alpha_1=0,\ -\alpha_2^{\lim}<\alpha_2<\alpha_2^{\lim}\}$. Therefore, the rectangle $\mathcal{R}$ contains two open discs on which $\Delta_{\alpha_1,\alpha_2}>0$, namely

$$\mathcal{Z}^+=\{(\alpha_1,\alpha_2)\in\mathcal{Z}:\alpha_2>0\},\quad \mathcal{Z}^-=\{(\alpha_1,\alpha_2)\in\mathcal{Z}:\alpha_2<0\}.$$

These two sets each contain an open interval of the α_2 axis. We saw in the second part of Lemma 7.7 that on both these intervals $L_{\alpha_1,\alpha_2}(T)$ has no real roots, that is its roots are of type (a). Therefore it has no real roots on all of $\mathcal{Z}^+$ and $\mathcal{Z}^-$.

Only those points of $\mathcal{Z}$ in the interval $\left\{\alpha_2=0,\ -\frac{\pi}{6}<\alpha_1<\frac{\pi}{6}\right\}$ still need to be considered. We saw in the first part of Lemma 7.7 that, for such points, $L_{\alpha_1,\alpha_2}(T)$ has no roots with $-1\leq T\leq 1$. This completes the proof of Proposition 4.5. □

7.4 Pairwise intersection: proof of Proposition 4.8

Proposition 4.8 will follow from the next lemma.

Lemma 7.9 *If $0 < x \le 4$ and $\mathcal{D}(x,y) \ge 0$ then $xy \ge 6$, with equality if and only if $(x,y) = \left(4, \frac{3}{2}\right)$.*

Proof Substituting $y = 6/x$ in (35) and simplifying, we obtain

$$\mathcal{D}\left(x, \frac{6}{x}\right) = -\frac{27(x-4)(x-9)}{x}.$$

When $0 < x \le 4$ we see immediately that this is nonpositive and equals zero if and only if $x = 4$. This means that $xy - 6$ has a constant sign on the region where $\mathcal{D}(x,y) > 0$. Checking at $(x,y) = (4,4)$, we see that it is positive. □

Proof of Proposition 4.8 To prove the disjointness of the given isometric spheres we calculate the Cygan distance between their centres. Since all the isometric spheres have radius 1, if we can show their centres are a Cygan distance at least 2 apart, then the spheres are disjoint. (Note that the Cygan distance is not a path metric, so it may be the distance is less than 2 but the spheres are still disjoint. This will not be the case in our examples.)

The centre of I_k^+ is

$$A^k(p_B) = \left[\tfrac{1}{\sqrt{2}} k x_1 x_2^2, k x_1^2 x_2^2 \sin\alpha_2\right];$$

see Proposition 4.2. We will show that $d_{\mathrm{Cyg}}(A^k(p_B), p_B)^4 > 16$ when $k^2 \ge 4$ and $(\alpha_1, \alpha_2) \in \mathcal{R}$, that is, $(x_1^4, x_2^4) \in [3,4] \times \left[\frac{3}{2}, 4\right]$:

$$d_{\mathrm{Cyg}}(A^k(p_B), p_B)^4 = \tfrac{1}{4}(k^4 x_1^4 x_2^8 + k^2 x_1^4 x_2^4 (4 - x_2^4)) \ge \tfrac{27}{16} k^4.$$

This number is greater than 16 when $k \ge 2$ or $k \le -2$ as claimed. Using Proposition 4.2 again, the centre of $\mathcal{I}_k^-$ is

$$A^k(p_{AB}) = \left[\tfrac{1}{\sqrt{2}}(k x_1 x_2^2 + x_1 e^{i\alpha_2}), -\sin\alpha_1\right].$$

We suppose that the pair $(x_1^4, x_2^4) \in [3,4] \times \left[\frac{3}{2}, 4\right]$ satisfies $x_1^4 x_2^4 \ge 6$, which is valid for $(\alpha_1, \alpha_2) \in \mathcal{Z}$ by Lemma 7.9. Then

$$\begin{aligned} d_{\mathrm{Cyg}}(A^k(p_{AB}), p_B)^4 &= \tfrac{1}{4}((k(k+1)x_1^2 x_2^4 + x_1^2)^2 + 4 - x_1^4) \\ &= 1 + \tfrac{1}{4}(k^2(k+1)^2 x_1^4 x_2^8 + 2k(k+1) x_1^4 x_2^4) \\ &\ge \left(\tfrac{3}{2}k(k+1) + 1\right)^2. \end{aligned}$$

This number is at least 16 when $k \ge 1$ or $k \le -2$, as claimed. Moreover, we have equality exactly when $k = 1$ or $k = -2$ and when $x_1^4 x_2^4 = 6$ and $x_2^4 = \frac{3}{2}$; that is, when $(x_1^4, x_2^4) = \left(4, \frac{3}{2}\right)$. □

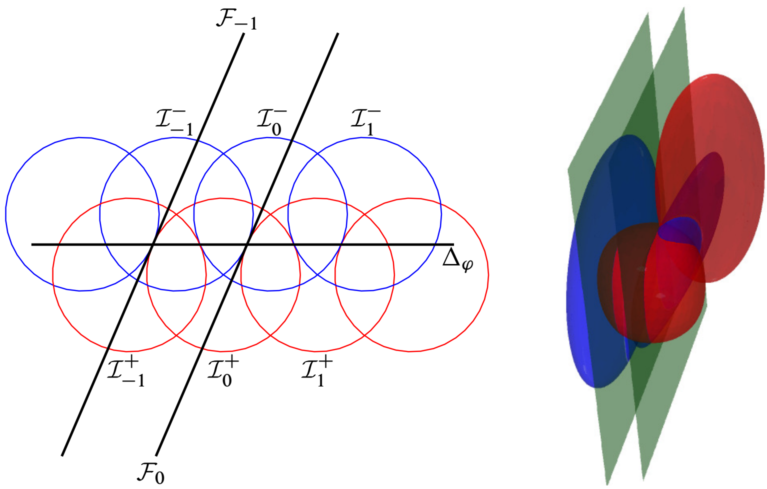


Figure 11: Vertical projection and realistic view of the isometric spheres and the fans F_0 and F_{-1} for the parameter values $\alpha_1 = 0$, $\alpha_2 = \alpha_2^{\mathrm{lim}}$. Compare with Figure 5.

7.5 The ideal boundary $\partial_\infty D$ is a cylinder: proof of Proposition 6.8

To prove Proposition 6.8, we adopt the following strategy:

Step 1 First, we intersect D with a fundamental domain D_A for the action of A on the Heisenberg group. The domain D_A is bounded by two parallel vertical planes F_{-1} and F_0 that are boundaries of *fans* in the sense of [17]. These two fans are such that $A(F_{-1}) = F_0$ (see Figure 11 for a view of the situation in vertical projection). We analyse the intersections of F_0 and F_{-1} with D, and show that they are topological circles, denoted by c_{-1} and c_0 with $A(c_{-1}) = c_0$.

Step 2 Secondly, we consider the subset of the complement of D which is contained in D_A, and prove that it is a 3–dimensional ball that intersects F_{-1} and F_0 along topological discs (bounded by c_{-1} and c_0). This proves that $D \cap D_A$ is the complement a solid tube in D_A, which is unknotted using Lemma 6.7. Finally, we prove that, gluing together copies by powers of A of $D \cap D_A$, we indeed obtain the complement of a solid cylinder.

We construct a fundamental domain D_A for the cyclic group $\langle A \rangle$ of Heisenberg translations. The domain D_A will be bounded by two fans, chosen to intersect as few bisectors as possible. The fan F_0 will pass through $p_{ST^{-1}}$ and will be tangent to both $\mathcal{I}_1^+$ and $\mathcal{I}_{-1}^-$; compare Figure 11. Similarly, $F_{-1} = A^{-1}(F_0)$ will pass through $A^{-1}(p_{ST^{-1}}) = p_{TST}$ and be tangent to both $\mathcal{I}_0^+$ and $\mathcal{I}_{-2}^-$. We first give F_0 and F_{-1}

in terms of horospherical coordinates and then we give them in terms of their own geographical coordinates (see [17]). In horospherical coordinates they are

$$F_0 = \left\{[x+iy,t] : 3x\sqrt{3} - y\sqrt{5} = \tfrac{\sqrt{2}}{2}\right\}, \tag{41}$$

$$F_{-1} = \{[x+iy,t] : 3x\sqrt{3} - y\sqrt{5} = -4\sqrt{2}\}. \tag{42}$$

This leads to the definition of D_A:

$$D_A = \left\{[x+iy,t] : -4\sqrt{2} \le 3x\sqrt{3} - y\sqrt{5} \le \tfrac{\sqrt{2}}{2}\right\}. \tag{43}$$

We choose geographical coordinates (ξ,η) on F_0: the lines where ξ is constant (resp. η is constant) are boundaries of complex lines (resp. Lagrangian planes). These coordinates correspond to the double foliation of fans by real planes and complex lines, which is described in [17, Section 5.2]. The particular choice is made so that the origin is the midpoint of the centres of $\mathcal{I}_0^+$ and $\mathcal{I}_0^-$. Doing so gives the fan F_0 as the set of points $f(\xi,\eta)$:

$$f(\xi,\eta) = \left\{\left[\tfrac{1}{4\sqrt{2}}(\sqrt{5}\xi + \sqrt{3} + 3i\sqrt{3}\xi + i\sqrt{5}), \eta - \tfrac{\xi}{4}\right] : \xi,\ \eta \in \mathbb{R}\right\}.$$

The standard lift of $f(\xi,\eta)$ is given by

$$\boldsymbol{f}(\xi,\eta) = \begin{bmatrix} -\xi^2 - \frac{\sqrt{15}}{4}\xi - \frac{1}{4} + i\eta - \frac{i}{4}\xi \\ \frac{\sqrt{5}}{4}\xi + \frac{\sqrt{3}}{4} + \frac{3i\sqrt{3}}{4}\xi + \frac{i\sqrt{5}}{4} \\ 1 \end{bmatrix}.$$

Using the convexity of Cygan spheres, we see that their intersection with F_0 (or F_{-1}) is one of: empty, a point or a topological circle. For the particular fans and isometric spheres of interest to us, the possible intersections are summarised in the following result:

Proposition 7.10 *The intersections of the fans F_{-1} and F_0 with the isometric spheres $\mathcal{I}_k^\pm$ are empty, except for those indicated in the following table:*

$\cap$	$\mathcal{I}_{-2}^-$	$\mathcal{I}_{-2}^+$	$\mathcal{I}_{-1}^-$	$\mathcal{I}_{-1}^+$	$\mathcal{I}_0^-$	$\mathcal{I}_0^+$	$\mathcal{I}_1^-$	$\mathcal{I}_1^+$
F_0	$\varnothing$	$\varnothing$	$\{p_{ST^{-1}}\}$	$\varnothing$	a circle	a circle	$\varnothing$	$\{p_{ST^{-1}}\}$
F_{-1}	$\{p_{TST}\}$	$\varnothing$	a circle	a circle	$\varnothing$	$\{p_{TST}\}$	$\varnothing$	$\varnothing$

Moreover, the point $p_{S^{-1}T}$ belongs to the interior of D_A. The parabolic fixed points $A^k(p_{ST^{-1}})$ lie outside D_A for all $k \ge 1$ and $k \le -1$; parabolic fixed points $A^k(p_{S^{-1}T})$ lie outside D_A for all $k \ne 0$.

A direct consequence of this proposition is that the only point in the closure of the quadrilateral $\mathcal{Q}_{-1}^-$ and the bigon $\mathcal{B}_{-1}^-$ that lie on F_0 is their vertex p_{TST}.

Proof The part about intersections of fans and isometric spheres is proved easily by projecting vertically onto $\mathbb{C}$, as in the proof of Proposition 4.8 (see Figure 11). Note that as isometric spheres are strictly convex, their intersections with a plane is either empty or a point or a topological circle. The part about the parabolic fixed points is a direct verification using (41) as well as (27). □

We need to be slightly more precise about the intersection of F_0 with $\mathcal{I}_0^+$ and $\mathcal{I}_0^-$.

Proposition 7.11 *The intersection of F_0 with $\mathcal{I}_0^+ \cup \mathcal{I}_0^-$ (and thus with ∂D) is a topological circle c_0, which is the union of two topological segments c_0^+ and c_0^-, where the segment $c_0^\pm$ is the part of $F_0 \cap \mathcal{I}_0^\pm$ that is outside $\mathcal{I}_0^\mp$. The two segments c_0^+ and c_0^- have the same endpoints; one of them is $p_{ST^{-1}}$, and we will denote the other by q_0. Moreover, the point q_0 lies on the segment $[p_{STS}, p_{S^{-1}T}]$ of $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$.*

The point q_0 appears in Figures 12, 13 and 14.

Proof The point $f(\xi,\eta)$ of the fan F_0 lies on $\mathcal{I}_0^+$ whenever $1 = |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_B\rangle|$ and on $\mathcal{I}_0^-$ whenever $1 = |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_{AB}\rangle|$. We first find all points on $F_0 \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^-$. These correspond to simultaneous solutions to

$$1 = |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_B\rangle| = |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_{AB}\rangle| \tag{44}$$

Computing these products and rearranging, we obtain

$$|\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_B\rangle|^2 = \left(\xi^2 + \tfrac{1}{4}\right)^2 + \xi^2 + \eta^2 + \tfrac{1}{2}\xi\left(\sqrt{15}\xi^2 + \tfrac{\sqrt{15}}{4} - \eta\right),$$
$$|\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p}_{AB}\rangle|^2 = \left(\xi^2 + \tfrac{1}{4}\right)^2 + \xi^2 + \eta^2 - \tfrac{1}{2}\xi\left(\sqrt{15}\xi^2 + \tfrac{\sqrt{15}}{4} - \eta\right).$$

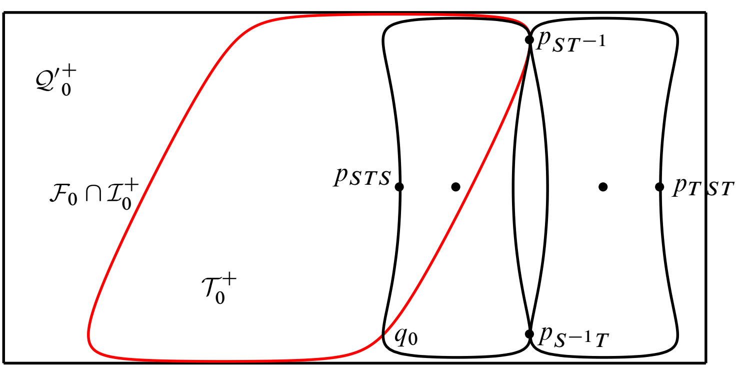


Figure 12: The intersection of F_0 with $\mathcal{I}_0^+$ drawn on $\mathcal{I}_0^+$ in geographical coordinates

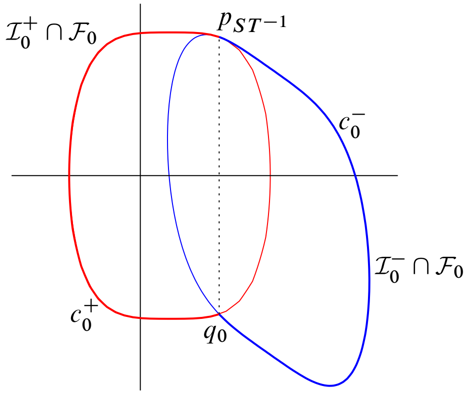


Figure 13: The intersection of F_0 with $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$. The disc $\mathcal{D}_0$ is the interior of $c_0 = c_0^+ \cap c_0^-$. The two segments c_0^+ and c_0^- are the thicker parts of $F_0 \cap \mathcal{I}_0^+$ and $F_0 \cap \mathcal{I}_0^-$.

Subtracting, we see that solutions to (44) must either have $\xi = 0$ or $\eta = \sqrt{15}\left(\xi^2 + \frac{1}{4}\right)$. Substituting these solutions into $1 = |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p_B}\rangle|^2$, we see first that $\xi = 0$ implies $1 = \eta^2 + \frac{1}{16}$, and secondly that $\eta = \sqrt{15}\left(\xi^2 + \frac{1}{4}\right)$ implies

$$1 = \left(\xi^2 + \tfrac{1}{4}\right)^2 + \xi^2 + 15\left(\xi^2 + \tfrac{1}{4}\right)^2 = (4\xi^2 + 1)^2 + \xi^2.$$

Clearly the only solution to this equation is $\xi = 0$. So both cases lead to the solutions $(\xi,\eta) = \left(0, \pm\frac{\sqrt{15}}{4}\right)$. Thus the only points satisfying (44), that is, the points in $F_0 \cap \mathcal{I}_0^+ \cap \mathcal{I}_0^-$, are

$$f\left(0, \tfrac{\sqrt{15}}{4}\right) = \left[\tfrac{\sqrt{3}+i\sqrt{5}}{4\sqrt{2}}, \tfrac{\sqrt{15}}{4}\right] \quad \text{and} \quad f\left(0, -\tfrac{\sqrt{15}}{4}\right) = \left[\tfrac{\sqrt{3}+i\sqrt{5}}{4\sqrt{2}}, -\tfrac{\sqrt{15}}{4}\right].$$

Note that the first of these points is $p_{ST^{-1}}$. We call the other point q_0.

These two points divide $F_0 \cap \mathcal{I}_0^+$ and $F_0 \cap \mathcal{I}_0^-$ into two arcs. It remains to decide which of these arcs is outside the other isometric sphere. Clearly

$$|\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p_B}\rangle| > |\langle \boldsymbol{f}(\xi,\eta), \boldsymbol{p_{AB}}\rangle| \quad \text{if and only if} \quad \xi\left(\sqrt{15}\xi^2 + \tfrac{\sqrt{15}}{4} - \eta\right) > 0.$$

Close to $\eta = -\frac{\sqrt{15}}{4}$ we see this quantity changes sign only when ξ does. This means that if $f(\xi,\eta) \in \mathcal{I}_0^-$ with $\xi > 0$ then $f(\xi,\eta)$ is in the exterior of $\mathcal{I}_0^+$. Similarly, if $f(\xi,\eta) \in \mathcal{I}_0^+$ with $\xi < 0$ then $f(\xi,\eta)$ is in the exterior of $\mathcal{I}_0^-$. In other words, c_0^+ is the segment of $F_0 \cap \mathcal{I}_0^+$ where $\xi < 0$ and c_0^- is the segment of $F_0 \cap \mathcal{I}_0^-$ where $\xi > 0$.

Finally, consider the involution $I_2 = SI_1$ in $\mathrm{PU}(2,1)$ from the proof of Proposition 3.8. (Note that since $\alpha_1 = 0$, this involution conjugates $\Gamma^{\lim}$ to itself.) The involution I_2 preserves F_0, acting on it by sending $f(\xi,\eta)$ to $f(-\xi,\eta)$, and hence interchanging

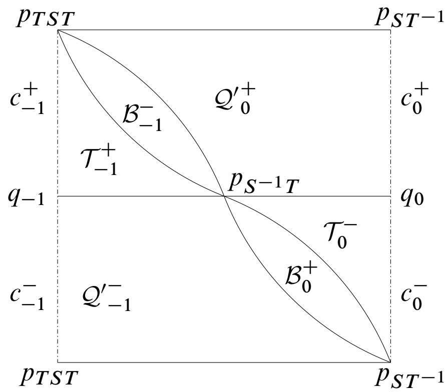


Figure 14: A combinatorial picture of the intersection of ∂D with D_A. The top and bottom lines are identified. The curve c_0 corresponds to the right-hand side of the figure.

the components of its complement. In Heisenberg coordinates I_2 is given by

$$I_2\colon [x+iy,\, t] \longleftrightarrow \left[-x-iy+\sqrt{3/8}+i\sqrt{5/8},\, t-\sqrt{5/2}\,x+\sqrt{3/2}\,y\right]. \tag{45}$$

As I_2 is elliptic and fixes the point q_∞, it is a Cygan isometry (see Section 2.4). Since it interchanges p_B and p_{AB}, it also interchanges $\mathcal{I}_0^+$ and $\mathcal{I}_0^-$. Hence their intersection is preserved setwise. The involution I_2 also interchanges $p_{S^{-1}T}$ and p_{STS} contained in $\mathcal{I}_0^+ \cap \mathcal{I}_0^-$ (but not on F_0). Therefore, these two points lie in different components of the complement of F_0. Hence there must be a point of F_0 on the segment $[p_{S^{-1}T}, p_{STS}]$. This point cannot be $p_{ST^{-1}}$, and so must be q_0 (see Figure 12). □

Let D^c denote the closure of the complement of D in $\partial \boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$.

Proposition 7.12 *The closure of the intersection $D^c \cap D_A$ is a solid tube homeomorphic to a 3–ball.*

Proof We describe the combinatorial cell structure of $D^c \cap D_A$; see Figure 14. Using Proposition 7.11, it is clear D^c intersects F_0 in a topological disc whose boundary circle is made up of two edges $c_0^\pm$ and two vertices $p_{ST^{-1}}$ and q_0. Combinatorially, this is a bigon. Applying A^{-1} we see D^c intersects F_{-1} in a bigon with boundary made up of edges $c_{-1}^\pm$ and two vertices p_{TST} and q_{-1}.

Moreover, Proposition 7.11 immediately implies that c_0 cuts $\mathcal{Q}_0^\pm$ into a quadrilateral and a triangle, which we denote by $\mathcal{Q}_0'^\pm$ and $\mathcal{T}_0^\pm$. Since D_A contains $p_{S^{-1}T}$ and p_{TST}, we see that D_A contains $\mathcal{Q}_0'^+$ and $\mathcal{T}_0^-$. These have vertex sets

$$\{p_{ST^{-1}},\, p_{TST},\, p_{S^{-1}T},\, q_0\} \quad \text{and} \quad \{p_{S^{-1}T},\, p_{ST^{-1}},\, q_0\},$$

respectively. Applying A^{-1} we see that c_{-1} cuts $\mathcal{Q}_{-1}^\pm$ into a quadrilateral, denoted by $\mathcal{Q}'^{\pm}_{-1}$, and a triangle, denoted by $\mathcal{T}^{\pm}_{-1}$. Of these, the quadrilateral $\mathcal{Q}'^{-}_{-1}$ and the triangle $\mathcal{T}^{+}_{-1}$ lie in D_A. Finally, the bigons $\mathcal{B}^{+}_{0}$ and $\mathcal{B}^{-}_{-1}$ also lie in D_A.

In summary, the boundary of $D^c \cap D_A$ has a combinatorial cell structure with five vertices $\{p_{ST^{-1}}, p_{S^{-1}T}, p_{TST}, q_0, q_{-1}\}$ and eight faces,

$$\{\mathcal{Q}'^{+}_{0}, \mathcal{Q}'^{-}_{-1}, \mathcal{T}^{-}_{0}, \mathcal{T}^{+}_{-1}, \mathcal{B}^{+}_{0}, \mathcal{B}^{-}_{-1}, F_0 \cap D^c, F_{-1} \cap D^c\}.$$

These are respectively two quadrilaterals, two triangles and four bigons. Therefore, in total the cell structure has $\frac{1}{2}(2\times 4 + 2\times 3 + 4\times 2) = 11$ edges. Therefore the Euler characteristic of $\partial(D^c \cap D_A)$ is

$$\chi(\partial(D^c \cap D_A)) = 5 - 11 + 8 = 2.$$

Hence $\partial(D^c \cap D_A)$ is indeed a sphere. This means $D^c \cap D_A$ is a ball, as claimed. □

Remark 7.13 The combinatorial structure described on Figure 14 is quite simple. However, the geometric realisation of this structure is much more intricate. As an example, there are fans F parallel to F_0 and F_{-1} whose intersections with D^c are disconnected. This means that the foliation described right after Proposition 6.8 that is used in the proof of Theorem 6.4 is actually quite "distorted".

Proposition 7.14 *There is a homeomorphism $\Psi_A\colon \mathbb{R}^2 \times [0,1] \to D_A$ that satisfies $\Psi_A(x,y,1) = A\Psi_A(x,y,0)$ and such that Ψ_A restricts to a homeomorphism from the exterior of $S^1 \times [0,1]$, that is, $\{(x,y,z) : x^2+y^2 \geq 1,\ 0 \leq z \leq 1\}$, to $D \cap D_A$.*

Proof We have shown in Proposition 7.12 that $D^c \cap D_A$ is a solid tube homeomorphic to a 3–ball and (using Proposition 7.11) that D^c intersects ∂D_A in two discs, one in F_0 bounded by c_0 and the other in F_{-1} bounded by c_{-1}. This means we can construct a homeomorphism Ψ^c_A from the solid cylinder $\{(x,y,z) : x^2+y^2 \leq 1,\ 0 \leq z \leq 1\}$ to $D^c \cap D_A$ such that the restriction of Ψ^c_A to $S^1 \times [0,1]$ is a homeomorphism to $\partial D \cap D_A$, with $\Psi^c_A\colon S^1 \times \{0\} \mapsto c_{-1}$ and $\Psi^c_A\colon S^1 \times \{1\} \mapsto c_0$. Adjusting Ψ^c_A if necessary, we can assume that $\Psi^c_A(x,y,1) = A\Psi^c_A(x,y,0)$.

Furthermore, in Lemma 6.7, we showed that D^c contains the invariant line Δ_φ of φ. This means that the cylinder $D^c \cap D_A$ is a thickening of $\Delta_\varphi \cap D_A$ and so, in particular, it cannot be knotted. Hence Ψ^c_A can be extended to a homeomorphism $\Psi_A\colon \mathbb{R}^2 \times [0,1] \to D_A$ satisfying $\Psi_A(x,y,1) = A\Psi_A(x,y,0)$. In particular, Ψ maps $\{(x,y,z) : x^2+y^2 \geq 1,\ 0 \leq z \leq 1\}$ homeomorphically to $D \cap D_A$, as claimed. □

Finally, we prove Proposition 6.8 by extending $\Psi_A\colon \mathbb{R}^2 \times [0,1] \to D_A$ equivariantly to a homeomorphism $\Psi\colon \mathbb{R}^3 \mapsto \partial \boldsymbol{H}^2_{\mathbb{C}} - \{q_\infty\}$. That is, if $(x,y,z+k) \in \mathbb{R}^3$ with $k \in \mathbb{Z}$ and $z \in [0,1]$, we define $\Psi(x,y,z+k) = A^k(x,y,z)$. Since $\Psi(x,y,1) = A\Psi(x,y,0)$, there is no ambiguity at the boundary.

References

[1] **H Akiyoshi**, **M Sakuma**, **M Wada**, **Y Yamashita**, *Punctured torus groups and 2–bridge knot groups, I*, Lecture Notes in Math. 1909, Springer (2007) MR

[2] **A F Beardon**, *The geometry of discrete groups*, Graduate Texts in Math. 91, Springer (1983) MR

[3] **S S Chen**, **L Greenberg**, *Hyperbolic spaces*, from "Contributions to analysis: a collection of papers dedicated to Lipman Bers" (L V Ahlfors, I Kra, B Maskit, L Nirenberg, editors), Academic Press, New York (1974) 49–87 MR

[4] **D Cooper**, **D D Long**, **M B Thistlethwaite**, *Flexing closed hyperbolic manifolds*, Geom. Topol. 11 (2007) 2413–2440 MR

[5] **M Culler**, **N M Dunfield**, **M Goerner**, **J R Weeks**, *SnapPy, a computer program for studying the geometry and topology of 3–manifolds* Available at `http://snappy.computop.org`

[6] **M Deraux**, *On spherical CR uniformization of 3–manifolds*, Exp. Math. 24 (2015) 355–370 MR

[7] **M Deraux**, *A 1–parameter family of spherical CR uniformizations of the figure eight knot complement*, Geom. Topol. 20 (2016) 3571–3621 MR

[8] **M Deraux**, **E Falbel**, *Complex hyperbolic geometry of the figure-eight knot*, Geom. Topol. 19 (2015) 237–293 MR

[9] **M Deraux**, **J R Parker**, **J Paupert**, *New non-arithmetic complex hyperbolic lattices*, Invent. Math. 203 (2016) 681–771 MR

[10] **D B A Epstein**, **C Petronio**, *An exposition of Poincaré's polyhedron theorem*, Enseign. Math. 40 (1994) 113–170 MR

[11] **E Falbel**, *A spherical CR structure on the complement of the figure eight knot with discrete holonomy*, J. Differential Geom. 79 (2008) 69–110 MR

[12] **E Falbel**, **A Guilloux**, **P-V Koseleff**, **F Rouillier**, **M Thistlethwaite**, *Character varieties for* $\mathrm{SL}(3,\mathbb{C})$*: the figure eight knot*, Exp. Math. 25 (2016) 219–235 MR

[13] **E Falbel**, **P-V Koseleff**, **F Rouillier**, *Representations of fundamental groups of 3–manifolds into* $\mathrm{PGL}(3,\mathbb{C})$*: exact computations in low complexity*, Geom. Dedicata 177 (2015) 229–255 MR

[14] **E Falbel**, **J R Parker**, *The geometry of the Eisenstein–Picard modular group*, Duke Math. J. 131 (2006) 249–289 MR

[15] **W M Goldman**, *Complex hyperbolic geometry*, Clarendon, New York (1999) MR

[16] **W M Goldman**, **J R Parker**, *Complex hyperbolic ideal triangle groups*, J. Reine Angew. Math. 425 (1992) 71–86 MR

[17] **W M Goldman**, **J R Parker**, *Dirichlet polyhedra for dihedral groups acting on complex hyperbolic space*, J. Geom. Anal. 2 (1992) 517–554 MR

[18] **M Heusener**, **V Munoz**, **J Porti**, *The* $\mathrm{SL}(3,\mathbb{C})$*–character variety of the figure eight knot*, preprint (2015) arXiv

[19] **L Keen**, **C Series**, *The Riley slice of Schottky space*, Proc. London Math. Soc. 69 (1994) 72–90 MR

[20] **B Martelli**, **C Petronio**, *Dehn filling of the "magic" 3–manifold*, Comm. Anal. Geom. 14 (2006) 969–1026 MR

[21] **G J Martin**, *On discrete isometry groups of negative curvature*, Pacific J. Math. 160 (1993) 109–127 MR

[22] **G D Mostow**, *On a remarkable class of polyhedra in complex hyperbolic space*, Pacific J. Math. 86 (1980) 171–276 MR

[23] **J R Parker**, *Complex hyperbolic Kleinian groups*, preprint To appear under Cambridge Univ. Press

[24] **J R Parker**, **I D Platis**, *Complex hyperbolic quasi-Fuchsian groups*, from "Geometry of Riemann surfaces" (F P Gardiner, G González-Diez, C Kourouniotis, editors), London Math. Soc. Lecture Note Ser. 368, Cambridge Univ. Press (2010) 309–355 MR

[25] **J R Parker**, **J Wang**, **B Xie**, *Complex hyperbolic* $(3,3,n)$ *triangle groups*, Pacific J. Math. 280 (2016) 433–453 MR

[26] **J R Parker**, **P Will**, *Complex hyperbolic free groups with many parabolic elements*, from "Geometry, groups and dynamics" (C S Aravinda, W M Goldman, K Gongopadhyay, A Lubotzky, M Mj, A Weaver, editors), Contemp. Math. 639, Amer. Math. Soc., Providence, RI (2015) 327–348 MR

[27] **J Paupert**, **P Will**, *Real reflections, commutators and cross-ratios in complex hyperbolic space*, Groups Geom. Dyn. 11 (2013) 311–352

[28] **A Pratoussevitch**, *Traces in complex hyperbolic triangle groups*, Geom. Dedicata 111 (2005) 159–185 MR

[29] **J G Ratcliffe**, *Foundations of hyperbolic manifolds*, 2nd edition, Graduate Texts in Math. 149, Springer (2006) MR

[30] **R E Schwartz**, *Degenerating the complex hyperbolic ideal triangle groups*, Acta Math. 186 (2001) 105–154 MR

[31] **R E Schwartz**, *Ideal triangle groups, dented tori, and numerical analysis*, Ann. of Math. 153 (2001) 533–598 MR

[32] **R E Schwartz**, *Complex hyperbolic triangle groups*, from "Proceedings of the International Congress of Mathematicians, II" (T Li, editor), Higher Ed. Press, Beijing (2002) 339–349 MR

[33] **R E Schwartz**, *A better proof of the Goldman–Parker conjecture*, Geom. Topol. 9 (2005) 1539–1601 MR

[34] **R E Schwartz**, *Spherical CR geometry and Dehn surgery*, Annals of Mathematics Studies 165, Princeton Univ. Press (2007) MR

[35] **W P Thurston**, *The geometry and topology of three-manifolds*, lecture notes, Princeton University (1979) Available at `http://msri.org/publications/books/gt3m`

[36] **P Will**, *Traces, cross-ratios and 2–generator subgroups of* $\mathrm{SU}(2,1)$, Canad. J. Math. 61 (2009) 1407–1436 MR

[37] **P Will**, *Bending Fuchsian representations of fundamental groups of cusped surfaces in* $\mathrm{PU}(2,1)$, J. Differential Geom. 90 (2012) 473–520 MR

[38] **P Will**, *Two-generator groups acting on the complex hyperbolic plane*, from "Handbook of Teichmüller theory, VI" (A Papadopoulos, editor), IMRA Lect. Math. Theor. Phys. 27, Eur. Math. Soc., Zürich (2016) 276–334

Department of Mathematical Sciences, Durham University
Durham, United Kingdom

Université Grenoble Alpes, Institut Fourier
Saint-Martin-d'Hères, France

`j.r.parker@durham.ac.uk`, `pierre.will@univ-grenoble-alpes.fr`

`http://maths.dur.ac.uk/~dma0jrp/`,
`https://www-fourier.ujf-grenoble.fr/~will/`

Proposed: Walter Neumann Received: 2 October 2015
Seconded: Danny Calegari, Jean-Pierre Otal Revised: 17 May 2016

Geometry & Topology 21 (2017) 3453–3466

The nilpotence theorem for the algebraic K–theory of the sphere spectrum

ANDREW J BLUMBERG
MICHAEL A MANDELL

We prove that in the graded commutative ring $K_*(\mathbb{S})$, all positive degree elements are multiplicatively nilpotent. The analogous statements also hold for $\mathrm{TC}_*(\mathbb{S})^\wedge_p$ and $K_*(\mathbb{Z})$.

19D10

1 Introduction

Much of the most exciting work in algebraic K–theory over the past 15 years has been aimed at the verification of the Quillen–Lichtenbaum conjecture. The successful affirmation of this conjecture has led to the identification of the homotopy types of the K–theory of the integers $\mathbb{Z}$ and the K–theory of the sphere spectrum $\mathbb{S}$ at regular primes; see Dwyer and Mitchell [15], Rognes [29; 30] and Rognes and Weibel [31]. Since $H\mathbb{Z}$ and $\mathbb{S}$ are E_∞ ring spectra, $K(\mathbb{Z})$ and $K(\mathbb{S})$ are E_∞ ring spectra and the graded rings $K_*(\mathbb{S}) = \pi_* K(\mathbb{S})$ and $K_*(\mathbb{Z}) = \pi_* K(\mathbb{Z})$ are commutative. However, almost nothing is known about the multiplicative structure. The only work in this direction so far is the investigation of Bergsaker and Rognes [4] of the Dyer–Lashof operations on $\mathrm{TC}_*(\mathbb{S})$ at the prime 2. In this paper, we begin the study of the multiplicative structure on the homotopy groups of $K(\mathbb{S})$ by proving the analogue of Nishida's nilpotence theorem.

Theorem 1 *Positive degree elements of $K_*(\mathbb{S})$ are nilpotent.*

On the way to proving the preceding theorem, we show the corresponding nilpotence result for $K_*(\mathbb{Z})$. We deduce this by observing that $K_{2n(p-1)}(\mathbb{Z}) \otimes \mathbb{Z}_{(p)} = 0$ for odd primes p and $n > 0$; it can also be deduced from the multiplicative properties of the Quillen–Lichtenbaum spectral sequence.

Theorem 2 *Positive degree elements of $K_*(\mathbb{Z})$ are nilpotent.*

Much of the interest in $K(\mathbb{S})$ comes from its identification as $A(*)$, Waldhausen's algebraic K–theory of the one-point space. Work of Waldhausen and collaborators shows that $A(X)$ controls high-dimensional manifold theory (eg see Waldhausen,

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3453

Jahren and Rognes [37] and Weiss and Williams [40]) via the connection to the stable pseudoisotopy spectrum $\mathrm{Wh}(X)$. Rognes shows that the infinite loop space structure on $\mathrm{Wh}(*)$ that is relevant to the Hatcher–Waldhausen map $G/O \to \Omega\mathrm{Wh}(*)$, where G/O denotes the classifying spectrum for smooth normal invariants, is induced by the ring structure on $A(*)$; see Rognes [28]. Moreover, $A(X)$ is a module over $A(*)$; more generally, for any ring spectrum (or even any Waldhausen category that admits factorization; see Blumberg and Mandell [7; 8]), the algebraic K–theory spectrum is a module over $A(*)$.

Theorem 1 also has direct implications in the context of Kontsevich's noncommutative motives. The work of Blumberg, Gepner and Tabuada [5; 6] produces a candidate category of spectral motives $\mathcal{M}ot_{\mathrm{ex}}$, which is a symmetric monoidal category with objects the smooth and proper small stable idempotent-complete ∞–categories. The category of spectral motives is stable, which in particular implies that it has a tensor-triangulated homotopy category and is enriched over spectra; the mapping spectra are essentially bivariant algebraic K–theory. The endomorphism spectrum of the unit is precisely $K(\mathbb{S})$ (as an E_∞ ring spectrum).

The Devinatz–Hopkins–Smith nilpotence theorem and the Hopkins–Smith thick subcategory theorem teach us that to understand a triangulated category, we should look to its thick subcategories, which play the role of prime ideals in derived algebraic geometry; see Hopkins [21], Neeman [25] and Thomason [35]. More recently, Balmer [1; 2] proposes a systematic study of this in the setting of "tensor-triangulated geometry", defining the *triangulated spectrum* to be the space of prime proper thick triangulated tensor ideals (with the Zariski topology). Balmer observes that there is a canonical map from the triangulated spectrum to the spectrum of the graded ring of endomorphisms of the unit and that in many known examples, the spectrum of the endomorphism ring controls the triangulated spectrum of the tensor-triangulated category. Our main theorem is the first step in realizing this program for spectral motives.

In a different direction, Morava has developed a conjectural program for studying a homotopy-theoretic analogue of Kontsevich's Grothendieck–Teichmüller group — see Kitchloo and Morava [22] and Morava [24] — in terms of homotopical descent for the category of spectral motives. These ideas revolve around understanding the structure of $\mathbb{S}\wedge^{\mathbf{L}}_{K(\mathbb{S})}\mathbb{S}$, which of course depends on the ring structure of $K(\mathbb{S})$. Morava notes that the calculation of this object is straightforward rationally and results in a concise description as a polynomial algebra on even degree generators: it is the polynomial algebra on the free Lie coalgebra $L\langle x_6, x_{10}, x_{14}, \dots\rangle$ on generators in degrees 6, 10, 14, etc. (It is a Hopf algebra with coalgebra the tensor coalgebra on $\pi_*\Sigma\mathrm{Wh}(*)_{\mathbb{Q}} \cong \pi_*\Sigma^6 ko_{\mathbb{Q}}$, where $\mathrm{Wh}(*)$ is the fiber of the map $K(S)\to S$.) Our results give the first progress in the direction of the torsion part of this theory.

Acknowledgments The authors thank Bill Dwyer, Mike Hopkins, Michael Larsen, Tyler Lawson, Barry Mazur, Jack Morava and Justin Noel for helpful conversations, and the IMA and MSRI for their hospitality while some of this work was done. The authors especially thank Lars Hesselholt; we learned much of what we know in this area through conversations with him and from his unpublished notes [19] on p–adic L–functions and the cyclotomic trace. Blumberg was supported in part by NSF grant DMS-1151577.

2 Reduction of Theorems 1 and 2

Consider the arithmetic square

$$\begin{array}{ccc} K(\mathbb{S}) & \longrightarrow & \prod_p K(\mathbb{S})^\wedge_p \\ \downarrow & & \downarrow \\ K(\mathbb{S})_\mathbb{Q} & \longrightarrow & \big(\prod_p K(\mathbb{S})^\wedge_p\big)_\mathbb{Q} \end{array}$$

where $(-)^\wedge_p$ denotes p–completion (localization with respect to the mod p Moore spectrum) and $(-)_\mathbb{Q}$ denotes rationalization. To prove Theorem 1, it suffices to prove the analogous nilpotence results for $K(\mathbb{S})_\mathbb{Q}$ and $K(\mathbb{S})^\wedge_p$ for each prime p; this is easy to see for $K(\mathbb{S})$ because $\pi_* K(\mathbb{S})$ is finitely generated in each degree [14, 1.2], which implies that $\pi_*(K(\mathbb{S})^\wedge_p) \cong (\pi_* K(\mathbb{S})) \otimes \mathbb{Z}^\wedge_p$; see [12, 2.5]. (Similar observations apply to $K(\mathbb{Z})$ for Theorem 2; see [27].). The rational part is well understood: the natural map $K(\mathbb{S})_\mathbb{Q} \to K(\mathbb{Z})_\mathbb{Q}$ is an equivalence [36, 2.3.8], and classical results of Borel [11, 12.2] imply that the positive degree elements of $\pi_* K(\mathbb{Z})_\mathbb{Q}$ are concentrated in odd degrees and therefore square to zero. It remains to study the situation after p–completion.

Our strategy for studying the multiplicative structure on $K(\mathbb{S})^\wedge_p$ uses the cyclotomic trace map, which is a map of E_∞ ring spectra from $K(\mathbb{S})$ to the topological cyclic homology $\mathrm{TC}(\mathbb{S})$. The homotopy type of $\mathrm{TC}(\mathbb{S})^\wedge_p$ (as a spectrum) is known by work of [9].

Theorem 2.1 [9, 5.16] *There is an equivalence of p–complete spectra*

$$\mathrm{TC}(\mathbb{S})^\wedge_p \simeq \mathbb{S}^\wedge_p \vee \mathrm{hofib}(\Sigma(\Sigma^\infty_+ \mathbb{C}P^\infty) \to \mathbb{S})^\wedge_p \simeq \mathbb{S}^\wedge_p \vee (\mathbb{C}P^\infty_{-1})^\wedge_p.$$

The Devinatz–Hopkins–Smith nilpotence theorems provide a criterion for determining when elements in the homotopy groups of a ring spectrum R are multiplicatively nilpotent. Specifically, an element $x \in \pi_* R$ is nilpotent if and only if the Hurewicz map takes it to a nilpotent element of $K(n)_* R$ for all $0 \le n \le \infty$ (and all primes p).

Although the previous theorem only identifies the homotopy type of the underlying spectrum and says nothing about the multiplication, it is enough to deduce a nilpotence result for $\mathrm{TC}(\mathbb{S})^{\wedge}_p$.

Proposition 2.2 *Let p be a prime, let $0 \le n \le \infty$, and let $\widetilde{\mathrm{TC}}(\mathbb{S}; p)$ be the homotopy fiber of the augmentation map $\mathrm{TC}(\mathbb{S})^{\wedge}_p \to \mathbb{S}^{\wedge}_p$ (obtained from the canonical map $\mathrm{TC}(\mathbb{S})^{\wedge}_p \to \mathrm{THH}(\mathbb{S})^{\wedge}_p \simeq \mathbb{S}^{\wedge}_p$). Then $K(n)_*(\widetilde{\mathrm{TC}}(\mathbb{S}; p))$ is concentrated in odd degrees.*

Proof As a consequence of Theorem 2.1, $\widetilde{\mathrm{TC}}(\mathbb{S}; p) \simeq \Sigma(\mathbb{C}P^{\infty}_{-1})^{\wedge}_p$. The spectrum $\mathbb{C}P^{\infty}_{-1}$ is the Thom spectrum of the virtual bundle $-\gamma$, for γ the tautological line bundle over $\mathbb{C}P^{\infty}$. The spectra $K(n)$ are all complex oriented; the proposition now follows from the Thom isomorphism. □

Since $\pi_*(\mathrm{TC}(\mathbb{S})^{\wedge}_p)$ splits as $\pi_*\mathbb{S}^{\wedge}_p \oplus \pi_*\widetilde{\mathrm{TC}}(\mathbb{S}; p)$, with the first factor the image of the inclusion of the unit, we obtain the following as an immediate corollary of the previous proposition and the nilpotence theorem.

Theorem 2.3 *For any prime p, all the nonzero degree elements of $\pi_*\mathrm{TC}(\mathbb{S})^{\wedge}_p$ are nilpotent.*

In light of the previous result, Theorem 1 becomes an immediate consequence of the following lemma. We prove this lemma for odd p in later sections; for $p = 2$ it is a special case of [29, 3.16].

Lemma 1 *For $p = 2$, let $d = 8$, and for p odd, let $d = 2(p-1)$. The homotopy fiber of the cyclotomic trace map $trc_p\colon K(\mathbb{S})^{\wedge}_p \to \mathrm{TC}(\mathbb{S})^{\wedge}_p$ has trivial homotopy groups in degrees kd for $k > 0$.*

Proof of Theorem 1 from Lemma 1 Given $x \in \pi_k K(\mathbb{S})^{\wedge}_p$, $x^d \in \pi_{kd} K(\mathbb{S})^{\wedge}_p$. When $k > 0$, we then know that for some power n, $(x^d)^n$ maps to zero in $\pi_{kdn}(\mathrm{TC}(\mathbb{S})^{\wedge}_p)$ under the trace map by Theorem 2.3. By Lemma 1, the kernel of the trace is zero in degree kdn, and so $x^{kdn} = 0$. □

As we used in the proof, Lemma 1 implies that the cyclotomic trace $K(\mathbb{S}) \to \mathrm{TC}(\mathbb{S})$ is injective in certain degrees. In fact, for odd regular primes, the cyclotomic trace is injective in all degrees. This follows from the work of Rognes on $\mathrm{Wh}(*)$ at odd regular primes, specifically [30, 3.6 and 3.8]. In the case of irregular primes, we expect that the trace fails to be injective; we hope to return to this question in a future paper.

On the way to proving Lemma 1, we also prove the following lemma. It is well known that $\pi_{4k} K(\mathbb{Z}) \otimes \mathbb{Z}_{(p)} = 0$ at regular primes, including $p = 2$ (see [39, 10.1], for example), and this combined with the following lemma now proves Theorem 2.

Lemma 2 *For p an odd prime, $\pi_{2(p-1)k} K(\mathbb{Z}) \otimes \mathbb{Z}_{(p)} = 0$ for $k > 0$.*

3 Reduction of Lemmas 1 and 2

The basic strategy for the proof of Lemmas 1 and 2 is to reduce the study of the homotopy fiber of the cyclotomic trace $K(\mathbb{S})^\wedge_p \to \mathrm{TC}(\mathbb{S})^\wedge_p$ to the study of the p–completion map $\mathbb{Z}[1/p] \to \mathbb{Q}^\wedge_p$ in étale cohomology. (This is now a fairly standard approach; for instance, see [30, Sections 2–3; 17; 19].) As indicated above, from here on we assume that p is odd (though all of what we say would also apply in the case $p=2$ until (3.6)). First, we apply Dundas' theorem [13] about the cyclotomic trace: the square

$$\begin{tikzcd} K(\mathbb{S})^\wedge_p \arrow[r] \arrow[d, "trc_p"'] & K(\mathbb{Z})^\wedge_p \arrow[d, "trc^{\mathbb{Z}}_p"] \\ \mathrm{TC}(\mathbb{S})^\wedge_p \arrow[r] & \mathrm{TC}(\mathbb{Z})^\wedge_p \end{tikzcd}$$

is homotopy cocartesian, where the horizontal maps arise from linearization. As a consequence, we have the following lemma:

Proposition 3.1 (Dundas [13]) *The induced map* $\mathrm{hofib}(trc_p) \to \mathrm{hofib}(trc^{\mathbb{Z}}_p)$ *is an equivalence.*

To understand $\mathrm{hofib}(trc^{\mathbb{Z}}_p)$, consider the commutative diagram

$$\begin{tikzcd} K(\mathbb{Z})^\wedge_p \arrow[r, "cmp"] \arrow[d, "trc^{\mathbb{Z}}_p"'] & K(\mathbb{Z}^\wedge_p)^\wedge_p \arrow[d, "trc^{\mathbb{Z}^\wedge_p}_p"] \\ \mathrm{TC}(\mathbb{Z})^\wedge_p \arrow[r, "cmp^{\mathrm{TC}}"'] & \mathrm{TC}(\mathbb{Z}^\wedge_p)^\wedge_p \end{tikzcd}$$

where the horizontal maps cmp and cmp^{TC} are induced by the map of rings $\mathbb{Z} \to \mathbb{Z}^\wedge_p$. By work of Hesselholt and Madsen [20], the bottom map is a weak equivalence [20, Addendum 6.2] and the right-hand map induces a weak equivalence [20, Theorem D]

$$K(\mathbb{Z}^\wedge_p)^\wedge_p \to \mathrm{TC}(\mathbb{Z}^\wedge_p)^\wedge_p[0,\infty) \tag{3.2}$$

(where $[0,\infty)$ denotes the connective cover). Thus, up to passing to a connective cover, we can identify the trace map $trc^{\mathbb{Z}}_p$ as the map $cmp\colon K(\mathbb{Z})^\wedge_p \to K(\mathbb{Z}^\wedge_p)^\wedge_p$. We then have the following relationship between $\mathrm{hofib}(trc_p) \simeq \mathrm{hofib}(trc^{\mathbb{Z}}_p)$ and $\mathrm{hofib}(cmp)$.

Proposition 3.3 *There is a cofiber sequence*

$$\mathrm{hofib}(cmp) \to \mathrm{hofib}(trc_p) \to \Sigma^{-2} H\mathbb{Z}^\wedge_p \to \Sigma \cdots$$

Proof Using the equivalence of $\mathrm{hofib}(trc_p)$ and $\mathrm{hofib}(trc_p^{\mathbb{Z}})$ above, we get a diagram of cofiber sequences

$$\begin{tikzcd}
\mathrm{hofib}(cmp) \arrow[r] \arrow[d] & K(\mathbb{Z})^\wedge_p \arrow[r, "cmp"] \arrow[d, equal] & K(\mathbb{Z}^\wedge_p)^\wedge_p \arrow[r] \arrow[d, "trc_p^{\mathbb{Z}^\wedge_p}"'] & \Sigma\,\mathrm{hofib}(cmp) \arrow[d] \\
\mathrm{hofib}(trc_p) \arrow[r] & K(\mathbb{Z})^\wedge_p \arrow[r] & \mathrm{TC}(\mathbb{Z}^\wedge_p)^\wedge_p \arrow[r] & \Sigma\,\mathrm{hofib}(trc_p)
\end{tikzcd}$$

identifying the right-hand square as homotopy (co)cartesian. Since $\pi_{-1}\mathrm{TC}(\mathbb{Z})^\wedge_p = \mathbb{Z}^\wedge_p$ and $\pi_n\mathrm{TC}(\mathbb{Z})^\wedge_p = 0$ for $n < -1$, the homotopy cofiber of the map $trc_p^{\mathbb{Z}^\wedge_p}$ in the diagram is $\Sigma^{-1}H\mathbb{Z}^\wedge_p$. Desuspending, we see that the homotopy cofiber of

$$\mathrm{hofib}(cmp) \to \mathrm{hofib}(trc_p)$$

is $\Sigma^{-2}H\mathbb{Z}^\wedge_p$. □

For Lemma 1 then, $\mathrm{hofib}(cmp)$ works just as well as $\mathrm{hofib}(trc_p)$. Quillen's localization sequence [26] gives cofiber sequences

$$\begin{tikzcd}
K(\mathbb{Z}/p) \arrow[r] \arrow[d, "\mathrm{id}"'] & K(\mathbb{Z}) \arrow[r] \arrow[d] & K(\mathbb{Z}[1/p]) \arrow[r] \arrow[d] & \Sigma\cdots \\
K(\mathbb{Z}/p) \arrow[r] & K(\mathbb{Z}^\wedge_p) \arrow[r] & K(\mathbb{Q}^\wedge_p) \arrow[r] & \Sigma\cdots
\end{tikzcd} \tag{3.4}$$

from which we can see that $\mathrm{hofib}(cmp)$ is equivalent to the homotopy fiber of the map

$$cmp'\colon K(\mathbb{Z}[1/p])^\wedge_p \to K(\mathbb{Q}^\wedge_p)^\wedge_p.$$

Proposition 3.5 *There is a homotopy equivalence* $\mathrm{hofib}(cmp) \to \mathrm{hofib}(cmp')$.

The advantage of this approach is that étale cohomology methods at the prime p can be applied in rings where p is a unit. Let R denote either $\mathbb{Z}[1/p]$ or $\mathbb{Q}^\wedge_p$; then R satisfies the "mild extra hypotheses" of Thomason [34, 0.1], which gives a spectral sequence

$$E_2^{s,t} = H^s_{\text{ét}}\big(\mathrm{Spec}\,R; \mathbb{Z}/p^n\big(\tfrac{1}{2}t\big)\big) \implies \pi_{t-s}(K_{\text{ét}}(R); \mathbb{Z}/p^n) \tag{3.6}$$

from étale cohomology to the mod p^n homotopy groups of (Dwyer–Friedlander) étale K–theory. In the formula above

$$\mathbb{Z}/p^n\big(\tfrac{1}{2}t\big) = \begin{cases} \mu_{p^n}^{\otimes(t/2)} & \text{if } t \text{ is even,} \\ 0 & \text{if } t \text{ is odd,} \end{cases}$$

where μ_{p^n} denotes the $(p^n)^{\text{th}}$ roots of 1 (ie $\mu_{p^n}(A) = \{x \in A \mid x^{p^n} = 1\}$, a sheaf in the étale topology). In this case the affirmed Quillen–Lichtenbaum conjecture [39, VI.8.2] identifies

$$\pi_*(K(R); \mathbb{Z}/p^n) = \pi_*(K_{\text{ét}}(R); \mathbb{Z}/p^n)$$

for $* \geq 2$. Also, because we have assumed that p is odd, $H^*_{\text{ét}}(R; \mathbb{Z}/p^n(k)) = 0$ for $* > 2$ [32, Section III.1.3], and the spectral sequence collapses to give an isomorphism and a short exact sequence

$$\pi_{2k-1}(K(R); \mathbb{Z}/p^n) \cong H^1_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}/p^n(k)),$$

$$(3.7)\quad 0 \to H^2_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}/p^n(k+1)) \to \pi_{2k}(K(R); \mathbb{Z}/p^n) \to H^0_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}/p^n(k)) \to 0$$

for $k > 1$. In fact, the calculation of the $H^0_{\text{ét}}$ term is well known:

Proposition 3.8 *Let $R = \mathbb{Z}[1/p]$ or $\mathbb{Q}^\wedge_p$. Then $H^0_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}/p^n(k)) = 0$ unless $(p-1) \mid k$. If $k = m(p-1)$, then $H^0_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}/p^n(k)) \cong \mu^{\otimes k}_{p^i}(\bar{Q})$, where $p^i = \gcd(|m|p, p^n)$ (and $i = n$ if $m = 0$) and $\bar{Q}$ is the algebraic closure of the field of fractions of R.*

Proof The inclusion of the generic point $\operatorname{Spec} \mathbb{Q} \to \operatorname{Spec} \mathbb{Z}[1/p]$ induces an isomorphism

$$H^0_{\text{ét}}(\operatorname{Spec} \mathbb{Z}[1/p], \mathbb{Z}/p^n(k)) \to H^0_{\text{ét}}(\operatorname{Spec} \mathbb{Q}, \mathbb{Z}/p^n(k));$$

see [32, Proposition 1]. This reduces to the case $Q = \mathbb{Q}$ or $\mathbb{Q}^\wedge_p$ and the étale cohomology $H^0_{\text{ét}}(\operatorname{Spec} Q; \mathbb{Z}/p^n(k))$ becomes the Galois cohomology $H^0_{\text{Gal}}(Q; \mu^{\otimes k}_{p^n}(\bar{Q}))$. (We will now fix $\bar{Q}$ and write μ_{p^n} for $\mu_{p^n}(\bar{Q})$.) Letting $G = \operatorname{Gal}(Q(\mu_{p^n})/Q)$, the action of $\operatorname{Gal}(\bar{Q}/Q)$ on $\mu^{\otimes k}_{p^n}$ factors through G, and we can identify $H^0_{\text{Gal}}(Q; \mu^{\otimes k}_{p^n})$ as the G–fixed point subgroup of $\mu^{\otimes k}_{p^n}$. We have a canonical isomorphism $G = (\mathbb{Z}/p^n)^\times$ given by letting $r \in (\mathbb{Z}/p^n)^\times$ act on $\alpha \in \mu_{p^n}$ by $\alpha \mapsto \alpha^r$; then r acts on $\mu^{\otimes k}_{p^n}$ by the r^k power map (ie multiplication by r^k when we write the group operation additively). Choosing r to be a generator of $(\mathbb{Z}/p^n)^\times$, the G–fixed point subgroup of $\mu^{\otimes k}_{p^n}$ is the subset where r acts by the identity, or equivalently, the subset $\alpha \in \mu^{\otimes k}_{p^n}$ such that $\alpha^{r^k-1} = 1$. If $p-1$ does not divide k, then $r^k - 1$ is not congruent to 0 mod p, and the only fixed point is the identity. On the other hand, $r^{m(p-1)} - 1$ is divisible by p^i (and for $i < n$ not p^{i+1}) where $p^i = \gcd(|m|p, p^n)$ (for $m \neq 0$ or $i = n$ if $m = 0$), and the G–fixed point subgroup is exactly the subgroup $\mu^{\otimes k}_{p^i}$. □

Defining $H^*_{\text{ét}}(-; \mathbb{Z}^\wedge_p(k))$ as the inverse limit of $H^*_{\text{ét}}(-; \mathbb{Z}/p^n(k))$, we see from the preceding proposition that for $R = \mathbb{Z}[1/p]$ or $\mathbb{Q}^\wedge_p$, we have $H^0_{\text{ét}}(R; \mathbb{Z}^\wedge_p(k)) = 0$ for

$k \neq 0$. Since in these cases the homotopy groups of $K(R)$ are finitely generated $\mathbb{Z}^{\wedge}_p$–modules in each degree (see [14, Section 4; 20, Theorem D; 10, 0.7]), $\pi_* K(R)^{\wedge}_p \cong \lim \pi_*(K(R); \mathbb{Z}/p^n)$. Combining these observations and the left exactness of lim, we then get isomorphisms

$$\begin{aligned} \pi_{2k-1}(K(R)^{\wedge}_p) &\cong H^1_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}^{\wedge}_p(k)), \\ \pi_{2k}(K(R)^{\wedge}_p) &\cong H^2_{\text{ét}}(\operatorname{Spec} R; \mathbb{Z}^{\wedge}_p(k+1)), \end{aligned} \tag{3.9}$$

for $k > 1$. Combining these isomorphisms with the fact that

$$\pi_{2m(p-1)} \operatorname{hofib}(trc_p) \cong \pi_{2m(p-1)} \operatorname{hofib}(cmp) \cong \pi_{2m(p-1)} \operatorname{hofib}(cmp')$$

and $\pi_{2m(p-1)} \operatorname{hofib}(cmp')$ fits in an exact sequence

$$\pi_{2m(p-1)+1} K(\mathbb{Z}[1/p])^{\wedge}_p \to \pi_{2m(p-1)+1} K(\mathbb{Q}^{\wedge}_p)^{\wedge}_p \to \pi_{2m(p-1)} \operatorname{hofib}(cmp') \longrightarrow \pi_{2m(p-1)} K(\mathbb{Z}[1/p])^{\wedge}_p,$$

Lemma 1 is now an immediate consequence of the following pair of lemmas, proved in the next section.

Lemma 3 *Let p be an odd prime. The map of rings $\mathbb{Z}[1/p] \to \mathbb{Q}^{\wedge}_p$ induces a surjection*

$$H^1_{\text{ét}}\big(\operatorname{Spec} \mathbb{Z}[1/p]; \mathbb{Z}^{\wedge}_p(m(p-1)+1)\big) \to H^1_{\text{ét}}\big(\operatorname{Spec} \mathbb{Q}^{\wedge}_p; \mathbb{Z}^{\wedge}_p(m(p-1)+1)\big)$$

for all $m > 0$.

Lemma 4 *Let p be an odd prime.* $H^2_{\text{ét}}(\operatorname{Spec} \mathbb{Z}[1/p]; \mathbb{Z}^{\wedge}_p(m(p-1)+1)) = 0$ *for all $m > 0$.*

We can also deduce Lemma 2: Quillen's computation of the K–theory of finite fields implies in particular that $K(\mathbb{Z}/p)^{\wedge}_p \simeq H\mathbb{Z}^{\wedge}_p$. It then follows from Quillen's localization sequence (3.4) that the map $K(\mathbb{Z})^{\wedge}_p \to K(\mathbb{Z}[1/p])^{\wedge}_p$ induces an isomorphism in homotopy groups above degree 1. Lemma 2 now follows from the isomorphisms (3.9) and Lemma 4.

4 Proof of Lemmas 3 and 4

In this section, we prove Lemmas 3 and 4. Lemma 3 is about the p–completion map in étale cohomology and the basic tool for studying this is the Tate–Poitou duality long exact sequence [33]. (Again, for examples applied to K–theory, see [30, 3.1; 17, Section 4; 19].)

In our context, the Tate–Poitou sequence takes the following form. Let M be a finite abelian p–group with an action of the Galois group G of the maximal extension of $\mathbb{Q}$ unramified except at p (eg $M = \mathbb{Z}/p^n(k)$) and let $(-)^*$ denote the Pontryagin dual, $A^* = \operatorname{Hom}(A, \mathbb{Q}/\mathbb{Z})$; then $M^*(1)$ is the G–module $\operatorname{Hom}(M, \mu_\infty)$, where μ_∞ denotes the G–module of all roots of 1 in the algebraic closure of $\mathbb{Q}$. The low-dimensional part of Tate–Poitou duality in the case at hand is then summarized by the following long exact sequence [33, 3.1]:

$$
\begin{aligned}
&0 \longrightarrow H^0_{\text{ét}}(\mathbb{Z}[1/p]; M) \longrightarrow H^0_{\text{ét}}(\mathbb{Q}_p^\wedge; M) \longrightarrow \big(H^2_{\text{ét}}(\mathbb{Z}[1/p], M^*(1))\big)^* \\
&\longrightarrow H^1_{\text{ét}}(\mathbb{Z}[1/p]; M) \longrightarrow H^1_{\text{ét}}(\mathbb{Q}_p^\wedge; M) \longrightarrow \big(H^1_{\text{ét}}(\mathbb{Z}[1/p], M^*(1))\big)^* \\
&\longrightarrow H^2_{\text{ét}}(\mathbb{Z}[1/p]; M) \longrightarrow H^2_{\text{ét}}(\mathbb{Q}_p^\wedge; M) \longrightarrow \big(H^0_{\text{ét}}(\mathbb{Z}[1/p], M^*(1))\big)^* \longrightarrow 0
\end{aligned}
\tag{4.1}
$$

When $M = \mathbb{Z}/p^n(k)$, the first map in the sequence above,

$$H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^n(k)) \to H^0_{\text{ét}}(\mathbb{Q}_p^\wedge; \mathbb{Z}/p^n(k)),$$

is an isomorphism by Proposition 3.8. Likewise, when $M = \mathbb{Z}/p^n(k)$ for $k > 1$, we see from (3.7) that $H^i_{\text{ét}}(R; \mathbb{Z}/p^n(k))$ is finite for $R = \mathbb{Z}[1/p]$ or $\mathbb{Q}_p^\wedge$, and it follows that the above is an exact sequence of finite groups. Taking the inverse limit over n is then exact and we get the following Tate–Poitou sequence:

$$
\begin{aligned}
&0 \to \big(H^2_{\text{ét}}(\mathbb{Z}[1/p], \mathbb{Z}/p^\infty(1-k))\big)^* \\
&\to H^1_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}_p^\wedge(k)) \to H^1_{\text{ét}}(\mathbb{Q}_p^\wedge; \mathbb{Z}_p^\wedge(k)) \to \big(H^1_{\text{ét}}(\mathbb{Z}[1/p], \mathbb{Z}/p^\infty(1-k))\big)^* \\
&\to H^2_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}_p^\wedge(k)) \to H^2_{\text{ét}}(\mathbb{Q}_p^\wedge; \mathbb{Z}_p^\wedge(k)) \to \big(H^0_{\text{ét}}(\mathbb{Z}[1/p], \mathbb{Z}/p^\infty(1-k))\big)^* \to 0
\end{aligned}
\tag{4.2}
$$

For Lemmas 3 and 4, we apply (4.2) with $k = m(p-1)+1$, combined with the main theorem of Bayer and Neukirch [3], which relates the values of the Iwasawa p–adic ζ–function with the size of étale cohomology groups. In the following theorem, $|\cdot|_p$ denotes the p–adic valuation on $\mathbb{Q}_p^\wedge$, normalized so that $|p^n u|_p = p^{-n}$, where u is a unit in $\mathbb{Z}_p^\wedge$.

Theorem 4.3 (Bayer and Neukirch [3, 6.1]) *Let $\zeta_I(\omega^0, s)$ denote the Iwasawa zeta function of [3, 5.1] associated to the trivial character ω^0 and the field $\mathbb{Q}$. Let $k =$*

$m(p-1)+1$ for $m \neq 0$. If $\zeta_I(\omega^0,k) \neq 0$ then the groups $H^*_{\text{ét}}(\mathbb{Z}[1/p];\mathbb{Z}/p^\infty(1-k))$ are all finite (zero for $* \geq 2$) and

$$|\zeta_I(\omega^0,k)|_p = \frac{\#\big(H^0_{\text{ét}}(\mathbb{Z}[1/p];\mathbb{Z}/p^\infty(1-k))\big)}{\#\big(H^1_{\text{ét}}(\mathbb{Z}[1/p];\mathbb{Z}/p^\infty(1-k))\big)}.$$

The following computation of $|\zeta_I(\omega^0, m(p-1)+1)|_p$ is well known.

Proposition 4.4 *For $m \neq 0$ and $k = m(p-1)+1$,*

$$|\zeta_I(\omega^0,k)|_p = \left|\frac{1}{mp}\right|_p.$$

Proof The Iwasawa zeta function used by Bayer and Neukirch [3, 5.1] depends on a choice of $q \in \mathbb{Z}^\wedge_p$ with $q \equiv 1 \bmod p$. For the trivial character, the formula is then

$$\zeta_I(\omega^0,s) = \frac{p^{\mu_0} g_0(q^{1-s}-1)}{1-q^{1-s}},$$

where, for $\mathbb{Q}$ (and any abelian extension thereof), $\mu_0 = 0$ as a case of the Iwasawa "$\mu = 0$" conjecture proved by Ferrero and Washington [16] (see [3, 5.3]) and $g_0(x)$ is the characteristic polynomial of the action of $T \in \mathbb{Z}^\wedge_p[\![T]\!] \cong \Lambda$ on a Λ–module denoted as $e_0\mathcal{M}$ in [3]. (Here Λ is the Iwasawa algebra [38, 7.1] for $\mathbb{Z}^\wedge_p \cong \Gamma < \mathrm{Gal}(\mathbb{Q}(\mu_{p^\infty})/\mathbb{Q})$ with topological generator $\gamma \leftrightarrow 1+T$ acting by $\alpha \mapsto \alpha^q$ for $x \in \mu_{p^\infty}$.) Since, for $k = m(p-1)+1$ with $m \neq 0$,

$$|1-q^{1-k}|_p = |1-q^{-m(p-1)}|_p = |1-q^{|m|(p-1)}|_p = \left|\frac{1}{m(p-1)p}\right|_p = \left|\frac{1}{mp}\right|_p,$$

it suffices to show that $g_0(x) = 1$. This is a special case of the main conjecture of Iwasawa theory [23, Section 6, Conjecture] for the trivial character. Though the exposition preceding [23, Section 9, Theorem] makes the statement appear ambiguous in the case of the trivial character, this case was known at least as far back as [18], as we now discuss for the benefit of those (like the authors) not expert in this theory.

Washington [38, 15.37] denotes $e_0\mathcal{M}$ as $\epsilon_0\mathcal{X}$ and $\epsilon_0\mathcal{X}_\infty$ and shows that

$$g_0(q(1+T)^{-1}-1) = f(T)u(T)$$

in $\mathbb{Z}^\wedge_p[\![T]\!]$ for $u(T)$ a unit power series and $f(x)$ the characteristic polynomial of $\epsilon_1 X$, where X is the inverse limit of X_n and $X_n \cong A_n$ is the p–Sylow subgroup of the class group of $\mathbb{Q}(\mu_{p^n})$. Greenberg [18] denotes X as X_K, $\epsilon_1 X$ as $X_K^{[1]}$, and defines $V^{[1]} = \epsilon_1 X \otimes_{\mathbb{Z}^\wedge_p} \Omega_p$ where $\Omega_p = \overline{\mathbb{Q}}_p$ is the algebraic closure of $\mathbb{Q}_p$. The characteristic polynomial of $\epsilon_1 X$ and $V^{[1]}$ are therefore equal, and Greenberg [18, Corollary 1] shows that $V^{[1]} = 0$. Thus, $f(x) = 1$ and we conclude that $g_0(x) = 1$. (In fact, $\epsilon_1 X = 0$ and $\epsilon_1 X_n = 0$ for all n as can be seen from [38, 6.16, 13.22] and Nakayama's lemma.) □

We can also compute $H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k))$ using Proposition 3.8, and for $k = m(p-1)+1$ we get

$$\begin{aligned} H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k)) &= H^0_{\text{ét}}\big(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(-m(p-1))\big) \\ &= \mu_{p^i}^{\otimes(1-k)}(\overline{\mathbb{Q}}) \cong \mathbb{Z}^\wedge_p/(mp), \end{aligned}$$

where $mp = p^i r$ for r relatively prime to p, or more concisely, $p^i = |1/(mp)|_p$. The following proposition is now immediate.

Proposition 4.5 $H^1_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k)) = 0$ *for* $m \neq 0$ *and* $k = m(p-1)+1$.

Combining the previous proposition with the Tate–Poitou sequence (4.2), the proof of Lemma 3 is now clear. For Lemma 4, we need the following K–theory computation of Bökstedt and Madsen [10] and Hesselholt and Madsen [20].

Theorem 4.6 (Hesselholt and Madsen [20, Theorem D], Bökstedt and Madsen [10, 0.7]) *For* $m > 0$,

$$\pi_{2m(p-1)}(K(\mathbb{Q}^\wedge_p)^\wedge_p) \cong \mathbb{Z}^\wedge_p/(mp).$$

Proof of Lemma 4 Let $k = m(p-1)+1$. By the previous theorem and (3.9), we have

$$\#\big(H^2_{\text{ét}}(\mathbb{Q}^\wedge_p; \mathbb{Z}^\wedge_p(k))\big) = \left|\frac{1}{mp}\right|_p$$

and by Proposition 3.8, we have

$$\#\Big(\big(H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k))\big)^*\Big) = \#\big(H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k))\big) = \left|\frac{1}{mp}\right|_p.$$

Because the map

$$H^2_{\text{ét}}(\mathbb{Q}^\wedge_p; \mathbb{Z}^\wedge_p(k)) \to \big(H^0_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k))\big)^*$$

in the Tate–Poitou sequence (4.2) is surjective and the groups are the same finite cardinality, it must therefore also be injective. The map

$$\big(H^1_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}/p^\infty(1-k))\big)^* \to H^2_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}^\wedge_p(k))$$

is therefore surjective, and Proposition 4.5 then shows that $H^2_{\text{ét}}(\mathbb{Z}[1/p]; \mathbb{Z}^\wedge_p(k))$ is zero. □

References

[1] **P Balmer**, *The spectrum of prime ideals in tensor triangulated categories*, J. Reine Angew. Math. 588 (2005) 149–168 MR

[2] **P Balmer**, *Spectra, spectra, spectra—tensor triangular spectra versus Zariski spectra of endomorphism rings*, Algebr. Geom. Topol. 10 (2010) 1521–1563 MR

[3] **P Báyer**, **J Neukirch**, *On values of zeta functions and l–adic Euler characteristics*, Invent. Math. 50 (1978) 35–64 MR

[4] **H S Bergsaker**, **J Rognes**, *Homology operations in the topological cyclic homology of a point*, Geom. Topol. 14 (2010) 735–772 MR

[5] **A J Blumberg**, **D Gepner**, **G Tabuada**, *A universal characterization of higher algebraic K–theory*, Geom. Topol. 17 (2013) 733–838 MR

[6] **A J Blumberg**, **D Gepner**, **G Tabuada**, *Uniqueness of the multiplicative cyclotomic trace*, Adv. Math. 260 (2014) 191–232 MR

[7] **A J Blumberg**, **M A Mandell**, *Algebraic K–theory and abstract homotopy theory*, Adv. Math. 226 (2011) 3760–3812 MR

[8] **A J Blumberg**, **M A Mandell**, *Localization for THH(ku) and the topological Hochschild and cyclic homology of Waldhausen categories*, preprint (2014) arXiv To appear in Mem. Amer. Math. Soc.

[9] **M Bökstedt**, **W C Hsiang**, **I Madsen**, *The cyclotomic trace and algebraic K–theory of spaces*, Invent. Math. 111 (1993) 465–539 MR

[10] **M Bökstedt**, **I Madsen**, *Topological cyclic homology of the integers*, from "K–theory" (C Kassel, J-L Loday, N Schappacher, editors), Astérisque 226, Soc. Math. France, Paris (1994) 57–143 MR

[11] **A Borel**, *Stable real cohomology of arithmetic groups*, Ann. Sci. École Norm. Sup. 7 (1974) 235–272 MR

[12] **A K Bousfield**, *The localization of spectra with respect to homology*, Topology 18 (1979) 257–281 MR

[13] **B I Dundas**, *Relative K–theory and topological cyclic homology*, Acta Math. 179 (1997) 223–242 MR

[14] **W G Dwyer**, *Twisted homological stability for general linear groups*, Ann. of Math. 111 (1980) 239–251 MR

[15] **W G Dwyer**, **S A Mitchell**, *On the K–theory spectrum of a ring of algebraic integers*, K-Theory 14 (1998) 201–263 MR

[16] **B Ferrero**, **L C Washington**, *The Iwasawa invariant μ_p vanishes for abelian number fields*, Ann. of Math. 109 (1979) 377–395 MR

[17] **T Geisser**, *The cyclotomic trace map and values of zeta functions*, from "Algebra and number theory" (R Tandon, editor), Hindustan Book Agency, Delhi (2005) 211–225 MR

[18] **R Greenberg**, *On p–adic L–functions and cyclotomic fields*, Nagoya Math. J. 56 (1975) 61–77 MR

[19] **L Hesselholt**, *Algebraic K–theory and the p–adic L–function*, unpublished notes

[20] **L Hesselholt**, **I Madsen**, *On the K–theory of finite algebras over Witt vectors of perfect fields*, Topology 36 (1997) 29–101 MR

[21] **M J Hopkins**, *Global methods in homotopy theory*, from "Homotopy theory" (E Rees, J D S Jones, editors), London Math. Soc. Lecture Note Ser. 117, Cambridge Univ. Press (1987) 73–96 MR

[22] **N Kitchloo**, **J Morava**, *The stable symplectic category and a conjecture of Kontsevich*, preprint (2012) arXiv

[23] **B Mazur**, **A Wiles**, *Class fields of abelian extensions of* $\mathbf{Q}$, Invent. Math. 76 (1984) 179–330 MR

[24] **J Morava**, *Homotopy-theoretically enriched categories of noncommutative motives*, Res. Math. Sci. 2 (2015) art. id. 8 MR

[25] **A Neeman**, *The chromatic tower for $D(R)$*, Topology 31 (1992) 519–532 MR

[26] **D Quillen**, *Higher algebraic K–theory, I*, from "Algebraic K–theory, I: Higher K–theories" (H Bass, editor), Lecture Notes in Math. 341, Springer (1973) 85–147 MR

[27] **D Quillen**, *Finite generation of the groups K_i of rings of algebraic integers*, from "Algebraic K–theory, I: Higher K–theories" (H Bass, editor), Lecture Notes in Math. 341, Springer (1973) 179–198 MR

[28] **J Rognes**, *The Hatcher–Waldhausen map is a spectrum map*, Math. Ann. 299 (1994) 529–549 MR

[29] **J Rognes**, *Two-primary algebraic K–theory of pointed spaces*, Topology 41 (2002) 873–926 MR

[30] **J Rognes**, *The smooth Whitehead spectrum of a point at odd regular primes*, Geom. Topol. 7 (2003) 155–184 MR

[31] **J Rognes**, **C Weibel**, *Two-primary algebraic K–theory of rings of integers in number fields*, J. Amer. Math. Soc. 13 (2000) 1–54 MR

[32] **C Soulé**, *K–théorie des anneaux d'entiers de corps de nombres et cohomologie étale*, Invent. Math. 55 (1979) 251–295 MR

[33] **J Tate**, *Duality theorems in Galois cohomology over number fields*, from "Proc. Internat. Congr. Mathematicians, I" (V Stenström, editor), Inst. Mittag-Leffler, Djursholm (1963) 288–295 MR

[34] **R W Thomason**, *Algebraic K–theory and étale cohomology*, Ann. Sci. École Norm. Sup. 18 (1985) 437–552 MR

[35] **R W Thomason**, *The classification of triangulated subcategories*, Compositio Math. 105 (1997) 1–27 MR

[36] **F Waldhausen**, *Algebraic K–theory of spaces*, from "Algebraic and geometric topology" (A Ranicki, N Levitt, F Quinn, editors), Lecture Notes in Math. 1126, Springer (1985) 318–419 MR

[37] **F Waldhausen**, **B Jahren**, **J Rognes**, *Spaces of PL manifolds and categories of simple maps*, Annals of Math. Studies 186, Princeton Univ. Press (2013) MR

[38] **L C Washington**, *Introduction to cyclotomic fields*, 2nd edition, Graduate Texts in Math. 83, Springer (1997) MR

[39] **C A Weibel**, *The K–book: an introduction to algebraic K-theory*, Graduate Studies in Math. 145, Amer. Math. Soc., Providence, RI (2013) MR

[40] **M Weiss**, **B Williams**, *Automorphisms of manifolds*, from "Surveys on surgery theory, II" (S Cappell, A Ranicki, J Rosenberg, editors), Ann. of Math. Stud. 149, Princeton Univ. Press (2001) 165–220 MR

Department of Mathematics, University of Texas
Austin, TX, United States

Department of Mathematics, Indiana University
Bloomington, IN, United States

blumberg@math.utexas.edu, mmandell@indiana.edu

Proposed: Mark Behrens — Received: 2 November 2015
Seconded: Peter Teichner, Ralph Cohen — Revised: 11 January 2017

msp

Geometry & Topology 21 (2017) 3467–3537

Quasi-isometric classification of right-angled Artin groups I: The finite out case

JINGYIN HUANG

Let G and G' be two right-angled Artin groups. We show they are quasi-isometric if and only if they are isomorphic, under the assumption that the outer automorphism groups $\mathrm{Out}(G)$ and $\mathrm{Out}(G')$ are finite. If we only assume $\mathrm{Out}(G)$ is finite, then G' is quasi-isometric to G if and only if G' is isomorphic to a subgroup of finite index in G. In this case, we give an algorithm to determine whether G and G' are quasi-isometric by looking at their defining graphs.

20F65, 20F67, 20F69

1 Introduction

1.1 Backgrounds and summary of results

Given a finite simplicial graph Γ with vertex set $\{v_i\}_{i\in I}$, the right-angled Artin group (RAAG) with defining graph Γ, denoted by $G(\Gamma)$, is given by the following presentation:

$$\{v_i \text{ for } i \in I \mid [v_i, v_j] = 1 \text{ if } v_i \text{ and } v_j \text{ are joined by an edge}\}.$$

We call $\{v_i\}_{i\in I}$ a *standard generating set* for $G(\Gamma)$; see Section 2.4.

The class of RAAGs enjoys a balance between simplicity and complexity. On one hand, RAAGs have many nice geometric, combinatorial and group theoretic properties (see Charney [16] for a summary); on the other hand, this class inherits the full complexity of the collection of finite simplicial graphs, and even a single RAAG could have very complicated subgroups (see, for example, Bestvina and Brady [8]).

In recent years, RAAGs have become important models to understand other unknown groups, either by (virtually) embedding the unknown groups into some RAAGs (such a program is outlined in Wise [61, Section 6]; see also Agol [2], Hagen and Wise [31; 32], Haglund and Wise [36], Ollivier and Wise [52], Przytycki and Wise [54; 55] and Wise [60; 62]), or by finding embedded copies of RAAGs in the unknown groups (see Baik, Kim and Koberda [3], Clay, Leininger and Mangahas [19], Kim and Koberda [44], Koberda [48] and Taylor [59]).

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3467

In this paper, we study the asymptotic geometry of RAAGs and classify a particular class of RAAGs by their quasi-isometric types. Previously, the quasi-isometric classification of RAAGs has been done for the following two classes:

(1) **Tree groups** It is shown by Behrstock and Neumann [7] that for any two trees Γ_1 and Γ_2 with diameter ≥ 3, the RAAGs $G(\Gamma_1)$ and $G(\Gamma_2)$ are quasi-isometric. Higher-dimensional analogs of tree groups are studied by Behrstock, Januszkiewicz and Neumann [5].

(2) **Atomic groups** A RAAG is atomic if its defining graph Γ is connected and does not contain valence-one vertices, cycles of length < 5 and separating closed stars. It is shown by Bestvina, Kleiner and Sageev [9] that two atomic RAAGs are quasi-isometric if and only if they are isomorphic.

Note that atomic groups are much more "rigid" than tree groups. We define the *dimension* of $G(\Gamma)$ to be the maximal n such that $G(\Gamma)$ contains a $\mathbb{Z}^n$ subgroup, and it coincides with the cohomological dimension of $G(\Gamma)$. All atomic groups are 2–dimensional; hence it is natural to ask what higher-dimensional RAAGs satisfy similar rigidity properties as atomic RAAGs. This is the starting point of the current paper.

Since we are looking for RAAGs which are rigid, those with small quasi-isometry groups would be reasonable candidates. However, even in the atomic case, the quasi-isometry group $\mathrm{QI}(G(\Gamma))$ is huge; see the discussion of quasi-isometry flexibility in [9, Section 11]. Then we turn to the outer automorphism group $\mathrm{Out}(G(\Gamma))$ for guidance.

Now we ask whether those RAAGs with "small" outer automorphism groups are also geometrically rigid in an appropriate sense. Actually, "small" outer automorphism groups and (quasi-isometric or commensurability) rigidity results come together in several other cases: for example, higher rank lattices (Eskin [25], Eskin and Farb [26], Kleiner and Leeb [45] and Mostow [51]), mapping class groups (Behrstock, Kleiner, Minsky and Mosher [6] and Hamenstädt [37]), $\mathrm{Out}(F_n)$ (Farb and Handel [27]), etc. Our first result is about the quasi-isometric classification for RAAGs with finite outer automorphism group.

Theorem 1.1 *Pick $G(\Gamma_1)$ and $G(\Gamma_2)$ such that $\mathrm{Out}(G(\Gamma_i))$ is finite for $i = 1, 2$. Then they are quasi-isomeric if and only if they are isomorphic.*

This theorem is proved in Section 4. See Theorem 4.13 for a more detailed version.

The collection of RAAGs with finite outer automorphism group is a reasonably large class. Recall that there is a one-to-one correspondence between finite simplicial graphs and RAAGs (see Droms [23]); thus it makes sense to talk about a random RAAG by considering the Erdös–Rényi model for random graphs. If the parameters of the model are in the right range, then almost all RAAGs have finite outer automorphism group; see Charney and Farber [18] and Day [22].

The class of 2–dimensional RAAGs with finite outer automorphism group is strictly larger than the class of atomic RAAGs; moreover, there are plenty of higher-dimensional RAAGs with finite outer automorphism group.

Whether $\mathrm{Out}(G(\Gamma))$ is finite or not can be easily read from Γ. We defined the *closed star* of a vertex v in Γ, denoted by $\mathrm{St}(v)$, to be the full subgraph (see Section 2.1) spanned by v and vertices adjacent to v. Similarly, $\mathrm{lk}(v)$ is defined to be the full subgraph spanned by vertices adjacent to v. Note that this definition is slightly different from the usual one.

By results from Laurence [49] and Servatius [57], $\mathrm{Out}(G(\Gamma))$ is generated by the following four types of elements (we identify the vertex set of Γ with a standard generating set of $G(\Gamma)$):

(1) Given a vertex $v \in \Gamma$, the group automorphism defined by sending $v \to v^{-1}$ and fixing all other generators.

(2) Graph automorphisms of Γ.

(3) If $\mathrm{lk}(w) \subset \mathrm{St}(v)$ for vertices $w, v \in \Gamma$, sending $w \to wv$ and fixing all other generators induces a group automorphism, called a *transvection*. It is an *adjacent transvection* when $d(v,w) = 1$, and a *nonadjacent transvection* otherwise.

(4) Suppose $\Gamma \setminus \mathrm{St}(v)$ is disconnected. Then one obtains a group automorphism by picking a connected component C and sending $w \to vwv^{-1}$ for each vertex $w \in C$ (all other generators are fixed). It is called a *partial conjugation*.

Elements of type (3) or (4) have infinite order in $\mathrm{Out}(G(\Gamma))$ while elements of type (1) or (2) are of finite order. $\mathrm{Out}(G(\Gamma))$ is finite if and only if Γ does not contain any separating closed star and there do not exist distinct vertices $v, w \in \Gamma$ such that $\mathrm{lk}(w) \subset \mathrm{St}(v)$.

Theorem 1.2 *Suppose* $\mathrm{Out}(G(\Gamma_1))$ *is finite. Then the following are equivalent:*

(1) $G(\Gamma_2)$ *is quasi-isometric to* $G(\Gamma_1)$.

(2) $G(\Gamma_2)$ *is isomorphic to a subgroup of finite index in* $G(\Gamma_1)$.

(3) Γ_2^e *is isomorphic to* Γ_1^e.

Here Γ^e denotes the *extension graph* introduced by Kim and Koberda in [42]; see Definition 2.11. Extension graphs can be viewed as "curve graphs" for RAAGs; see Kim and Koberda [43]. This analog carries on to the aspect of quasi-isometric rigidity. Namely, if G is a mapping class group and $q\colon G' \to G$ is a quasi-isometry, then it is shown in [6] that G' naturally acts on the curve graph associated with G. This is still true if G is a RAAG with some restriction on its outer automorphism group, for example, if $\mathrm{Out}(G)$ is finite.

However, in general, there exists a pair of commensurable RAAGs with different extension graphs; see Example 3.22. There also exists a pair of RAAGs, not quasi-isometric, with isomorphic extension graphs; see Huang [38, Section 5.3].

Motivated by Theorem 1.2(2), we now look at finite-index *RAAG subgroups* (ie subgroups which are also RAAGs) of $G(\Gamma_1)$.

Given a RAAG $G(\Gamma)$ (not necessarily having a finite outer automorphism group) with a standard generating set S, let d_S be the word metric on $G(\Gamma)$ with respect to S. A subset $K \subset G(\Gamma)$ is S*–convex* if for any three points $x, y \in K$ and $z \in G(\Gamma)$ such that $d_S(x,y) = d_S(x,z) + d_S(z,y)$, we must have $z \in K$. Every finite S–convex subset K naturally gives rise to a finite-index RAAG subgroup $G \le G(\Gamma)$ such that K is the fundamental domain of the left action $G \curvearrowright G(\Gamma)$. For example, if $G(\Gamma) = \mathbb{Z} \oplus \mathbb{Z}$ and K is a rectangle of size n by m, then the corresponding subgroup is of the form $n\mathbb{Z} \oplus m\mathbb{Z}$. The detailed construction in the more general case is given in Section 6.1. G is called an S*–special* subgroup of $G(\Gamma)$. A subgroup of $G(\Gamma)$ is *special* if it is S–special for some standard generating set S. A similar construction in the case of right-angled Coxeter groups is in Haglund [34].

Here is an alternative description in terms of the canonical completion introduced by Haglund and Wise [35]. Let $S(\Gamma)$ be the Salvetti complex of $G(\Gamma)$ (see Section 2.4) and let $X(\Gamma)$ be the universal cover. We pick an identification between $G(\Gamma)$ and the 0–skeleton of $X(\Gamma)$. The above subset K gives rise to a convex subcomplex $\bar{K} \subset X(\Gamma)$. Then the corresponding special subgroup is the fundamental group of the canonical completion with respect the local isometry $\bar{K} \to S(\Gamma)$.

Our next result says if $\mathrm{Out}(G(\Gamma))$ is finite, then this is the only way to obtain finite-index RAAG subgroups of $G(\Gamma)$.

Theorem 1.3 *Suppose* $\mathrm{Out}(G(\Gamma))$ *is finite, and let S be a standard generating set for $G(\Gamma)$. Then all finite-index RAAG subgroups are S–special. Moreover, there is a one-to-one correspondence between nonnegative finite S–convex subsets of $G(\Gamma)$ based at the identity and finite-index RAAG subgroups of $G(\Gamma)$.*

See Theorem 6.13 for a slight reformulation of Theorem 1.3.

We need to explain two terms: nonnegative and based at the identity. For example, take $G = n\mathbb{Z} \oplus m\mathbb{Z}$ inside $\mathbb{Z} \oplus \mathbb{Z}$; then any n by m rectangle could be the fundamental domain for the action of G. We naturally require the rectangle to be in the first quadrant and contain the identity, which would give us a unique choice. Similar things can be done in all RAAGs, and these two terms will be defined precisely in Section 6.

The most simple example is when $G(\Gamma) = \mathbb{Z}$; we have a one-to-one correspondence between finite-index subgroups of the form $n\mathbb{Z}$ and intervals of the form $[0, n-1]$.

Corollary 1.4 *If* $\mathrm{Out}(G(\Gamma_1))$ *is finite, then* $G(\Gamma_2)$ *is quasi-isometric to* $G(\Gamma_1)$ *if and only if* $G(\Gamma_2)$ *is isomorphic to a special subgroup of* $G(\Gamma_1)$*.*

It turns out that there is an algorithm to enumerate the defining graphs of all special subgroups of a RAAG:

Theorem 1.5 *If* $\mathrm{Out}(G(\Gamma))$ *is finite, then* $G(\Gamma')$ *is quasi-isometric to* $G(\Gamma)$ *if and only if* Γ' *can be obtained from* Γ *by finitely many* GSEs*. In particular, there is an algorithm to determine whether* $G(\Gamma')$ *and* $G(\Gamma)$ *are quasi-isometric by looking at the graphs* Γ *and* Γ'*.*

A GSE is a generalized version of a star extension in [9, Example 1.4]; see also [42, Lemma 50]. It will be defined in Section 6.

A question motivated by Theorem 1.2 is the following:

Question 1.6 *Let* $G(\Gamma)$ *be a RAAG such that* $\mathrm{Out}(G(\Gamma))$ *is finite, and let* H *be a finitely generated group quasi-isometric to* $G(\Gamma)$*. What can we say about* H*?*

As a partial answer to this question, we have the following:

Theorem 1.7 (Huang and Kleiner [40]) *Let* $G(\Gamma)$ *and* H *be as in Question 1.6. Then the induced quasi-action* $H \curvearrowright X(\Gamma)$ *is quasi-isometrically conjugate to a geometric action* $H \curvearrowright X'$*. Here* X' *is a* $\mathrm{CAT}(0)$ *cube complex which is closely related to* $X(\Gamma)$*.*

1.2 Comments on the proof

1.2.1 Proof of Theorem 1.1 We start by introducing notation. The Salvetti complex of $G(\Gamma)$ is denoted by $S(\Gamma)$, the universal cover of $S(\Gamma)$ is denoted by $X(\Gamma)$, and flats in $X(\Gamma)$ that cover standard tori in $S(\Gamma)$ are called *standard flats*. See Section 2.4 for precise definition of these terms.

Let $q\colon X(\Gamma) \to X(\Gamma')$ be a quasi-isometry. The proof of Theorem 1.1 follows the scheme of the proof of the main theorem in [9]. Similar schemes can also be found in [6; 45]. There are three steps in [9]. First they show that q maps top-dimensional flats to top-dimensional flats up to finite Hausdorff distance. However, the collection of all top-dimensional flats is too large to be linked directly to the combinatorics of RAAGs, so the second step is to show that quasi-isometries preserve standard flats up to finite Hausdorff distance. The third step is to straighten the quasi-isometry such that it actually maps standard flats to standard flats exactly, not just up to finite Hausdorff distance, and the conclusion follows automatically.

In our case, the first step has been done in Huang [39], where we show q still preserves top-dimensional flats up to finite Hausdorff distance in the higher-dimensional case. No assumption on the outer automorphism group is needed for this step.

The second step consists of two parts. First we show q preserves certain top-dimensional maximal products up to finite Hausdorff distance. Then one wishes to pass to standard flats by intersecting these top-dimensional objects. However, in the higher-dimensional case, a lower-dimensional standard flat may not be the intersection of top-dimensional objects, and even when it is an intersection, one may not be able to read this information directly from the defining graph Γ. This is quite different from the 2–dimensional situation in [9] and relies on several new ingredients.

A necessary condition for q to preserve the standard flats is that every element in $\mathrm{Out}(G(\Gamma))$ does so, which implies there could not be any transvections in $\mathrm{Out}(G(\Gamma))$. This condition is also sufficient.

Theorem 1.8 *Suppose* $\mathrm{Out}(G(\Gamma))$ *is transvection-free. Then there exists a positive constant* $D = D(L, A, \Gamma)$ *such that for any standard flat* $F \subset X(\Gamma)$*, there exists a standard flat* $F' \subset X(\Gamma')$ *such that* $d_H(q(F), F') < D$*.*

Here d_H denotes the Hausdorff distance.

In Step 3, we introduce an auxiliary simplicial complex $\mathcal{P}(\Gamma)$, which serves as a link between the asymptotic geometry of $X(\Gamma)$ and the combinatorial structure of $X(\Gamma)$. More precisely, on one hand, $\mathcal{P}(\Gamma)$ can be viewed as a simplified Tits boundary for $X(\Gamma)$; on the other hand, one can read certain information about the wall space structure of $X(\Gamma)$ from $\mathcal{P}(\Gamma)$. This complex turns out to coincide with the extension graph introduced in [42], where it was motivated from the viewpoint of the mapping class group.

Denote the Tits boundary of $X(\Gamma)$ by $\partial_T(X(\Gamma))$, and let $T(\Gamma) \subset \partial_T(X(\Gamma))$ be the union of Tits boundaries of standard flats in $X(\Gamma)$. Then $T(\Gamma)$ has a natural simplicial structure. However, $T(\Gamma)$ contains redundant information; this can be seen in the similar situation where the link of the base point of $S(\Gamma)$ looks more complicated than Γ, but they essentially contain the same information.

This redundancy can be resolved by replacing the spheres in $T(X)$ that arise from standard flats by simplexes of the same dimension. This gives rise to a well-defined simplicial complex $\mathcal{P}(\Gamma)$ since for any standard flats F_1 and F_2 with $\partial_T F_1 \cap \partial_T F_2 \neq \varnothing$, there exists a standard flat F such that $\partial_T F = \partial_T F_1 \cap \partial_T F_2$. See Section 4.1 for more properties of $\mathcal{P}(\Gamma)$.

By Theorem 1.8, if both $\mathrm{Out}(G(\Gamma))$ and $\mathrm{Out}(G(\Gamma'))$ are transvection-free, then q induces a boundary map $\partial q\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$, which is a simplicial isomorphism. Next we want to consider the converse and reconstruct a map $X(\Gamma) \to X(\Gamma')$ from the boundary map ∂q in the following sense. Pick a vertex $p \in X(\Gamma)$, and let $\{F_i\}_{i=1}^n$ be the collection of maximal standard flats containing p. By Theorem 1.8, for each i, there exists a unique maximal standard flat $F_i' \subset X(\Gamma')$ such that $d_H(q(F_i), F_i') < \infty$. One may wish to map p, which turns out to be the intersection of the F_i, to the intersection of all the F_i'. However, in general, $\bigcap_{i=1}^n F_i'$ may be empty, or it may contain more than one point; hence our map may not be well defined.

It turns out that if we also rule out partial conjugations in $\mathrm{Out}(G(\Gamma))$, then $\bigcap_{i=1}^n F_i'$ is exactly a point. This give rises to a well-defined map $\bar{q}\colon X(\Gamma)^{(0)} \to X(\Gamma')^{(0)}$ which maps vertices in a standard flat to vertices in a standard flat. If $\mathrm{Out}(G(\Gamma'))$ is also finite, then we can define an inverse map of $\bar{q}$, and this is enough to deduce Theorem 1.1.

1.2.2 Proof of Theorem 1.2 If only $\mathrm{Out}(G(\Gamma))$ is assumed to be finite, we can still recover the fact that ∂q is a simplicial isomorphism (this is nontrivial, since Theorem 1.8 does not say that for any standard flat $F' \subset X(\Gamma')$, we can find a standard flat $F \subset X(\Gamma)$ such that $d_H(q(F), F') < \infty$). Hence we can define $\bar{q}$ as before. However, the inverse of $\bar{q}$ does not exist in general.

The next step is to trying to extend $\bar{q}$ to a cubical map (Definition 2.1) from $X(\Gamma)$ to $X(\Gamma')$. There are obvious obstructions: though $\bar{q}$ maps vertices in a standard geodesic to vertices in a standard geodesic, $\bar{q}$ may not preserve the order of these vertices. A typical example is given in Figure 1, where one can permute the green level and the red level in a tree; then the order of vertices in the black line is not preserved.

A remedy is to "flip backwards". Namely we will precompose $\bar{q}$ with a sequence of permutations of "levels" such that the resulting map restricted to each standard geodesic respects the order. Then we can extend $\bar{q}$ to a cubical map. This argument relies on the understanding of quasi-isometric flexibility, namely how much room we have to perform these flips. One formulation of this aspect is the following.

Theorem 1.9 *If* $\mathrm{Out}(G(\Gamma))$ *is finite, then* $\mathrm{Aut}(\mathcal{P}(\Gamma)) \cong \mathrm{Isom}(G(\Gamma), d_r)$.

Here d_r denotes the syllable metric, defined in Section 4.3; see also [43, Section 5.2].

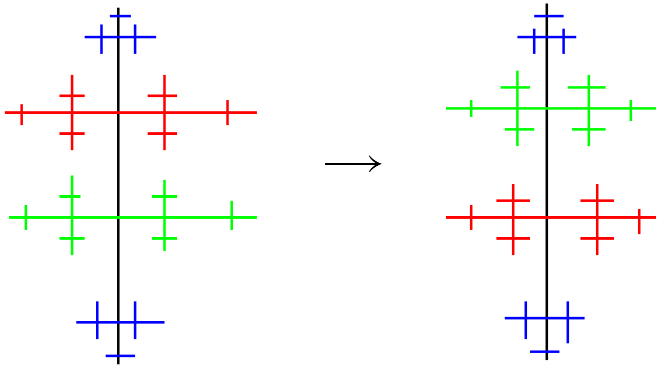


Figure 1: Flipping of levels

Theorem 1.2, Theorem 1.3, Corollary 1.4 and Theorem 1.5 rely on the cubical map $\overline{q}$. In particular, $\overline{q}^{-1}(x)$ ($x \in X(\Gamma')$ is a vertex) is a compact convex subcomplex, and this is how we obtain the S–convex subset in Theorem 1.3.

1.3 Organization of the paper

Section 2 contains basic notation used in this paper and some background material about CAT(0) cube complexes and RAAGs. In particular, Section 2.3 collects several technical lemmas about CAT(0) cube complexes. One can skip Section 2.3 on first reading and come back when needed.

In Section 3, we prove Theorem 1.8. Section 3.1 is about the stability of top-dimensional maximal product subcomplexes under quasi-isometries, and Section 3.2 deals with lower-dimensional standard flats. In Section 4, we prove Theorem 1.1. We will construct the extension complex from our viewpoint in Section 4.1 and explain how is this object is related to Tits boundary, flat space and contact graph. In Section 4.2, we describe how to reconstruct the quasi-isometry.

Sections 5.1 and 5.2 are devoted to proving Theorem 1.2. We prove Theorem 1.3, Corollary 1.4 and Theorem 1.5 in Section 6.

Acknowledgements I would like to thank Bruce Kleiner and Robert Young for reading parts of the paper and giving valuable comments. I thank Jason Behrstock, Ruth Charney and Walter Neumann for suggesting interesting questions, which led to Section 6.3 of this paper. I thank Pallavi Dani, Mark Hagen, Thomas Koberda, Do-gyeom Kim and Sang-hyun Kim for related discussions. I also thank the referee for carefully reading the paper and providing helpful comments.

2 Preliminaries

2.1 Notation and conventions

All graphs in this paper are simple.

The flag complex of a graph Γ is denoted by $F(\Gamma)$; ie $F(\Gamma)$ is a flag complex whose 1–skeleton is Γ.

A subcomplex K' in a combinatorial polyhedral complex K is *full* if K' contains all the subcomplexes of K which have the same vertex set as K'. If K is 1–dimensional, then we also call K' a *full subgraph*.

We use "$*$" to denote the join of two simplicial complexes and "$\circ$" to denote the join of two graphs. Let K be a simplicial complex or a graph. By viewing the 1–skeleton as a metric graph with edge length $=1$, we obtain a metric defined on the 0–skeleton, which we denote by d. Let $N \subset K$ be a subcomplex. We define the *orthogonal complement* of N, denoted by $N^{\perp}$, to be the set $\{w \in K^{(0)} \mid d(w,v)=1$ for any vertex $v \in N\}$; define the *link* of N, denoted by $\mathrm{lk}(N)$, to be the full subcomplex spanned by $N^{\perp}$; and define the *closed star* of N, denoted by $\mathrm{St}(N)$, to be the full subcomplex spanned by $N \cup \mathrm{lk}(N)$. Suppose L is a subcomplex such that $N \subset L \subset K$. We denote the closed star of N in L by $\mathrm{St}(N,L)$. If L is a full subcomplex, then $\mathrm{St}(N,L)=\mathrm{St}(N)\cap L$. We can define $\mathrm{lk}(N,L)$ in a similar way. Let $M \subset K$ be an arbitrary subset. We denote the collection of vertices inside M by $v(M)$.

We use id to denote the identity element of a group, and we use Id to denote the identity map from a space to itself.

Let (X,d) be a metric space. The open ball of radius r centered at p in X will be denoted by $B(p,r)$. Given subsets $A, B \subset X$, the open r–neighborhood of a subset A is denoted by $N_r(A)$. The diameter of A is denoted by $\mathrm{diam}(A)$. The Hausdorff distance between A and B is denoted by $d_H(A,B)$. We will also use the adapted notation of coarse set theory introduced in [50], displayed in the following table:

symbol	meaning
$A \subset_r B$	$A \subset N_r(B)$
$A \subset_\infty B$	$\exists r>0$ such that $A \subset N_r(B)$
$A \overset{r}{=} B$	$d_H(A,B) \le r$
$A \overset{\infty}{=} B$	$d_H(A,B) < \infty$
$A \cap_r B$	$N_r(A) \cap N_r(B)$

2.2 CAT(0) space and CAT(0) cube complex

The standard reference for CAT(0) spaces is [13].

Let (X,d) be a CAT(0) space. Pick $x, y \in X$, we denote the unique geodesic segment joining x and y by $\overline{xy}$. For $y, z \in X \setminus \{x\}$, denote the comparison angle between $\overline{xy}$ and $\overline{xz}$ at x by $\overline{\angle}_x(y,z)$ and the Alexandrov angle by $\angle_x(y,z)$.

The boundary of X, denoted by ∂X, is the collection of asymptotic classes of geodesic rays. The boundary ∂X has an angular metric, which is defined by

$$\angle(\xi_1, \xi_2) = \lim_{t,t' \to \infty} \overline{\angle}_p(l_1(t), l_2(t')),$$

where l_1 and l_2 are unit speed geodesic rays emanating from a base point p such that $l_i(\infty) = \xi_i$ for $i = 1, 2$. This metric does not depend on the choice of p, and the length metric associated to the angular metric, denoted by d_T, is called the Tits metric. The Tits boundary, denoted by $\partial_T X$, is the CAT(1) space $(\partial X, d_T)$; see [13, Chapters II.8 and II.9].

Given two metric spaces (X_1, d_1) and (X_2, d_2), denote the *Cartesian product* of X_1 and X_2 by $X_1 \times X_2$, ie $d = \sqrt{d_1^2 + d_2^2}$ on $X_1 \times X_2$. If X_1 and X_2 are CAT(0), then so is $X_1 \times X_2$.

An *n–flat* in a CAT(0) space X is the image of an isometric embedding $\mathbb{E}^n \to X$. Note that any flat is convex in X.

Pick a convex subset $C \subset X$; then C is also CAT(0). We use π_C to denote the nearest point projection from X to C; it is well defined and 1–Lipschitz. Moreover, pick $x \in X \setminus C$; then $\angle_{\pi_C(x)}(x, y) \geq \frac{\pi}{2}$ for any $y \in C$ such that $y \neq \pi_C(x)$; see [13, Proposition II.2.4].

If $C' \subset X$ is another convex set, then C' is *parallel* to C if $d(\cdot, C)|_{C'}$ and $d(\cdot, C')|_C$ are constant functions. There is a natural isomorphism between $C \times [0, d(C, C')]$ and the convex hull of C and C' in this case. We define the *parallel set* of C, denoted by P_C, to be the union of all convex subsets of X parallel to C. If C has the geodesic extension property, or more generally, C is boundary-minimal (see [14, Section 3.C]), then P_C is a convex subset in X. Moreover, P_C admits a canonical splitting $P_C = C \times C^{\perp}$, where $C^{\perp}$ is also a CAT(0) space.

Now we turn to CAT(0) cube complexes. All cube complexes in this paper are assumed to be finite dimensional.

A cube complex X is obtained by gluing a collection of unit Euclidean cubes isometrically along their faces, see [13, Definition II.7.32] for a precise definition. Then the cube complex has a natural piecewise Euclidean metric. This metric is complete and geodesic if X is finite dimensional [13, I.7.19] and is nonpositively curved if the link of each vertex is a flag complex [29]. If in addition X is simply connected, then this metric is CAT(0) and X is said to be a CAT(0) *cube complex*. We can put a different metric on the 1–skeleton $X^{(1)}$ by considering it as a metric graph with all edge lengths 1. This is called the ℓ^1 *metric*. We use d for the CAT(0) metric on X and d_{ℓ^1} for the ℓ^1 metric on $X^{(1)}$. The natural injection $(X^{(1)}, d_{\ell^1}) \hookrightarrow (X, d)$ is a quasi-isometry; see [13, I.7.31] or [15, Lemma 2.2]. In this paper, we will mainly use the CAT(0) metric unless otherwise specified. Also any notions which depend on the metric, like geodesic, convex subset, convex hull etc, will be understood automatically with respect to the CAT(0) metric unless otherwise specified.

Definition 2.1 [15, Section 2.1] A cellular map between CAT(0) cube complexes is *cubical* if its restriction $\sigma \to \tau$ between cubes factors as $\sigma \to \eta \to \tau$, where the first map $\sigma \to \eta$ is a natural projection onto a face of σ and the second map $\eta \to \tau$ is an isometry.

A *geodesic segment, geodesic ray* or *geodesic line* in X is an isometric embedding of $[a, b]$, $[0, \infty)$ or $\mathbb{R}$ into X with respect to the CAT(0) metric. A *combinatorial geodesic segment, combinatorial geodesic ray* or *combinatorial geodesic* is an ℓ^1–isometric embedding of $[a, b]$, $[0, \infty)$ or $\mathbb{R}$ into $X^{(1)}$ such that its image is a subcomplex.

Let X be a CAT(0) cube complex and let $Y \subset X$ be a subcomplex. Then the following are equivalent (see [34]):

(1) Y is convex with respect to the CAT(0) metric.

(2) Y is a full subcomplex and $Y^{(1)} \subset X^{(1)}$ is convex with respect to the ℓ^1 metric.

(3) $\mathrm{Lk}(p, Y)$ (the link of p in Y) is a full subcomplex of $\mathrm{Lk}(P, X)$ for every vertex $p \in Y$.

The collection of convex subcomplexes in a CAT(0) cube complex enjoys the following version of Helly's property [28]:

Lemma 2.2 *Let X be as above, and let $\{C_i\}_{i=1}^k$ be a collection of convex subcomplexes. If $C_i \cap C_j \neq \varnothing$ for any $1 \le i \neq j \le k$, then $\bigcap_{i=1}^k C_i \neq \varnothing$.*

Lemma 2.3 *Let X_1 and X_2 be two CAT(0) cube complexes, and let $K \subset X_1 \times X_2$ be a convex subcomplex. Then K admits a splitting $K = K_1 \times K_2$, where K_i is a convex subcomplex of X_i for $i = 1, 2$.*

The lemma is clear when $X_1 \cong [0, 1]$, and the general case follows from this special case.

Now we come to the notion of hyperplane, which is the cubical analog of "track" introduced in [24]. A *hyperplane* h in a cube complex X is a subset such that:

(1) h is connected.

(2) For each cube $C \subset X$, either $h \cap C$ is empty or it is a union of mid-cubes of C.

(3) h is minimal; ie if there exists $h' \subset h$ satisfying (1) and (2), then $h = h'$.

Recall that a *mid-cube* of $C = [0,1]^n$ is a subset of the form $f_i^{-1}\left(\frac{1}{2}\right)$, where f_i is one of the coordinate functions.

If X is a CAT(0) cube complex, then the following are true (see [56]):

(1) Each hyperplane is embedded; ie $h \cap C$ is either empty or a mid-cube of C (in more general cube complexes, it is possible that $h \cap C$ contains two or more mid-cubes of C).

(2) h is a convex subset in X, and h with the induced cell structure from X is also a CAT(0) cube complex.

(3) $X \setminus h$ has exactly two connected components; they are called *halfspaces*. The closure of a halfspace is called *closed halfspace*, which is also convex in X with respect to the CAT(0) metric.

(4) Let N_h be the smallest subcomplex of X that contains h. Then N_h is a convex subcomplex of X, and there is a natural isometry $i\colon N_h \to h \times [0,1]$ such that $i(h) = h \times \left\{\frac{1}{2}\right\}$. N_h is called the *carrier* of h.

(5) For every edge $e \subset X$, there exists a unique hyperplane h_e which intersects e in its midpoint. In this case, we say h_e is the hyperplane dual to e and e is an edge dual to the hyperplane h_e.

(6) Lemma 2.2 is also true for a collection of hyperplanes.

Now it is easy to see an edge path $\omega \subset X$ is a combinatorial geodesic segment if and only if there do not exist two different edges of ω such that they are dual to the same hyperplane. Moreover, for two vertices $v, w \in X$, their ℓ^1 distance is exactly the number of hyperplanes that separate v from w.

Pick an edge $e \subset X$, and let $\pi_e\colon X \to e \cong [0,1]$ be the CAT(0) projection. Then:

(1) The hyperplane dual to e is exactly $\pi_e^{-1}\left(\frac{1}{2}\right)$.

(2) $\pi_e^{-1}(t)$ is convex in X for any $0 \le t \le 1$; moreover, if $0 < t < t' < 1$, then $\pi_e^{-1}(t)$ and $\pi_e^{-1}(t')$ are parallel.

(3) Let N_{h_e} be the carrier of the hyperplane dual to e. Then N_{h_e} is the closure of $\pi_e^{-1}(0,1)$. Alternatively, we can describe N_{h_e} as the parallel set of e.

2.3 Coarse intersections of convex subcomplexes

Lemma 2.4 [39, Lemma 2.10] *Let X be a* CAT(0) *cube complex of dimension n, and let C_1, C_2 be convex subcomplexes. Put $\Delta = d(C_1, C_2)$. Let $Y_1 = \{y \in C_1 \mid d(y, C_2) = \Delta\}$ and $Y_2 = \{y \in C_2 \mid d(y, C_1) = \Delta\}$. Then:*

(1) *Y_1 and Y_2 are not empty.*

(2) *Y_1 and Y_2 are convex; π_{C_1} maps Y_2 isometrically onto Y_1 and π_{C_2} maps Y_1 isometrically onto Y_2; the* CAT(0) *convex hull of $Y_1 \cup Y_2$ is isometric to $Y_1 \times [0, \Delta]$ (since we are taking the* CAT(0) *convex hull, it does not has to be a subcomplex).*

(3) *Y_1 and Y_2 are subcomplexes, and $\pi_{C_2}|_{Y_1}$ is a cubical isomorphism from Y_1 to Y_2 with its inverse given by $\pi_{C_1}|_{Y_2}$.*

(4) *For any $\epsilon > 0$, there exists $A = A(\Delta, n, \epsilon)$ such that if $d(p_1, Y_1) \geq \epsilon > 0$ and $d(p_2, Y_2) \geq \epsilon > 0$ for $p_1 \in C_1$ and $p_2 \in C_2$, then*

$$d(p_1, C_2) \geq \Delta + \mathrm{A}d(p_1, Y_1) \quad \text{and} \quad d(p_2, C_1) \geq \Delta + \mathrm{A}d(p_2, Y_2). \tag{2-1}$$

Remark 2.5 Equation (2-1) implies for any $r > 0$, we have $(C_1 \cap_r C_2) \subset_{r'} Y_i$ ($i = 1, 2$), where $r' = \min(1, (2r - \Delta)/A) + r$ and $A = A(\Delta, n, 1)$. Moreover, $\partial_T C_1 \cap \partial_T C_2 = \partial_T Y_1 = \partial_T Y_2$.

The remark implies $Y_1 \overset{\infty}{=} Y_2 \overset{\infty}{=} (C_1 \cap_r C_2)$ for r large enough. We use $\mathcal{I}(C_1, C_2) = (Y_1, Y_2)$ to describe this situation, where $\mathcal{I}$ stands for the word "intersect". The next lemma gives a combinatorial description of Y_1 and Y_2.

Lemma 2.6 *Let X, C_1, C_2, Y_1 and Y_2 be as above. Pick an edge e in Y_1 (or Y_2), and let h be the hyperplane dual to e. Then $h \cap C_i \neq \varnothing$ for $i = 1, 2$. Conversely, if a hyperplane h' satisfies $h' \cap C_i \neq \varnothing$ for $i = 1, 2$, then h' is the dual hyperplane of some edge e' in Y_1 (or Y_2). Moreover, $\mathcal{I}(h' \cap C_1, h' \cap C_2) = (h' \cap Y_1, h' \cap Y_2)$.*

Proof The first part follows from Lemma 2.4. Let $\mathcal{I}(h' \cap C_1, h' \cap C_2) = (Y_1', Y_2')$. Pick $x \in Y_1'$ and let $x' = \pi_{h' \cap C_2}(x) \in Y_2'$. Then $\pi_{h' \cap C_1}(x') = x$. Let $N_{h'} = h' \times [0, 1]$ be the carrier of h'. Then $(h' \cap C_i) \times \left(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right) = C_i \cap \left(h' \times \left(\frac{1}{2} - \epsilon, \frac{1}{2} + \epsilon\right)\right)$ for $i = 1, 2$ and $\epsilon < \frac{1}{2}$. Thus for any $y \in C_2$, we have $\angle_{x'}(x, y) \geq \frac{\pi}{2}$, which implies $x' = \pi_{C_2}(x)$. Similarly, $x = \pi_{C_1}(x') = \pi_{C_1} \circ \pi_{C_2}(x)$; hence $x \in Y_1$ and $Y_1' \subset Y_1$. By the same argument, $Y_2' \subset Y_2$; thus $Y_i' = Y_i \cap h'$ for $i = 1, 2$, and the lemma follows. □

Lemma 2.4, Remark 2.5 and Lemma 2.6 can also be applied to CAT(0) rectangle complexes of finite type, whose cells are of the form $\prod_{i=1}^{n} [0, a_i]$. "Finite type" means there are only finitely many isometry types of rectangle cells in the rectangle complex.

Lemma 2.7 *Let X, C_1, C_2, Y_1 and Y_2 be as above. If h is a hyperplane separating C_1 from C_2, then there exists a convex set $Y \subset h$ such that Y is parallel to Y_1 (or Y_2).*

Proof Let $\Delta = d(C_1, C_2)$, and let $M = Y_1 \times [0, \Delta]$ be the convex hull of Y_1 and Y_2. We want to prove $M \cap h = Y_1 \times \{t\} \subset Y_1 \times [0, \Delta]$ for some $t \in [0, \Delta]$. It suffices to show for any edge $e \subset Y_1$, we have $(e \times [0, \Delta]) \cap h = e \times \{t\}$ for some t.

Pick a point $x \in e$, and consider the point $\{x\} \times \{t\}$ in $M = Y_1 \times [0, \Delta]$. Since $e \times \{t\}$ and e are parallel, $e \times \{t\}$ sits inside a cube and is parallel to an edge of this cube. Thus either $e \times \{t\} \subset h$ or $e \times \{t\}$ is parallel to some edge dual to h. But the second case implies that h is dual to e and $h \cap Y_1 \neq \emptyset$, which is impossible, so $e \times \{t\} \subset (e \times [0, \Delta]) \cap h$. Now we are done since $(\{x\} \times [0, \Delta]) \cap h$ is exactly one point for each $x \in e$. □

2.4 Right-angled Artin groups

Pick a finite simplicial graph Γ. Let $G(\Gamma)$ be a RAAG. A generating set $S \subset G(\Gamma)$ is called a *standard generating set* if all relators in the associated group presentation are commutators. Each standard generating set S determines a graph Γ_S whose vertices are elements in S, and two vertices are adjacent if the corresponding group elements commute. It follows from [23] that the isomorphism type of Γ_S does not depend on the choice of the standard generating set S; in particular, Γ_S is isomorphic to Γ.

Let S be a standard generating set for $G(\Gamma)$. We label the vertices of Γ by elements in S. The RAAG $G(\Gamma)$ has a nice Eilenberg–Mac Lane space $S(\Gamma)$, called the Salvetti complex; see [17; 16]. This is a nonpositively curved cube complex. The 2–skeleton of $S(\Gamma)$ is the usual presentation complex of $G(\Gamma)$. If the presentation complex contains a copy of 2–skeleton of a 3–torus, then we attach a 3–cell to obtain a 3–torus. We can build $S(\Gamma)$ inductively in this manner, and this process will stop after finitely many steps. The closure of each k–cell in $S(\Gamma)$ is a k–torus. A torus of this kind is called a *standard torus*. There is a one-to-one correspondence between the k–cells (or standard tori of dimension k) in $S(\Gamma)$ and k–cliques (complete subgraphs of k vertices) in Γ; thus $\dim(S(\Gamma)) = \dim(F(\Gamma)) + 1$. We define the *dimension* of $G(\Gamma)$ to be the dimension of $S(\Gamma)$.

Denote the universal cover of $S(\Gamma)$ by $X(\Gamma)$, which is a CAT(0) cube complex. Our previous labeling of vertices of Γ induces a labeling of the standard circles of $S(\Gamma)$, which lifts to a labeling of edges of $X(\Gamma)$. We choose an orientation for each standard circle of $S(\Gamma)$, and this gives us a directed labeling of the edges in $X(\Gamma)$. If we pick a base point $v \in X(\Gamma)$ (v is a vertex), then there is a one-to-one correspondence between words in $G(\Gamma)$ and edge paths in $X(\Gamma)$ which start at v.

Each full subgraph $\Gamma' \subset \Gamma$ gives rise to a subgroup $G(\Gamma') \hookrightarrow G(\Gamma)$. A subgroup of this kind is called a S*–standard subgroup* and a left coset of an S–standard subgroup is called an S*–standard coset* (we will omit S when the generating set is clear). There is also an embedding $S(\Gamma') \hookrightarrow S(\Gamma)$ which is locally isometric. Let $p\colon X(\Gamma) \to S(\Gamma)$ be the covering map. Then each connected component of $p^{-1}(S(\Gamma'))$ is a convex subcomplex isometric to $X(\Gamma')$. We will call these components *standard subcomplexes with defining graph* Γ'. A *standard* k*–flat* is a standard complex which covers a standard k–torus in $S(\Gamma)$. When $k = 1$, we also call it a *standard geodesic*.

We pick an identification of the Cayley graph of $G(\Gamma)$ with the 1–skeleton of $X(\Gamma)$; hence $G(\Gamma)$ is identified with the vertices of $X(\Gamma)$. Let $v \in X(\Gamma)$ be the base vertex which corresponds to the identity in the Cayley graph of $G(\Gamma)$. Then for any $h \in G(\Gamma)$, the convex hull of $\{hgv\}_{g\in G(\Gamma')}$ is a standard subcomplex associated with Γ'. Thus there is a one-to-one correspondence between standard subcomplexes with defining graph Γ' in $X(\Gamma)$ and left cosets of $G(\Gamma')$ in $G(\Gamma)$.

Note that for every edge $e \in X(\Gamma)$, there is a vertex in Γ which shares the same label as e, and we denote this vertex by V_e. If $K \subset X(\Gamma)$ is a subcomplex, we define V_K to be $\{V_e \mid e \text{ is an edge in } K\}$ and Γ_K to be the full subgraph spanned by V_K. This subgraph is called the *support* of K. In particular, if K is a standard subcomplex, then the *defining graph* of K is Γ_K.

Every finite simplicial graph Γ admits a canonical join decomposition

$$\Gamma = \Gamma_1 \circ \Gamma_2 \circ \cdots \circ \Gamma_k,$$

where Γ_1 is the maximal clique join factor and Γ_i does not allow any nontrivial join decomposition and is not a point, for $2 \le i \le k$. The graph Γ is *irreducible* if this join decomposition is trivial. This decomposition induces a product decomposition $X(\Gamma) = \mathbb{E}^n \times \prod_{i=2}^{k} X(\Gamma_i)$, which is called the *De Rahm decomposition* of $X(\Gamma)$. This is consistent with the canonical product decomposition of CAT(0) cube complex discussed in [15, Section 2.5].

We turn to the asymptotic geometry of RAAGs. A right-angled Artin group $G(\Gamma)$ is one-ended if and only if Γ is connected. Moreover, the n–connectivity at infinity of $G(\Gamma)$ can be read off from Γ; see [11]. In order to classify all RAAGs up to quasi-isometry, it suffices to consider those one-ended RAAGs. This follows from the main results in [53]. Moreover, we deduce the following lemma from [53, Lemma 3.2].

Lemma 2.8 *If $q\colon X(\Gamma) \to X(\Gamma')$ is an (L, A)–quasi-isometry, then there exists $D = D(L, A) > 0$ such that for any connected component $\Gamma_1 \subset \Gamma$ where Γ_1 is not a point and any standard subcomplex $K_1 \subset X(\Gamma)$ with defining graph Γ_1, there is a unique connected component $\Gamma'_1 \subset \Gamma'$ and a unique standard subcomplex $K'_1 \subset X(\Gamma')$ with defining graph Γ'_1 such that $d_H(q(K_1), K'_1) < D$.*

It is shown in [4] and [1] that $G(\Gamma)$ has linear divergence if and only if Γ is either a join or a point, which implies Γ being a join is a quasi-isometric invariant. Moreover, their results together with Theorem B of [41] implies that the De Rahm decomposition is stable under quasi-isometry:

Theorem 2.9 *Given $X = X(\Gamma)$ and $X' = X(\Gamma')$, let $X = \mathbb{R}^n \times \prod_{i=1}^k X(\Gamma_i)$ and $X' = \mathbb{R}^{n'} \times \prod_{j=1}^{k'} X(\Gamma'_j)$ be the corresponding De Rahm decompositions. If $\phi\colon X \to X'$ is an (L, A)–quasi-isometry, then $n = n'$, $k = k'$ and there exist constants $L' = L'(L, A)$, $A' = A'(L, A)$ and $D = D(L, A)$ such that after reindexing the factors in X', we have (L', A')–quasi-isometry $\phi_i\colon X(\Gamma_i) \to X(\Gamma'_i)$ with $d(p' \circ \phi, \prod_{i=1}^k \phi_i \circ p) < D$, where $p\colon X \to \prod_{i=1}^k X(\Gamma_i)$ and $p'\colon X' \to \prod_{i=1}^k X(\Gamma'_i)$ are the projections.*

Thus in order to study the quasi-isometric classification of RAAGs, it suffices to study those RAAGs which are one-ended and irreducible, but this will rely on finer quasi-isometric invariants of RAAGs.

Recall that in the case of Gromov hyperbolic spaces, quasi-isometries map geodesics to geodesics up to finite Hausdorff distance, hence induce a well-defined boundary map. The analog of this fact for 2–dimensional RAAGs has been established in [10], ie quasi-isometries map 2–flats to 2–flats up to finite Hausdorff distance. The following is a higher-dimensional generalization of [10, Theorem 3.10].

Theorem 2.10 [39, Theorem 5.20] *If $\phi\colon X(\Gamma_1) \to X(\Gamma_2)$ is an (L, A)–quasi-isometry, then $\dim(X(\Gamma_1)) = \dim(X(\Gamma_2))$, and there is a constant $D = D(L, A)$ such that for any top-dimensional flat $F_1 \subset X(\Gamma_1)$, there is a unique flat $F_2 \subset X(\Gamma_2)$ with $d_H(\phi(F_1), F_2) < D$.*

For each right-angled Artin group $G(\Gamma)$, there is a simplicial graph Γ^e, called the *extension graph*, which is introduced in [42]. Extension graphs can be viewed as "curve graphs" for RAAGs [43].

Definition 2.11 [42, Definition 1] The vertex set of Γ^e consists of words in $G(\Gamma)$ that are conjugate to elements in S (recall S is a standard generating set for $G(\Gamma)$), and two vertices are adjacent in Γ^e if and only if the corresponding words commute in $G(\Gamma)$.

The flag complex of the extension graph is called the *extension complex*.

Since the curve graph captures the combinatorial pattern of how Dehn twist flats intersect in a mapping class group, it plays an important role in the quasi-isometric rigidity of a mapping class group [37; 6]. Similarly, we will see in Section 4 that the extension graph captures the combinatorial pattern of the coarse intersection of a certain collection of flats in a RAAG, and it is a quasi-isometric invariant for certain classes of RAAGs.

3 Stable subgraph

We now study the behavior of certain standard subcomplexes under quasi-isometries.

3.1 Coarse intersection of standard subcomplexes and flats

Lemma 3.1 *Let Γ be a finite simplicial graph and let K_1, K_2 be two standard subcomplexes of $X(\Gamma)$. If $(Y_1, Y_2) = \mathcal{I}(K_1, K_2)$, then Y_1 and Y_2 are also standard subcomplexes.*

Proof The lemma is clear if $K_1 \cap K_2 \neq \varnothing$. Now we assume $d(K_1, K_2) = c > 0$. Pick a vertex $v_1 \in K_1$. By Lemma 2.4, there exists a vertex $v_2 \in K_2$ such that $d(v_1, v_2) = c$. Let $l\colon [0, c] \to X(\Gamma)$ be the unit speed geodesic from v_1 to v_2. We can find a sequence of cubes $\{B_i\}_{i=1}^N$ and $0 = t_0 < t_1 < \cdots < t_{N-1} < t_N = c$ such that each B_i contains $\{l(t) \mid t_{i-1} < t < t_i\}$ as interior points.

Let $V_l = \bigcup_{i=1}^N V_{B_i}$ (recall that V_{B_i} is the collection of the labels of edges in B_i; see Section 2.4) and let $V_i = V_{K_i}$ for $i = 1, 2$. Put $V' = V_1 \cap V_2 \cap V_l^\perp$ (recall that $V_l^\perp$ denotes the orthogonal complement of V_l; see Section 2.1) and let Γ' be the full subgraph spanned by V'. Let Y_1' be the standard subcomplex that has defining graph Γ' and contains v_1 (if V' is empty, then $Y_1' = v_1$). We claim $Y_1 = Y_1'$.

Pick an edge $e \subset K_1$ such that $v_1 \in e$ and $V_e \in V'$. Let h be the hyperplane dual to e and $N_h \cong h \times [0, 1]$ be the carrier of h. Since $d(V_e, w) = 1$ for any $w \in V_l$, we can assume $l \subset h \times \{1\} \subset N_h$. By our definition of V', there is an edge $e' \in K_2$ with $v_2 \in e'$ and h dual to e'; thus e and e' cobound an isometrically embedded flat rectangle (one side of the rectangle is l), which implies $e \subset Y_1$. Let l' be the side of the rectangle opposite to l. We can define $V_{l'}$ similarly as we define V_l; then $V_{l'} = V_l$. Now let ω be any edge path starting at v_1 such that $V_{e'} \in V'$ for any edge $e' \subset \omega$. Then it follows from the above argument and induction on the combinatorial length of ω that $\omega \subset Y_1$, thus $Y_1' \subset Y_1$.

For the other direction, since Y_1 is a convex subcomplex by Lemma 2.4, it suffices to prove every vertex of Y_1 belongs to Y_1'. By the induction argument as above, we only need to show that, for an edge e_1 with $v_1 \in e_1$, if $e_1 \subset Y_1$, then $e_1 \subset Y_1'$. Lemma 2.4 implies that there exists an edge $e_2 \subset Y_2$ such that e_1 and e_2 cobound an isometrically embedded flat rectangle (one side of the rectangle is l). So l is in the carrier of the hyperplane dual to e_1. It follows that $V_{e_1} \in V'$ and $e_1 \subset Y_1'$. □

Corollary 3.2 *Let K_1, K_2, Y_1 and Y_2 be as above. Let h be a hyperplane separating K_1 and K_2 and let e be an edge dual to h. Then $V_e \in V_{Y_1}^\perp = V_{Y_2}^\perp$. In particular, pick a vertex $v \in \Gamma$. Then $v \in V_{Y_1}$ if and only if*

(1) $v \in V_{K_1} \cap V_{K_2}$, *and*

(2) *for any hyperplane h' separating K_1 from K_2 and any edge e' dual to h', $d(v, V_{e'}) = 1$.*

Proof Let l and V_l be as in the proof of Lemma 3.1. Let V_l' be a collection of vertices of Γ such that $v \in V_l'$ if and only if $v = V_{e'}$ for some edge $e' \subset X(\Gamma)$ satisfying (2). It suffices to prove $V_l' = V_l$.

It is clear that $V_l' \subset V_l$ since if a hyperplane h separates K_1 from K_2, then l intersects h transversally at one point. To see $V_l \subset V_l'$, it suffices to show $h \cap K_i = \varnothing$ for $i = 1, 2$, where h is a hyperplane that intersects l transversally. Let $x = l \cap h$. Suppose $h \cap K_1 \neq \varnothing$ and let $x' = \pi_{h \cap K_1}(x)$. Now consider the triangle $\Delta(v_1, x, x')$ (recall that $v_1 = l(0)$). We have $\angle_{v_1}(x, x') \geq \frac{\pi}{2}$ (since $\pi_{K_1}(x) = v_1$), $\angle_{x'}(v_1, x) \geq \frac{\pi}{2}$ (see the proof of Lemma 2.6) and $\angle_x(v_1, x') > 0$, which is a contradiction, so $h \cap K_1 = \varnothing$. Similarly, $h \cap K_2 = \varnothing$. □

Remark 3.3 Recall that a standard coset of $G(\Gamma)$ is a left coset of a standard subgroup of $G(\Gamma)$. Lemma 3.1 implies that for each pair of standard cosets of $G(\Gamma)$, we can associated another standard coset which captures the coarse intersection of the pair. Moreover, we can also define a notion of distance between two standard cosets, which takes values in $G(\Gamma)$.

Recall that Γ_K is the support of K (see Section 2.4), and that $\mathrm{lk}(\Gamma_K)$ is the full subgraph spanned by $V_K^{\perp}$ (see Section 2.1).

Lemma 3.4 *Let $K \subset X(\Gamma)$ be a convex subcomplex and let $\Gamma' = \mathrm{lk}(\Gamma_K)$. Then the parallel set P_K of K is a convex subcomplex and canonically splits as $K \times X(\Gamma')$.*

Note that we do not require K to satisfy the geodesic extension property.

Proof Pick a vertex $v \in K$. Let $\Gamma'' = \Gamma_K$ and let P_1 be the unique standard subcomplex that passes through v and has defining graph $\Gamma' \circ \Gamma''$ (recall that $\circ$ denotes the graph join). Then $K \subset P_1$. Let P' be the natural copy of $K \times X(\Gamma')$ inside P_1. It is clear that $P' \subset P_K$.

Let K' be a convex subset parallel to K, and let $\phi\colon K \to K'$ be the isometry induced by CAT(0) projection onto K'. Pick a vertex $v \in K$, and let l be the geodesic segment connecting v and $\phi(v)$. We define V_l as in the proof of Lemma 3.1 (note that $\phi(v)$ is not necessarily a vertex). Let e be any edge such that $v \in e \subset K$. Then there is a flat rectangle with e, $\phi(e)$ and l as its three sides. Thus l is contained in the carrier of the hyperplane dual to e, and $V_l \subset V_e^{\perp}$. Note that if l' is the side opposite to l, then $V_{l'} = V_l$. For any given edge $e' \subset K$, we can find an edge path $\omega \subset K$ such that e is the first and e' is the last edge in ω. By induction on the combinatorial length of w and the same argument as above, we can show $V_l \subset V_{e'}^{\perp}$, thus $V_l \subset V_K^{\perp}$ and $K' \subset P'$. It follows that $P_K \subset P'$, so $P_K = P'$. $\square$

Remark 3.5 The following is a generalization of the above lemma for general CAT(0) cube complexes. Let X be a CAT(0) cube complex. A convex set $K \subset X$ is *regular* if for any $x \in K$, the space of direction $\Sigma_x K$ of x in K [13, Chapter II.3] satisfies:

(1) $\Sigma_x K$ is a subcomplex of $\Sigma_x X$ with respect to the canonical all-right spherical complex structure on $\Sigma_x X$.

(2) There exists $r > 0$ such that $B(x, r) \cap K$ is isometric to the r–ball centered at the cone point in the Euclidean cone over $\Sigma_x K$.

If $K \subset X$ is a regular convex subset, then P_K is convex and admits a splitting $P_K \cong K \times N$, where N has an induced cubical structure from X (N is CAT(0)).

Lemma 3.6 *Let $q\colon X(\Gamma_1) \to X(\Gamma_2)$ be an (L, A)–quasi-isometry and let $F \subset X(\Gamma_1)$ be a subcomplex isometric to $\mathbb{E}^k$. Suppose $n = \dim(X(\Gamma_1)) = \dim(X(\Gamma_2))$. If there exist $R_1, R_2 > 0$ and top-dimensional flats F_1 and F_2 such that*

$$F \overset{R_2}{=} F_1 \cap_{R_1} F_2 \quad \text{and} \quad F \overset{\infty}{=} F_1 \cap_R F_2$$

for any $R \geq R_1$, then:

(1) *There exist a constant $D = D(L, A, R_1, R_2, n)$ and a subcomplex $F' \subset X(\Gamma_2)$ isometric to $\mathbb{E}^k$ such that $q(F) \overset{D}{=} F'$.*

(2) *There exists a constant $D' = D'(L, A)$ such that $q(P_F) \overset{D'}{=} P_{F'}$.*

Proof By Theorem 2.10, there exist top-dimensional flats $F_1', F_2' \subset X(\Gamma_2)$ such that $q(F_i) \overset{D_1}{=} F_i'$ for $D_1 = D_1(L, A)$ and $i = 1, 2$. Thus there exist $R' = R'(L, A, R_1, R_2)$ and $R_3 = R_3(L, A, R_1, R_2) > R_1$ such that $q(F_1 \cap_{R_1} F_2) \subset F_1' \cap_{R'} F_2' \subset q(F_1 \cap_{R_3} F_2)$; this and Remark 2.5 imply

$$q(F_1 \cap_{R_1} F_2) \overset{D_2}{=} F_1' \cap_{R'} F_2' \tag{3-1}$$

for $D_2 = D_2(n, d(F_1, F_2)) = D_2(L, A, R_1, R_2, n)$.

Let $(Y_1, Y_2) = \mathcal{I}(F_1', F_2')$. Then there exists $D_3 = D_3(L, A, R_1, R_2, n)$ such that

$$Y_1 \overset{D_3}{=} F_1' \cap_{R'} F_2'. \tag{3-2}$$

From (3-1) and (3-2), we have

$$q(F) \overset{D_4}{=} Y_1 \tag{3-3}$$

for $D_4 = D_4(L, A, R_1, R_2, n)$. By Lemma 2.4, Y_1 is a convex subcomplex of F_1'. This together with (3-3) implies $Y_1 = F' \times \prod_{i=1}^{k'} I_i$, where F' is isometric to $\mathbb{E}^k$ and $\{I_i\}_{i=1}^{k'}$ are finite intervals. Moreover, by (3-3), $\operatorname{diam}\bigl(\prod_{i=1}^{k'} I_i\bigr)$ must be bounded in terms of D_4, L and A; thus (1) follows.

Let $\{F_\lambda\}_{\lambda\in\Lambda}$ be the collection of top-dimensional flats in $X(\Gamma_1)$ which are contained in the parallel set P_F of F. Lemma 3.4 implies

$$d_H\Bigl(\bigcup_{\lambda\in\Lambda} F_\lambda, P_F\Bigr) \le 1. \tag{3-4}$$

For $\lambda\in\Lambda$, there exists $R_\lambda > 0$ such that $F \subset_{R_\lambda} F_\lambda$. Let F_λ' be the top-dimensional flat in $X(\Gamma_2)$ such that $q(F_\lambda) \overset{D_1}{=} F_\lambda'$. Then by (1), there exists $R_\lambda' > 0$ such that $F' \subset_{R_\lambda'} (F_\lambda')$. This and Lemma 2.4 imply $F_\lambda' \subset P_{F'}$ for any $\lambda\in\Lambda$. Thus by (3-4), there exists $D' = D'(L, A)$ such that $q(P_F) \subset_{D'} P_{F'}$. And (2) follows by running the same argument for the quasi-isometry inverse of q. □

A *tree product* is a convex subcomplex $K \subset X(\Gamma)$ such that K splits as a product of trees, ie there exists a cubical isomorphism $K \cong \prod_{i=1}^{n} T_i$ where the T_i are trees. A *standard tree product* is a tree product which is also a standard subcomplex.

One can check that K is a standard tree product if and only if the defining graph Γ_K of K has a join decomposition $\Gamma_K = \Gamma_1 \circ \Gamma_2 \circ \cdots \circ \Gamma_n$, where each Γ_i is discrete. Thus one can choose the above T_i to be standard subcomplexes of K. Note that every standard flat is a standard tree product, and every subcomplex isometric to $\mathbb{E}^k$ is a tree product.

Lemma 3.7 *Suppose* $\dim(X(\Gamma)) = n$. *Let* $q\colon X(\Gamma) \to X(\Gamma')$ *be a quasi-isometry. Let* $K = \prod_{i=1}^{n} T_i$ *be a top-dimensional tree product with its tree factors. Then there exists a standard tree product* K' *in* $X(\Gamma')$ *such that* $q(K) \subset_\infty K'$.

The proof essentially follows [9, Theorem 4.2].

Proof For $1 \le i \le n$, let $V_i = V_{T_i} \in \Gamma$ be the collection of labels of edges in T_i. The case where all the V_i are consist of one point follows from Theorem 2.10. If each V_i contains at least two points, then by Lemma 3.6, for any geodesic $l \subset T_i$, there exists a subcomplex $l' \subset X(\Gamma')$ isometric to $\mathbb{R}$ such that $q(l) \overset{\infty}{=} l'$. Since l' is unique up to parallelism, the collection of labels of edges in l' does not depend on the choice of l' and will be denoted by $V_{q(l)}$. For $1 \le i \le n$, define $V_i' = \bigcup_{l\subset T_i} V_{q(l)}$ where l varies among all geodesics in T_i.

We claim $V_i' \subset (V_j')^{\perp}$ for $i \neq j$. To see this, pick geodesic $l_i \in T_i$ and let $F = \prod_{i=1}^{n} l_i$. Then there exist top-dimensional flat F' and geodesic lines $\{l_i'\}_{i=1}^{n}$ (each l_i' is a subcomplex) in $X(\Gamma')$ such that $q(F) \overset{\infty}{=} F'$ and $q(l_i) \overset{\infty}{=} l_i'$. Since $l_i' \subset_\infty F'$, by Lemma 2.4, we can assume l_i' is a subcomplex of F'. Pick $i \neq j$. Since l_i and l_j are orthogonal, they have infinite Hausdorff distance. Thus l_i' and l_j' have infinite Hausdorff distance. By our assumption, l_i' and l_j' are isometric to $\mathbb{E}^1$, and they are convex subcomplexes of $F' \cong \mathbb{E}^n$. Thus either l_i' and l_j' are parallel, or they are orthogonal. The former is impossible since l_i' and l_j' have infinite Hausdorff distance. Thus $\{l_i'\}_{i=1}^{n}$ is a mutually orthogonal collection.

Let $\Gamma_1' = V_1' \circ V_2' \circ \cdots \circ V_n' \subset \Gamma'$. Then each V_i' has to be a discrete full subgraph by our dimension assumption. Let $\{F_\lambda\}_{\lambda \in \Lambda}$ be the collection of top-dimensional flats in K and let F_λ' be the unique flat such that $q(F_\lambda) \overset{\infty}{=} F_\lambda'$. Note that for arbitrary F_{λ_1} and F_{λ_2}, there exists a finite chain in $\{F_\lambda\}_{\lambda \in \Lambda}$ which starts with F_{λ_1} and ends with F_{λ_2} such that the intersection of adjacent elements in the chain contains a top-dimensional orthant. Thus the collection $\{F_\lambda'\}_{\lambda \in \Lambda}$ also has this property. Then $\bigcup_{\lambda \in \Lambda} F_\lambda'$ is contained in a standard subcomplex of $X(\Gamma')$ with defining graph Γ_1'.

It remains to deal with the case where there exist $i \neq j$ such that $|V_i| = 1$ and $|V_j| \geq 2$. We suppose $|V_i| = 1$ for $1 \leq i \leq m$ and $|V_i| \geq 2$ for $i > m$. By applying Lemma 3.6 with $F = \prod_{i=1}^{m} T_i$, we can reduce to a lower-dimensional case, and the lemma follows by induction on dimension. □

Corollary 3.8 *Let $q\colon X(\Gamma) \to X(\Gamma')$ be a quasi-isometry, and let K be a top-dimensional maximal standard tree product; ie K is not properly contained in another tree product. Then there exists a standard tree product $K' \subset X(\Gamma')$ such that $q(K) \overset{\infty}{=} K'$.*

3.2 Standard flats in transvection-free RAAGs

Up to now, we have only dealt with top-dimensional standard subcomplexes. The next goal is to study those standard subcomplexes which are not necessarily top dimensional. In particular, we are interested in whether quasi-isometries will preserve standard flats up to finite Hausdorff distance. The answer turns out to be related to the outer automorphism group of $G(\Gamma)$.

One direction is obvious: namely, if every quasi-isometry $q\colon X(\Gamma) \to X(\Gamma')$ maps any standard flat in $X(\Gamma)$ to a standard flat in $X(\Gamma')$ up to finite Hausdorff distance, then $\mathrm{Out}(G(\Gamma))$ must be transvection-free (ie $\mathrm{Out}(G(\Gamma))$ does not contain any transvections). The converse is also true. Now we set up several necessary tools to prove the converse.

In this section, Γ will always be a finite simplicial graph.

Definition 3.9 A subgraph $\Gamma_1 \subset \Gamma$ is *stable* in Γ if the following are true:

(1) Γ_1 is a full subgraph.

(2) Let $K \subset X(\Gamma)$ be a standard subcomplex such that $\Gamma_K = \Gamma_1$, and let Γ' be a finite simplicial graph such that, for some L and A, there is an (L, A)–quasi-isometry $q\colon X(\Gamma) \to X(\Gamma')$. Then there exists $D = D(L, A, \Gamma_1, \Gamma) > 0$ and a standard subcomplex $K' \subset X(\Gamma')$ such that $d_H(q(K), K') < D$.

For simplicity, we will also say the pair (Γ_1, Γ) is *stable* in this case. A standard subcomplex $K \subset X(\Gamma)$ is *stable* if it arises from a stable subgraph of Γ.

We claim the defining graph $\Gamma_{K'}$ of K' is stable in Γ'. To see this, pick any graph Γ'' so that there is an (L, A)–quasi-isometry $q'\colon X(\Gamma') \to X(\Gamma'')$, and pick any standard subcomplex $K'_1 \subset X(\Gamma')$ with defining graph $\Gamma_{K'}$. Note that there is an isometry $i\colon X(\Gamma') \to X(\Gamma')$ such that $i(K') = K'_1$. Since the map $q' \circ i \circ q$ is a quasi-isometry from $X(\Gamma)$ to $X(\Gamma'')$, we have that $q' \circ i \circ q(K)$ is Hausdorff close to a standard subcomplex in $X(\Gamma'')$ by the stability of Γ_1; hence the same is true for $q'(K'_1)$. It follows from this claim that one can obtain quasi-isometric invariants by identifying certain classes of stable subgraphs.

It is immediate from the definition that for finite simplicial graphs $\Gamma_1 \subset \Gamma_2 \subset \Gamma_3$, if (Γ_1, Γ_2) is stable and (Γ_2, Γ_3) is stable, then (Γ_1, Γ_3) is stable. However, it is not necessarily true that if (Γ_1, Γ_3) and (Γ_2, Γ_3) are stable, then (Γ_1, Γ_2) is stable. In the sequel, we will investigate several other properties of stable subgraphs. The following lemma is an easy consequence of Lemma 3.1 and Remark 2.5:

Lemma 3.10 *Suppose Γ_1 and Γ_2 are stable in Γ. Then $\Gamma_1 \cap \Gamma_2$ is also stable in Γ.*

The following result follows from Lemma 2.8.

Lemma 3.11 *If Γ_1 is stable in Γ, then every connected component of Γ_1 that contains more than one point is also stable in Γ.*

Lemma 3.12 *Suppose Γ_1 is stable in Γ. Let V be the vertex set of Γ_1 and let Γ_2 be the full subgraph spanned by V and the orthogonal complement $V^{\perp}$. Then Γ_2 is also stable in Γ.*

Proof Let $K_2 \subset X(\Gamma)$ be a standard subcomplex with defining graph $\Gamma_{K_2} = \Gamma_2$, and let $K_1 \subset K_2$ be any standard subcomplex satisfying $\Gamma_{K_1} = \Gamma_1$. Lemma 3.4 implies $K_2 = P_{K_1} = K_1 \times K_1^{\perp}$. For a vertex $x \in K_1^{\perp}$, let $K_x = K_1 \times \{x\}$. Let $q\colon X(\Gamma) \to X(\Gamma')$

be an (L,A)–quasi-isometry. Then there exists standard subcomplex K'_x such that $d_H(q(K_x),K'_x) < D = D(L,A,\Gamma_1,\Gamma)$. Thus $K'_x \overset{\infty}{=} K'_y$ for vertices $x,y \in K_1^{\perp}$. It follows from Lemma 3.1 that K'_x and K'_y are parallel. Thus $q(P_{K_1}) \subset_R P_{K'_x}$ for $R = D+L+A$. Moreover, $P_{K'_x}$ is also a standard subcomplex by Lemma 3.4. By considering the quasi-isometry inverse and repeating the previous argument, we know $q(P_{K_1}) \overset{\infty}{=} P_{K'_x}$; thus Γ_2 is also stable in Γ. □

Lemma 3.13 *Suppose Γ_1 is stable in Γ. Pick a vertex $v \notin \Gamma_1$. Then the full subgraph spanned by $v^{\perp} \cap \Gamma_1$ is stable in Γ.*

Proof We use Γ_2 to denote the full subgraph spanned by $v^{\perp} \cap \Gamma_1$. Let $K_2 \subset X(\Gamma)$ be a standard subcomplex such that $\Gamma_{K_2} = \Gamma_2$, and let $K_1 \subset X(\Gamma)$ be the unique standard subcomplex such that $\Gamma_{K_1} = \Gamma_1$ and $K_2 \subset K_1$. Pick a vertex $x \in K_2$, and let $e \subset X(\Gamma)$ be the edge such that $V_e = v$ and $x \in e$. Suppose $\bar{x}$ is the other end point of e. Let $\bar{K}_i$ be the standard subcomplex that contains $\bar{x}$ and has defining graph Γ_i for $i = 1,2$. Denote the hyperplane dual to e by h. Since $v \notin \Gamma_1$, we have $h \cap K_1 = \varnothing$ and $h \cap \bar{K}_1 = \varnothing$; thus h separates K_1 and $\bar{K}_1$, and $d(K_1,\bar{K}_1) = 1$. It follows from Corollary 3.2 that $\mathcal{I}(K_1,\bar{K}_1) = (K_2,\bar{K}_2)$; in particular $K_2 \overset{D}{=} K_1 \cap_R \bar{K}_1$ for D depending on R and the dimension of $X(\Gamma)$. Now the lemma follows since Γ_1 is stable. □

The next result is a direct consequence of Corollary 3.8.

Lemma 3.14 *If Γ_1 is stable in Γ, then there exists Γ_2 which is stable in Γ_1 such that*

(1) *Γ_2 is a graph join $\bar{\Gamma}_1 \circ \bar{\Gamma}_2 \circ \cdots \circ \bar{\Gamma}_k$, where $\bar{\Gamma}_i$ is discrete for $1 \le i \le k$;*

(2) *$k = \dim(X(\Gamma_1))$.*

Lemma 3.15 *Let Γ be a finite simplicial graph such that there do not exist vertices $v \neq w$ of Γ with $v^{\perp} \subset \mathrm{St}(w)$. Then every stable subgraph of Γ contains a stable vertex.*

Proof Let Γ_1 be a minimal stable subgraph; ie it does not properly contain any stable subgraph of Γ. It suffices to show Γ_1 is a point. We argue by contradiction and assume Γ_1 contains more than one point.

First we claim Γ_1 cannot be discrete. Suppose the contrary is true. Pick vertices $v,w \in \Gamma_1$ and pick a vertex $u \in v^{\perp} \setminus \mathrm{St}(w)$. By Lemma 3.13, $u^{\perp} \cap \Gamma_1$ is also stable. Note that $v \in u^{\perp} \cap \Gamma_1$ and $w \notin u^{\perp} \cap \Gamma_1$, which contradicts the minimality of Γ_1.

We claim Γ_1 must be a clique. Since Γ_1 is not discrete, by Lemma 3.14, we can find a stable subgraph

$$\Gamma_2 = \bar{\Gamma}_1 \circ \bar{\Gamma}_2 \circ \cdots \circ \bar{\Gamma}_m \subset \Gamma_1, \tag{3-5}$$

where $\{\overline{\Gamma}_i\}_{i=1}^m$ are discrete full subgraphs and $m \geq 2$. Then $\Gamma_2 = \Gamma_1$. Suppose some $\overline{\Gamma}_i$ contains more than one point, and let Γ_3 be the join of the remaining join factors. Then Theorem 2.9 implies that Γ_3 is stable, contradicting the minimality of Γ_1. Therefore, Γ_1 is a clique.

Pick distinct vertices $v_1, v_2 \in \Gamma_1$. By our assumption, there exists a vertex $w \in v_1^\perp \setminus \mathrm{St}(v_2)$. Since Γ_1 is a clique, $\Gamma_1 \subset \mathrm{St}(v_2)$, so $w \notin \Gamma_1$. Let Γ_4 be the full subgraph spanned by $w^\perp \cap \Gamma_1$. Then Γ_4 is stable by Lemma 3.13. Moreover, $\Gamma_4 \subsetneq \Gamma_1$ (since $v_2 \notin \Gamma_4$), which yields a contradiction. □

Lemma 3.16 *Let Γ be as in Lemma 3.15 and let Γ_1 be a stable subgraph of Γ. Then for any vertex $w \in \Gamma_1$, there exists a stable vertex $v \in \Gamma_1$ such that $d(v, w) \leq 1$.*

Proof Denote the combinatorial distances in Γ and Γ_1 by d and d_1, respectively. Since Γ_1 is a full subgraph, $d(x, y) = 1$ if and only if $d_1(x, y) = 1$, and $d(x, y) \geq 2$ if and only if $d_1(x, y) \geq 2$, for vertices $x, y \in \Gamma_1$. If w is isolated in Γ_1, then we can use the argument in the second paragraph of the proof of Lemma 3.15 to get rid of all vertices in Γ_1 except w, which implies w is a stable vertex. If w is not isolated, we can assume Γ_1 is connected by Lemma 3.11.

By Lemma 3.15, there exists a stable vertex $u \in \Gamma_1$. If $d_1(u, w) \leq 1$, then we are done. Otherwise, let ω be a geodesic in Γ_1 connecting u and w (note that ω might not be a geodesic in Γ), and let $\{v_i\}_{i=0}^n$ be the consecutive vertices in ω; here $v_0 = w$, $v_n = u$ and $n = d_1(w, u)$.

Since u is stable, by Lemma 3.12, $\mathrm{St}(u)$ is also stable. Note that $d_1(v_{n-2}, u) = 2$, so $d(v_{n-2}, u) = 2$ and $v_{n-2} \notin \mathrm{St}(u)$. Lemma 3.13 implies $v_{n-2}^\perp \cap \mathrm{St}(u)$ is stable, and by Lemma 3.10, $v_{n-2}^\perp \cap \mathrm{St}(u) \cap \Gamma_1$ is also stable. Note that $v_{n-2}^\perp \cap \mathrm{St}(u) \cap \Gamma_1 \neq \varnothing$ since it contains v_{n-1}. Lemma 3.15 implies there is a stable vertex $u' \in v_{n-2}^\perp \cap \mathrm{St}(u) \cap \Gamma_1$, and it is easy to see $d_1(w, u') = n-1$. Now the lemma follows by induction. □

Lemma 3.17 *Let Γ be as in Lemma 3.15. Then every vertex of Γ is stable.*

Proof Let Γ_w be the intersection of all the stable subgraphs that contain w. By Lemma 3.10, Γ_w is the minimal stable subgraph that contains w. It suffices to prove $\Gamma_w = \{w\}$. We argue by contradiction and denote the vertices in $\Gamma_w \setminus \{w\}$ by $\{v_i\}_{i=1}^k$. The minimality of Γ_w implies we cannot use Lemma 3.13 to get rid of some v_i while keeping w; thus $w^\perp \setminus \mathrm{St}(v_i) \subset \{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k\}$ for any i. In other words,

$$w^\perp \subset \mathrm{St}(v_i) \cup \{v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k\} \tag{3-6}$$

for $1 \leq i \leq k$. Then there does not exist i such that $\Gamma_w \subset \mathrm{St}(v_i)$: otherwise, we would have $w^\perp \subset \mathrm{St}(v_i)$ by (3-6).

On the other hand, Lemma 3.16 implies there exists a stable vertex $u \in \Gamma_w$ with $d(w,u) = 1$. Then $\mathrm{St}(u)$ is stable (Lemma 3.12) and $\mathrm{St}(u) \cap \Gamma_w$ is stable (Lemma 3.10). Note that $w \in \mathrm{St}(u) \cap \Gamma_w$. By the minimality of Γ_w, we have $\Gamma_w \subset \mathrm{St}(u)$, which yields a contradiction. □

Lemma 3.18 *Let Γ be a finite simplicial graph, and pick stable subgraphs Γ_1, Γ_2 of Γ. Let $\overline{\Gamma}$ be the full subgraph spanned by V and $V^{\perp}$, where $V = V_{\Gamma_1}$. If $\Gamma_2 \subset \overline{\Gamma}$, then the full subgraph spanned by $\Gamma_1 \cup \Gamma_2$ is stable in Γ.*

To simplify notation, in the following proof, we will write $q(K) \approx K'$, where q, K and K' are as in Definition 3.9. We will also assume without loss of generality that $q(K) \subset K'$.

Proof Let $q\colon X(\Gamma) \to X(\Gamma')$ be an (L,A)–quasi-isometry. Suppose K_1 and K are standard subcomplexes in $X(\Gamma)$ such that $\Gamma_{K_1} = \Gamma_1$, $\Gamma_K = \overline{\Gamma}$ and $K_1 \subset K$. Put $K' \approx q(K)$, $K_1' \approx q(K_1)$, $K = K_1 \times K_1^{\perp}$ and $K' = K_1' \times K_1'^{\perp}$. The proof of Lemma 3.12 implies there exist a quasi-isometry $q'\colon K_1^{\perp} \to K_1'^{\perp}$ and a constant $D_1 = D_1(L,A,\Gamma_1,\Gamma)$ such that

$$d(q' \circ p_2(x), p_2' \circ q(x)) < D_1 \tag{3-7}$$

for any $x \in K$, where $p_2\colon K \to K_1^{\perp}$ and $p_2'\colon K' \to K_1'^{\perp}$ are projections.

Let $\Gamma_2 = \Gamma_{21} \circ \Gamma_{22}$, where $\Gamma_{21} = \Gamma_1 \cap \Gamma_2$, and let K_{22}, K_2 be standard subcomplexes such that $\Gamma_{K_{22}} = \Gamma_{22}$, $\Gamma_{K_2} = \Gamma_2$ and $K_{22} \subset K_2 \subset K$. By (3-7), it suffices to prove there exist a standard subcomplex $K_{22}' \subset K'$ and a constant $D = D(L,A,\Gamma_1,\Gamma_2,\Gamma)$ such that $d_H(p_2' \circ q(K_{22}), K_{22}') < D$. Let $K_2' \approx q(K_2)$. Then $K_2' \subset K'$, and $p_2'(K_2')$ is a standard subcomplex. By (3-7), $p_2' \circ q(K_{22}) \overset{\infty}{=} p_2' \circ q(K_2) \overset{\infty}{=} p_2'(K_2')$; thus we can take $K_{22}' = p_2'(K_2')$. □

Remark 3.19 In general, the full subgraph spanned by $\Gamma_1 \cup \Gamma_2$ is not necessarily stable even if Γ_1 and Γ_2 are stable; see Remark 3.26.

The next theorem follows from Lemma 3.17 and Lemma 3.18.

Theorem 3.20 *Given a finite simplicial graph Γ, the following are equivalent:*

(1) $\mathrm{Out}(G(\Gamma))$ *is transvection-free.*

(2) *For any (L,A)–quasi-isometry $q\colon X(\Gamma) \to X(\Gamma')$, there exists a positive constant $D = D(L,A,\Gamma)$ such that for any standard flat $F \subset X(\Gamma)$, there exists a standard flat $F' \subset X(\Gamma')$ with $d_H(q(F), F') < D$.*

3.3 Standard flats in general RAAGs

At this point, we have the following natural questions:

(1) In Theorem 3.20, is it true that every standard flat in $X(\Gamma')$ comes from some standard flat in $X(\Gamma)$? A related question could be, is condition (1) in Theorem 3.20 a quasi-isometric invariant for right-angled Artin groups?

(2) What can we say about the stable subgraphs of Γ if we drop condition (1) in Theorem 3.20?

We will first give a negative answer to question (1) in Example 3.22 below. Then we will prove Theorem 3.21, which answers question (2). Section 4 and, in particular, the proof of Theorem 1.1 will not depend on this subsection. However, we will need Theorem 3.21 and Lemma 3.23 for Section 5.

Theorem 3.21 *Let Γ be an arbitrary finite simplicial graph. A clique $\Gamma_1 \subset \Gamma$ is stable if and only if there do not exist vertices $w \in \Gamma_1$ and $v \in \Gamma \setminus \Gamma_1$ such that $w^{\perp} \subset \mathrm{St}(v)$.*

In other words, the clique Γ_1 is stable if and only if the corresponding $\mathbb{Z}^n$ subgroup of $G(\Gamma_1)$ is invariant under all transvections.

Example 3.22 Let Γ_1 and Γ_2 be as indicated in Figure 2. It is easy to see $\mathrm{Out}(G(\Gamma_1))$ is transvection-free while $\mathrm{Out}(G(\Gamma_2))$ contains nontrivial transvection (Γ_2 has a dead end at vertex u). We claim $G(\Gamma_1)$ and $G(\Gamma_2)$ are commensurable and, in particular, quasi-isometric.

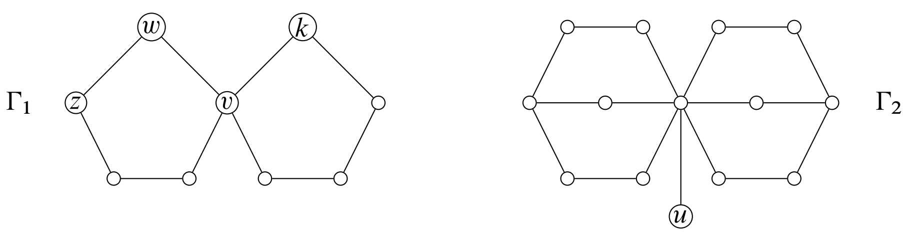


Figure 2: $\mathrm{Out}(G(\Gamma_1))$ is transvection-free while $\mathrm{Out}(G(\Gamma_2))$ contains nontrivial transvection.

Let $\Gamma \subset \Gamma_1$ be the pentagon on the left side and let Y be the Salvetti complex of Γ. Suppose $X_1 = Y \sqcup Y \sqcup (\mathbb{S}^1 \times [0,1])/\sim$; here the two boundary circles of the annulus are identified with two standard circles which are in different copies of Y. Then $\pi_1(X_1) = G(\Gamma_1)$. Define the homomorphism $h_1\colon G(\Gamma) \to \mathbb{Z}/2$ by sending w to the nontrivial element in $\mathbb{Z}/2$ and other generators to the identity element. Let Y' be the 2–sheeted cover of Y with respect to $\ker(h_1)$.

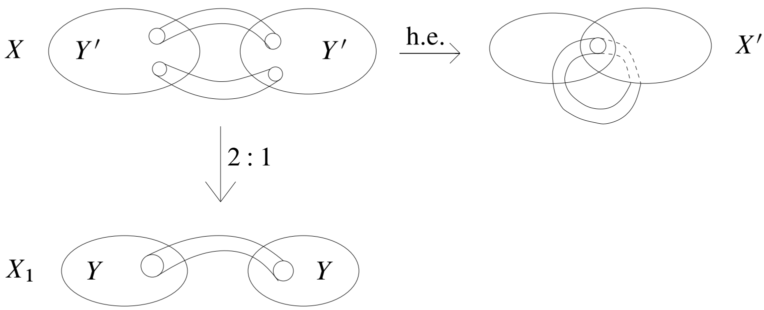


Figure 3

Define the homomorphism h_2: $G(\Gamma_1) \to \mathbb{Z}/2$ by sending w and k to the nontrivial element in $\mathbb{Z}/2$ and other generators to the identity element. Let X be the 2–sheeted cover of X_1 with respect to $\ker(h_2)$. Then X is made of two copies of Y' and two annuli with the boundaries of the annuli identified with the v–circles in Y' (each Y' has two v–circles which cover the v–circle in Y); see Figure 3.

The cover X is homotopy equivalent to a Salvetti complex. To see this, let S_w be the circle in Y' which covers the w–circle in Y two times and let $S_z \vee S_v$ be a wedge of the two circles in Y' which covers the wedge of the z–circle and the v–circle in Y. There is a copy of $S_w \times (S_z \vee S_v)$ inside Y'. Let I be a segment in S_w such that its end points are mapped to the base point of Y under the covering map. We collapse $I \times (S_z \vee S_v)$ to $\{\text{pt}\} \times (S_z \vee S_v)$ inside each copy of Y' in X, and collapse one of the annuli in X to a circle by killing the interval factor. Denote the resulting space by X'. Then X' is homotopy equivalent to X, and the uncollapsed annulus in X becomes a torus in X'. It is not hard to see X' is a Salvetti complex with defining graph Γ_2.

Any standard geodesic in $X(\Gamma_2)$ which comes from vertex u is not Hausdorff close to a quasi-isometric image of some standard geodesic in $X(\Gamma_1)$, since u is not a stable vertex while every vertex in Γ_1 is stable.

Here is a generalization of the above example. Suppose Γ is a finite simplicial graph with vertices $v_1, v_2 \in \Gamma$ such that $d(v_1, v_2) = 2$ and they are separated by the intersection of links $\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2)$. Define a homomorphism h: $G(\Gamma) \to \mathbb{Z}/2$ by sending v_1 and v_2 to the nontrivial element in $\mathbb{Z}/2$ and killing all other generators. Then $\ker(h)$ is also a right-angled Artin group by the same argument as before. To find its defining graph, let $\{C_i\}_{i=1}^n$ be the components of $\Gamma \setminus (\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2))$, and suppose $v_1 \in C_1$. Define $\Gamma_1 = C_1 \cup (\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2))$ and $\Gamma_2 = \left(\bigcup_{i=2}^n C_i\right) \cup (\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2))$. Then Γ_1 and Γ_2 are full subgraphs of Γ; moreover, $\mathrm{St}(v_i) \in C_i$. For $i = 1, 2$, let Γ_i' be the graph obtained by gluing two copies of Γ_i along $\mathrm{St}(v_i)$, and let Γ_3' be the join of one point and $\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2)$. Then the defining graph of $\ker(h)$ can be obtained by gluing Γ_1', Γ_2' and Γ_3' along $\mathrm{lk}(v_1) \cap \mathrm{lk}(v_2)$.

Note that we are taking advantage of separating closed stars while constructing the counterexample. If separating closed stars are not allowed in Γ, then we have a positive answer to question (1); see Section 5.

In the rest of this subsection, we will prove Theorem 3.21. Γ will be an arbitrary finite simplicial graph in the rest of this subsection. Theorem 3.21 is actually a consequence of the following more general result.

Lemma 3.23 *Pick a vertex $w \in \Gamma$, and let Γ_w be the intersection of all stable subgraphs of Γ that contain w. Define $W = \{w' \in \Gamma \mid w^{\perp} \subset \mathrm{St}(w')\}$. Then Γ_w is the full subgraph spanned by W.*

In other words, $G(\Gamma_w) \le G(\Gamma)$ is the minimal standard subgroup containing w with the property that $G(\Gamma_w)$ is invariant under any transvection.

Now we show how to deduce Theorem 3.21 from Lemma 3.23

Proof of Theorem 3.21 The "only if" part can be proved by contradiction (choose a transvection which does not preserve the subgroup $G(\Gamma_1)$). For the converse, let $\{v_i\}_{i=1}^n$ be the vertex set of Γ_1, and let Γ_{v_i} be the minimal stable subgraph that contains v_i for $1 \le i \le n$. By our assumption and Lemma 3.23, $\Gamma_{v_i} \subset \Gamma_1$. Thus the full subgraph spanned by $\bigcup_{i=1}^n \Gamma_{v_i}$ is stable by Lemma 3.18, which means Γ_1 is stable. □

It remains to prove Lemma 3.23. We first set up two auxiliary lemmas.

Lemma 3.24 *Let $v \in \Gamma$ be a vertex which is not isolated. Then at least one of the following is true:*

1. *v is contained in a stable discrete subgraph with more than one vertex.*
2. *v is contained in a stable clique subgraph.*
3. *There is a stable discrete subgraph with more than one vertex whose vertex set is in $v^{\perp}$.*
4. *There is a stable clique subgraph whose vertex set is in $v^{\perp}$.*

Proof Since v is not isolated, we can assume Γ is connected by Lemma 3.11. By Lemma 3.14, we can find a stable subgraph $\Gamma_1 = \bar{\Gamma}_1 \circ \bar{\Gamma}_2 \circ \cdots \circ \bar{\Gamma}_n$ where $\{\bar{\Gamma}_i\}_{i=1}^n$ are discrete full subgraphs and $n = \dim(X(\Gamma))$. If $v \in \Gamma_1$, then by the third paragraph of the proof of Lemma 3.15, we know either (1), (2) or (4) is true.

Suppose $d(v,\Gamma_1)=1$. Let Γ_2 be the full subgraph spanned by $v^\perp\cap\Gamma_1$. Then Γ_2 is stable by Lemma 3.13. The proof of Lemma 3.15 implies every stable subgraph of Γ contains either a stable discrete subgraph or a stable clique subgraph (this does not depend on the $v^\perp\not\subseteq \mathrm{St}(w)$ assumption); thus either (3) or (4) is true.

Suppose $d(v,\Gamma_1)\geq 2$. Pick vertex $u\in\Gamma_1$ such that $d(v,u)=d(v,\Gamma_1)=n$, and let ω be a geodesic connecting v and u. Suppose $\{v_i\}_{i=0}^n$ are the consecutive vertices in ω such that $v_0=v$ and $v_n=u$. Let Γ' be the full subgraph spanned by $v_{n-1}^\perp\cap\Gamma$, and let Γ'' be the full subgraph spanned by V and $V^\perp$, where $V=V_{\Gamma'}$ (the vertex set of Γ'). Then Γ' is stable by Lemma 3.13, and Γ'' is stable by Lemma 3.12. Note that $d(v,x)\geq n$ for any vertex $x\in V$, so $d(v,y)\geq n-1$ for any vertex $y\in V^\perp$. Thus $d(v,\Gamma'')\geq n-1$. However, $v_{n-1}\in\Gamma''$. So $d(v,\Gamma'')=n-1$. Now we can induct on n and reduce to the $d(v,\Gamma_1)=1$ case. □

It is interesting to see that if Γ has large diameter, then there are a lot of nontrivial stable subgraphs.

We record the following lemma which is an easy consequence of Theorem 2.9.

Lemma 3.25 *Suppose $\Gamma=\Gamma_1\circ\Gamma_2$, where Γ_1 is the maximal clique join factor of Γ. If Γ_2' is stable in Γ_2, then $\Gamma_1\circ\Gamma_2'$ is stable in Γ.*

Now we are ready to prove Lemma 3.23.

Proof of Lemma 3.23 By Lemma 3.10, Γ_w is the minimal stable subgraph that contains w. If there exists vertex $w'\in W$ such that $w'\notin\Gamma_w$, then sending $w\to ww'$ and fixing all other vertices would induce a group automorphism, which gives rise to a quasi-isometry from $X(\Gamma)$ to $X(\Gamma)$. The existence of such a quasi-isometry would contradict the stability of Γ_w; thus $W\subset\Gamma_w$.

Let W' be the vertex set of Γ_w. It remains to prove $W'\subset W$. Suppose $W\subsetneq W'$ and let $u\in W'\setminus W$. Then $\varnothing\neq w^\perp\setminus\mathrm{St}(u)$. The minimality of Γ_w implies we cannot use Lemma 3.13 to get rid of u while keeping w; thus $w^\perp\setminus\mathrm{St}(u)\subset W'\setminus\{u,w\}$. In summary,

$$\varnothing\neq w^\perp\setminus\mathrm{St}(u)\subset W'\setminus\{u,w\}. \tag{3-8}$$

In particular, w is not isolated in Γ_w, and

$$\Gamma_w\not\subseteq\mathrm{St}(u). \tag{3-9}$$

Now we apply Lemma 3.24 to Γ_w and w, and recall that if a subgraph is stable in Γ_w, then it is stable in Γ. If case (1) in Lemma 3.24 is true, then we will get a contradiction since w is not isolated in Γ_w. If case (2) is true, then Γ_w sits inside some clique, which is contradictory to (3-9).

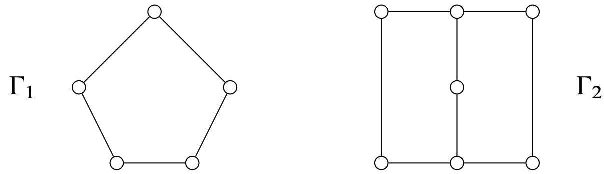


Figure 4: Failure of Theorem 3.21 in the more general case; see Remark 3.26

If case (3) is true, let $\Gamma_1 \subset \Gamma_w$ be the corresponding stable discrete subgraph. Let $V_1 = V_{\Gamma_1}$, and let $V_1' = \{u \in \Gamma_w \mid d(u,v) = 1 \text{ for any } v \in V_1\}$. Suppose Γ_w' is the full subgraph spanned by V_1 and V_1'. Then Γ_w' is stable by Lemma 3.12; hence $\Gamma_w' = \Gamma_w$. Let $\Gamma_w = \overline{\Gamma}_1 \circ \overline{\Gamma}_2 \circ \cdots \circ \overline{\Gamma}_k$ be the join decomposition induced by the De Rahm decomposition of $X(\Gamma_w)$. Then $k \geq 2$ and u does not sit inside the clique factor by (3-9).

If there is no clique factor, then each join factor is stable by Theorem 2.9, and w is inside one of the join factors, which contradicts the minimality of Γ_w. If the clique factor exists and w sits inside the clique factor, then by Theorem 2.9, the clique factor is stable, and we have the same contradiction as before. If the clique factor exists and w sits outside the clique factor, this reduces to the next case.

If case (4) is true, let $\Gamma_2 \subset \Gamma_w$ be the corresponding stable clique subgraph. We can also assume without loss of generality that w is not contained in a stable clique. Let $V_2 = V_{\Gamma_2}$ and $V_2' = \{u \in \Gamma_w \mid d(u,v) = 1 \text{ for any } v \in V_2\}$. Suppose Γ_w'' is the full subgraph spanned by V_2 and V_2'. Then $\Gamma_w'' = \Gamma_w$ as before. Let $\Gamma_w = \Gamma_1' \circ \Gamma_2'$ where Γ_1' corresponds to the Euclidean De Rahm factor of $X(\Gamma_w)$. Note that Γ_2' is nontrivial, and $w, u \in \Gamma_2'$ as in the discussion of case (3). Equation (3-8) implies that $w^\perp \not\subseteq \mathrm{St}(u)$ is still true if we take the orthogonal complement of w and the closed star of u in Γ_2'; in particular, w is not isolated in Γ_2'. Moreover, $\dim(X(\Gamma_2')) < \dim(X(\Gamma_w)) \leq \dim(X(\Gamma))$.

If $\dim(X(\Gamma)) = 2$, then Γ_2' has to be discrete, which is contradictory to the fact that w is not isolated in Γ_2'. If $\dim(X(\Gamma)) = n > 2$, then by induction, we can assume the lemma is true for all lower-dimensional graphs. Then there exists $\overline{\Gamma}_w$ stable in Γ_2' such that $w \in \overline{\Gamma}_w$ and $u \notin \overline{\Gamma}_w$. By Lemma 3.25, $\overline{\Gamma}_w \circ \Gamma_1'$ is stable in Γ_w, hence in Γ, which contradicts the minimality of Γ_w. □

Remark 3.26 It is nature to ask whether Theorem 3.21 is still true if we do not require Γ_1 to be a clique. It turns out there are counterexamples. Let Γ_1 and Γ_2 be as indicated in Figure 4. Then $G(\Gamma_1)$ is quasi-isometric to $G(\Gamma_2)$ by the discussion in Section 11 of [9]. Let $q: X(\Gamma_2) \to X(\Gamma_1)$ be a quasi-isometry, and let K be a standard subcomplex in $X(\Gamma_2)$ such that its defining graph Γ_K is a pentagon in Γ_2. Suppose $q(K)$ is

Hausdorff close to a standard subcomplex K' in $X(\Gamma)$. Then $\Gamma_{K'}$ must be a connected proper subgraph of Γ_1, hence a tree. But this is impossible by the results in [7].

4 From quasi-isometries to isomorphisms

4.1 The extension complexes

4.1.1 Extension complexes and standard flats Let $q\colon X(\Gamma) \to X(\Gamma')$ be a quasi-isometry. Usually q does not induce a well-defined boundary map; see [20]. However, Theorem 3.20 implies we still have control on a subset of the Tits boundaries when $\mathrm{Out}(G(\Gamma))$ and $\mathrm{Out}(G(\Gamma'))$ are transvection-free. In this subsection, we will reorganize this piece of information in terms of extension complexes.

Recall that we identify the vertex set of Γ with a standard generating set S of $G(\Gamma)$. We also label the standard circles in the Salvetti complex by elements in S. By choosing an orientation for each standard circle, we obtain a directed labeling of edges in $X(\Gamma)$.

Denote the extension complex of Γ by $\mathcal{P}(\Gamma)$. We give an alternative definition of $\mathcal{P}(\Gamma)$ here, which is natural for our purposes. The vertices of $\mathcal{P}(\Gamma)$ are in one-to-one correspondence with the parallel classes of standard geodesics in $X(\Gamma)$ (two standard geodesics are in the same parallel class if they are parallel). Two distinct vertices $v_1, v_2 \in \mathcal{P}(\Gamma)$ are connected by an edge if and only if we can find standard geodesics l_i in the parallel classes associated with v_i $(i = 1, 2)$ such that l_1 and l_2 span a standard 2–flat. The next observation follows from Lemmas 3.1 and 2.4:

Observation 4.1 *If $v_1 \neq v_2$, then v_1 and v_2 are joined by an edge if and only if there exist l_i' in the parallel classes associated with v_i $(i = 1, 2)$ and $R > 0$ such that $l_1' \subset N_R(P_{l_2'})$.*

We define $\mathcal{P}(\Gamma)$ to be the flag complex of its 1–skeleton.

Lemma 4.2 *$\mathcal{P}(\Gamma)$ is isomorphic to the extension complex of Γ.*

Proof It suffices to show the 1–skeleton of $\mathcal{P}(\Gamma)$ is isomorphic to the extension graph Γ^e. Pick vertex $v \in \mathcal{P}(\Gamma)$, and let l be a standard geodesic in the parallel class associated with v. We identify l with $\mathbb{R}$ in an orientation-preserving way (the orientation in l is induced by the directed labeling). Recall that $G(\Gamma) \curvearrowright X(\Gamma)$ by deck transformations. Let $\alpha_v \in G(\Gamma)$ be the element such that $\alpha_v(l) = l$ and $\alpha_v(x) = x + 1$ for any $x \in l$. It is easy to see α_v is conjugate to an element in S; thus α_v gives rise to a vertex $\alpha_v \in \Gamma^e$ by Definition 2.11. Note that α_v does not depend the choice of l in the parallel class, so we have a well-defined map from the vertex set of $\mathcal{P}(\Gamma)$ to the vertex set of Γ^e. Moreover, if v_1 and v_2 are adjacent, then α_{v_1} and α_{v_2} commute.

Now we define an inverse map. Pick $\alpha = gsg^{-1} \in \Gamma^e$ $(s \in S)$. Then all standard geodesics which are stabilized by α are in the same parallel class. Let v_α be the vertex in $\mathcal{P}(\Gamma)$ associated with this parallel class. We map the vertex α of Γ^e to the vertex v_α. Now we show this map extends to the 1–skeleton. For $i = 1, 2$, let $\alpha_i = g_i s_i g_i^{-1} \in \Gamma^e$. By the centralizer theorem of [58], α_1 and α_2 commute if and only if $[s_1, s_2] = 1$ and there exists $g \in G(\Gamma)$ such that $\alpha_i = g s_i g^{-1}$. Thus v_{α_1} and v_{α_2} are adjacent in $\mathcal{P}(\Gamma)$. □

Since every edge in the standard geodesics of the same parallel class has the same label, the labeling of the edges of $X(\Gamma)$ induces a labeling of the vertices of $\mathcal{P}(\Gamma)$. Moreover, since $G(\Gamma) \curvearrowright X(\Gamma)$ by label-preserving cubical isomorphisms, we obtain an induced action $G(\Gamma) \curvearrowright \mathcal{P}(\Gamma)$ by label-preserving simplicial isomorphisms. Moreover, the unique label-preserving map from the vertices of $\mathcal{P}(\Gamma)$ to the vertices of $F(\Gamma)$ extends to a simplicial map

$$\pi\colon \mathcal{P}(\Gamma) \to F(\Gamma). \tag{4-1}$$

Pick an arbitrary vertex $p \in X(\Gamma)$; one can obtain a simplicial embedding i_p from the flag complex $F(\Gamma)$ of Γ to $\mathcal{P}(\Gamma)$ by considering the collection of standard geodesics passing through p. We will denote the image of i_p by $(F(\Gamma))_p$. Note that for each vertex $p \in X(\Gamma)$, the composition $\pi \circ i_p\colon F(\Gamma) \to F(\Gamma)$ is the identity map.

Pick a $(k-1)$–simplex in $\mathcal{P}(\Gamma)$ with vertex set $\{v_i\}_{i=1}^k$, and pick a standard geodesic l_i in the parallel class associated with v_i for each $1 \le i \le k$. Since $P_{l_i} \cap P_{l_j} \neq \emptyset$ for $1 \le i \neq j \le k$, by Lemma 2.2, $\bigcap_{i=1}^k P_{l_i} \neq \emptyset$. By Corollary 3.2 and Lemma 3.4, there exist standard geodesics $\{l_i'\}_{i=1}^k$ satisfying:

1. (1) l_i' is parallel to l_i for each i.
2. (2) The convex hull of $\{l_i'\}_{i=1}^k$ is a standard k–flat denoted by F_k.
3. (3) $\bigcap_{i=1}^k P_{l_i} = P_{F_k}$.

Thus we have a one-to-one correspondence between the $(k-1)$–simplexes of $\mathcal{P}(\Gamma)$ and parallel classes of standard k–flats in $X(\Gamma)$. In particular, maximal simplexes in $\mathcal{P}(\Gamma)$, namely those simplexes which are not properly contained in some larger simplexes of $\mathcal{P}(\Gamma)$, are in one-to-one correspondence with maximal standard flats in $X(\Gamma)$. For standard flat $F \subset X(\Gamma)$, we denote the simplex in $\mathcal{P}(\Gamma)$ associated with the parallel class containing F by $\Delta(F)$.

Observation 4.3 *Let Δ_1, Δ_2 be two simplexes in $\mathcal{P}(\Gamma)$ such that $\Delta = \Delta_1 \cap \Delta_2 \neq \emptyset$. For $i = 1, 2$, let $F_i \subset X(\Gamma)$ be a standard flat such that $\Delta(F_i) = \Delta_i$. Set $(F_1', F_2') = \mathcal{I}(F_1, F_2)$. Then $\Delta(F_1') = \Delta(F_2') = \Delta$.*

We define the *reduced Tits boundary*, denoted $\overline{\partial}_T(X(\Gamma))$, to be the subset of $\partial_T(X(\Gamma))$ which is the union of Tits boundaries of standard flats in $X(\Gamma)$. For a standard flat $F \subset X(\Gamma)$, we triangulate $\partial_T F$ into all-right spherical simplexes which are the Tits boundaries of orthant subcomplexes in F. Pick another standard flat $F' \subset X(\Gamma)$; then $\partial_T F \cap \partial_T F'$ is a subcomplex in both $\partial_T F$ and $\partial_T F'$ by Lemma 3.1 and Remark 2.5. Thus we can endow $\overline{\partial}_T(X(\Gamma))$ with the structure of an all-right spherical complex.

Now we look at the relation between $\overline{\partial}_T(X(\Gamma))$ and $\mathcal{P}(\Gamma)$. For each standard flat $F \subset X(\Gamma)$, we can associate $\partial_T F$ with $\Delta(F) \subset \mathcal{P}(\Gamma)$. This induces a surjective simplicial map $s\colon \overline{\partial}_T(X(\Gamma)) \to \mathcal{P}(\Gamma)$ (s can be defined by induction on dimension). Note that the inverse image of each simplex in $\mathcal{P}(\Gamma)$ under s is a sphere in $\overline{\partial}_T(X(\Gamma))$. Then one can construct $\overline{\partial}_T(X(\Gamma))$ from $\mathcal{P}(\Gamma)$ as follows. We start with a collection of the $\mathbb{S}^0$ which are in one-to-one correspondence to vertices of $\mathcal{P}(\Gamma)$ and form a join of n copies of the $\mathbb{S}^0$ if and only if the corresponding n vertices in $\mathcal{P}(\Gamma)$ span an $(n-1)$–simplex. In other words, $\overline{\partial}_T(X(\Gamma))$ is obtained by applying the spherical complex construction in the sense of [12, Definition 2.1.22] to $\mathcal{P}(\Gamma)$.

Let $K_1 \subset X(\Gamma)$ be a standard subcomplex. We define $\overline{\partial}_T(K_1)$ to be the union of Tits boundaries of standard flats in K_1. Note that $\overline{\partial}_T(K_1) = \overline{\partial}_T(X(\Gamma)) \cap \partial_T K_1$, and it descends to a subcomplex in $\mathcal{P}(\Gamma)$, which will be denoted by $\Delta(K_1)$.

Lemma 4.4 *Let K_1 and K_2 be two standard subcomplexes of $X(\Gamma)$. Put $(K_1', K_2') = \mathcal{I}(K_1, K_2)$. Then $\Delta(K_1') = \Delta(K_2') = \Delta(K_1) \cap \Delta(K_2)$.*

Proof By Remark 2.5, we have $\partial_T K_1' = \partial_T K_2' = \partial_T K_1 \cap \partial_T K_2$; hence $\overline{\partial}_T K_1' = \overline{\partial}_T K_2' = \overline{\partial}_T K_1 \cap \overline{\partial}_T K_2$ and $\Delta(K_1') = \Delta(K_2') = \Delta(K_1) \cap \Delta(K_2)$. □

Now we study how the extension complexes behave under quasi-isometries.

Lemma 4.5 *Pick Γ_1 and Γ_2 such that $\mathrm{Out}(G(\Gamma_i))$ is transvection-free for $i = 1, 2$. Then any quasi-isometry $q\colon X(\Gamma_1) \to X(\Gamma_2)$ induces a simplicial isomorphism $q_*\colon \mathcal{P}(\Gamma_1) \to \mathcal{P}(\Gamma_2)$. If only $\mathrm{Out}(G(\Gamma_1))$ is assumed to be transvection-free, we still have a simplicial embedding $q_*\colon \mathcal{P}(\Gamma_1) \to \mathcal{P}(\Gamma_2)$.*

Proof We prove the case when both $\mathrm{Out}(G(\Gamma_1))$ and $\mathrm{Out}(G(\Gamma_2))$ are transvection-free. The other case is similar. By Theorem 3.20, every vertex in Γ_1 is stable; thus q sends any parallel class of standard geodesics in $X(\Gamma_1)$ to another parallel class of standard geodesics in $X(\Gamma_2)$ up to finite Hausdorff distance. This induces a well-defined map q_* from the 0–skeleton of $\mathcal{P}(\Gamma_1)$ to the 0–skeleton of $\mathcal{P}(\Gamma_2)$. The map q_* is a bijection by considering the quasi-isometry inverse. Moreover, Observation 4.1

implies two vertices in $\mathcal{P}(\Gamma_1)$ are adjacent if and only if their images under q_* are adjacent. So we can extend q_* to be a graph isomorphism between the 1–skeleton of $\mathcal{P}(\Gamma_1)$ and the 1–skeleton of $\mathcal{P}(\Gamma_2)$. Since both $\mathcal{P}(\Gamma_1)$ and $\mathcal{P}(\Gamma_2)$ are flag complexes, we can extend the isomorphism to the whole complex. □

4.1.2 Extension complexes and their relatives Now we discuss the relation between $\mathcal{P}(\Gamma)$ with several other objects in the literature. The material in this subsection will not be used later.

We can endow $F(\Gamma)$ with the structure of complex of groups, which gives us an alternative definition of $\mathcal{P}(\Gamma)$. More specifically, $\mathcal{P}(\Gamma) = F(\Gamma)\times G(\Gamma)/\sim$; here $\mathrm{St}(v)\times g_1$ and $\mathrm{St}(v)\times g_2$ ($v \in F(\Gamma)$ is a vertex) are identified if and only if there exists an integer m such that $g_1^{-1}g_2 = v^m$ (we also view v as one of the generators of $G(\Gamma)$). Hence for k–simplex $\Delta^k \subset F(\Gamma)$ with vertex set $\{v_i\}_{i=1}^k$, we have that $\mathrm{St}(\Delta^k)\times g_1$ and $\mathrm{St}(\Delta^k)\times g_2$ are identified if and only if $g_1^{-1}g_2$ belongs to the $\mathbb{Z}^k$ subgroup of $G(\Gamma)$ generated by $\{v_i\}_{i=1}^k$. One can compare this with a similar construction for a Coxeter group in [21].

There is another important object associated with a right-angled Artin group, called the modified Deligne complex in [17] and the *flat space* in [9].

Definition 4.6 Let $\mathbb{P}(\Gamma)$ be poset of left cosets of standard abelian subgroups of $G(\Gamma)$ (including the trivial subgroup) such that the partial order is induced by inclusion of sets. Then the *modified Deligne complex* is defined to be the geometric realization of the derived poset of $\mathbb{P}(\Gamma)$.

Recall that elements in the *derived poset* of a poset $\mathbb{P}$ are totally ordered finite chains in $\mathbb{P}$. It can be viewed as an abstract simplex.

The extension complex $\mathcal{P}(\Gamma)$ can be viewed as a coarse version of the modified Deligne complex. Let A and B be two subsets of a metric space. We say A and B are *coarsely equivalent* if $A \overset{\infty}{=} B$, and A is *coarsely contained* in B if $A \subset_\infty B$. Let $\mathbb{P}'(\Gamma)$ be the poset whose elements are coarsely equivalent classes of left cosets of nontrivial standard abelian subgroups of $G(\Gamma)$, and the partial order is induced by coarse inclusion of sets.

Observation 4.7 *The poset $\mathbb{P}'(\Gamma)$ is an abstract simplicial complex, and it is isomorphic to $\mathcal{P}(\Gamma)$.*

Roughly speaking, $\mathbb{P}(\Gamma)$ captures the combinatorial pattern of how standard flats in $X(\Gamma)$ intersect with each other, and $\mathcal{P}(\Gamma)$ is about how they coarsely intersect with each other; thus $\mathbb{P}(\Gamma)$ contains more information than $\mathcal{P}(\Gamma)$. However, in certain cases, it is possible to recover information about $\mathbb{P}(\Gamma)$ from $\mathcal{P}(\Gamma)$, and this enable us to prove quasi-isometric classification/rigidity results for RAAGs.

We can define the poset $\mathbb{P}'(\Gamma)$ for an arbitrary Artin group by considering the collection of coarse equivalent classes of spherical subgroups in an Artin group under coarse inclusion. Then the geometric realization of the derived poset of $\mathbb{P}'(\Gamma)$ would be a natural candidate for the extension complex of an Artin group. It is interesting to ask how much of the results in [43] can be generalized to this context.

There is also a link between $\mathcal{P}(\Gamma)$ and the structure of hyperplanes in $X(\Gamma)$. Recall that for every CAT(0) cube complex X, the *crossing graph* of X, denoted by $C(X)$, is a graph whose vertices are in one-to-one correspondence to the hyperplanes in X, and two vertices are adjacent if and only if the corresponding hyperplanes intersect. The *contact graph*, introduced in [30] and denoted by $\mathcal{C}(X)$, has the same vertex set as $C(X)$, and two vertices are adjacent if and only if the carriers of the corresponding hyperplanes intersect.

There is a natural surjective simplicial map $p\colon C(X(\Gamma)) \to \Gamma^e$ defined as follows. Pick a vertex $v \in C(X(\Gamma))$ and let h be the corresponding hyperplane. Since all standard geodesics which intersect h at one point are in the same parallel class, we define $p(v)$ to be the vertex in Γ^e associated with this parallel class; see Lemma 4.2. It is clear that if $v_1, v_2 \in C(X(\Gamma))$ are adjacent vertices, then $p(v_1)$ and $p(v_2)$ are adjacent, so p extends to a simplicial map. Pick a vertex $w \in \Gamma^e$; then $p^{-1}(e)$ is the collection of hyperplanes dual to a standard geodesic.

Theorem 4.8 [42; 30] *If Γ is connected, then $C(X(\Gamma))$, $\mathcal{C}(X(\Gamma))$ and $\mathcal{P}(\Gamma)$ are quasi-isometric to each other; moreover, they are quasi-isometric to a tree.*

From this viewpoint, $\mathcal{P}(\Gamma)$ captures both the geometric information of $X(\Gamma)$ (the standard flats) and the combinatorial information (the hyperplanes).

4.2 Reconstruction of quasi-isometries

We show the boundary map $q_*\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$ in Lemma 4.5 induces a well-defined map from $G(\Gamma)$ to $G(\Gamma')$.

Lemma 4.9 *Let F_1 and F_2 be two maximal standard flats in $X(\Gamma)$ and let Δ_1 and Δ_2 be the corresponding maximal simplexes in $\mathcal{P}(\Gamma)$. If F_1 and F_2 are separated by a hyperplane h, then there exist vertices $v_i \in \Delta_i$ for $i = 1, 2$ and $v \in \mathcal{P}(\Gamma)$ such that v_1 and v_2 are in different connected components of $\mathcal{P}(\Gamma) \setminus \mathrm{St}(v)$.*

Proof Let e be an edge dual to h and let l be the standard geodesic that contains e. Set $v = \Delta(l) \in \mathcal{P}(\Gamma)$. By Lemma 3.4, the parallel set P_l of l is isometric to $h \times \mathbb{E}^1$. Thus every standard geodesic parallel to l must have nontrivial intersection with h. Since $F_1 \cap h = \varnothing$, we see that F_1 cannot contain any standard geodesic parallel to l, which means $v \notin \Delta_1$. Moreover, $\Delta_1 \not\subseteq \mathrm{St}(v)$ since Δ_1 is a maximal simplex. Similarly, $\Delta_2 \not\subseteq \mathrm{St}(v)$; thus we can find vertices $v_i \in \Delta_i \setminus \mathrm{St}(v)$ for $i = 1, 2$. We claim v_1, v_2 and v are the vertices we are looking for.

If there is a path $\omega \subset \mathcal{P}(\Gamma) \setminus \mathrm{St}(v)$ connecting v_1 and v_2, we can assume ω consists of a sequence of edges $\{e_i\}_{i=1}^k$ with $v_1 \in e_1$ and $v_2 \in e_k$. For each e_i, pick a maximal simplex Δ'_i that contains e_i, and let F'_i be the maximal standard flat such that $\Delta(F'_i) = \Delta'_i$. Then $v \notin \Delta'_i$ for $1 \le i \le k$; hence $F'_i \cap h = \varnothing$.

Set $\Delta'_0 = \Delta_1$, $\Delta'_{k+1} = \Delta_2$, $F'_0 = F_1$ and $F'_{k+1} = F_2$. Since $\Delta'_i \cap \Delta'_{i+1}$ contains a vertex in ω, we have

$$(\Delta'_i \cap \Delta'_{i+1}) \setminus \mathrm{St}(v) \neq \varnothing \tag{4-2}$$

for $0 \le i \le k$. Since F'_0 and F'_{k+1} are in different sides of h, there exists i_0 such that h separates F'_{i_0} and F'_{i_0+1}. Let $(F''_{i_0}, F''_{i_0+1}) = \mathcal{I}(F'_{i_0}, F'_{i_0+1})$. By Observation 4.3, $\Delta(F''_{i_0}) = \Delta(F''_{i_0+1}) = \Delta'_i \cap \Delta'_{i+1}$. However, by Lemma 2.7, there exists a convex subset of h parallel to F''_{i_0}; thus $F''_{i_0} \subset_\infty h \subset P_l$. It follows from Observation 4.1 that $\Delta'_i \cap \Delta'_{i+1} \subset \mathrm{St}(v)$, which contradicts (4-2). □

Denote the Cayley graph of $G(\Gamma)$ with respect to the standard generating set S by $C(\Gamma)$. We pick an identification between $C(\Gamma)$ and the 1–skeleton of $X(\Gamma)$. Thus $G(\Gamma)$ is identified with the vertex set of $X(\Gamma)$.

Lemma 4.10 *Let Γ_1 be a simple graph such that:*

1. *There is no separating closed star in $F(\Gamma_1)$.*
2. *$F(\Gamma_1)$ is not contained in a union of two closed stars.*

Then any simplicial isomorphism $s\colon \mathcal{P}(\Gamma_1) \to \mathcal{P}(\Gamma_2)$ induces a unique map $s'\colon G(\Gamma_1) \to G(\Gamma_2)$ such that for any maximal standard flat $F_1 \subset X(\Gamma_1)$, vertices in F_1 are mapped by s' to vertices lying in a maximal standard flat $F_2 \subset X(\Gamma_2)$ with $\Delta(F_2) = s'(\Delta(F_1))$.

Proof Pick a vertex $p \in G(\Gamma_1)$. Let $\{F_i\}_{i=1}^k$ be the collection of maximal standard flats containing p. For $1 \le i \le k$, define $\Delta_i = \Delta(F_i)$ and $\Delta'_i = s(\Delta_i)$. Let $F'_i \subset X(\Gamma_2)$ be the maximal standard flat such that $\Delta(F'_i) = \Delta'_i$. Let $K_p = (F(\Gamma_1))_p = \bigcup_{i=1}^k \Delta_i$ (recall that $K_p \cong F(\Gamma_1)$). We claim

$$\bigcap_{i=1}^k F'_i \neq \varnothing. \tag{4-3}$$

The lemma will then follow from (4-3). To see this, we deduce from condition (2) that $\bigcap_{i=1}^k \Delta_i = \varnothing$. Hence $\bigcap_{i=1}^k F_i = \{p\}$. It follows that $\bigcap_{i=1}^k \Delta'_i = \varnothing$. This together with (4-3) imply that $\bigcap_{i=1}^{k} F'_i$ is exactly one point. We define s' by sending p to this point. One readily verifies that s' has the required properties.

It remains to prove (4-3).

Suppose that (4-3) is not true. Then by Lemma 2.2, there exist i_1 and i_2 such that $F'_{i_1} \cap F'_{i_2} = \varnothing$. Thus F'_{i_1} and F'_{i_2} are separated by a hyperplane. It follows from Lemma 4.9 that there exist vertices $v' \in \mathcal{P}(\Gamma_2)$, $v'_1 \in \Delta'_{i_1}$ and $v'_2 \in \Delta'_{i_2}$ such that v'_1 and v'_2 are in different connected components of $\mathcal{P}(\Gamma_2) \setminus \mathrm{St}(v')$. Let $v = s^{-1}(v')$, $v_1 = s^{-1}(v'_1)$ and $v_2 = s^{-1}(v'_2)$. Then $K_p \setminus (K_p \cap \mathrm{St}(v))$ is disconnected (since $v_1, v_2 \in K_p$ and they are separated by $\mathrm{St}(v)$).

If $v \in K_p$, then K_p would contain a separating closed star, which yields a contradiction; thus (4-3) is true in this case.

Suppose $v \notin K_p$. Pick a standard geodesic l such that $\Delta(l) = v$ and let $\{h_i\}_{i=1}^{n}$ be the collection of hyperplanes in $X(\Gamma)$ such that each h_i separates p from the parallel set P_l of l (note that $p \notin P_l$). For $1 \le i \le n$, pick an edge e_i dual to h_i and let w_i be the unique vertex in K_p that has the same label as e_i. Let $w_0 \in K_p$ be the unique vertex which has the same label as v. We claim

$$\mathrm{St}(v) \cap K_p = \bigcap_{i=0}^{n} (\mathrm{St}(w_i) \cap K_p). \tag{4-4}$$

For every $u \in K_p$, let l_u be the unique standard geodesic such that $\Delta(l_u) = u$ and $p \in l_u$.

Pick $u \in \mathrm{St}(v) \cap K_p$. Observation 4.1 implies $\mathcal{I}(l_u, P_l) = (l_u, l'_u)$, where l'_u is some standard geodesic in P_l. Then for $1 \le i \le n$, the hyperplane h_i separates l_u from P_l, otherwise $h_i \cap l_u \neq \varnothing$ and Lemma 2.6 implies $h_i \cap P_l \neq \varnothing$, which is a contradiction. It follows from Corollary 3.2 that u and w_i are adjacent for $0 \le i \le n$; thus $u \in \bigcap_{i=0}^{n} (\mathrm{St}(w_i) \cap K_p)$. Therefore, $\mathrm{St}(v) \cap K_p \subset \bigcap_{i=0}^{n} (\mathrm{St}(w_i) \cap K_p)$.

Pick $u \in \bigcap_{i=0}^{n} (\mathrm{St}(w_i) \cap K_p)$. First we show $l_u \cap P_l = \varnothing$. Suppose there is a vertex z in $l_u \cap P_l$. Since v and w_0 have the same label and $u \in \mathrm{St}(w_0)$, it follows that the edge in l_u which contains z belongs to the parallel set P_l. Then $l_u \subset P_l$, contradicting the fact that $p \notin P_l$. Therefore, $l_u \cap P_l = \varnothing$.

Now we pick an edge path ω of shortest combinatorial length that travels from l_u to P_l. Let $\{f_j\}_{j=1}^{m}$ be the consecutive edges in ω such that $f_1 \cap l_u \neq \varnothing$. For each f_j, let $\bar{h}_j$ be the hyperplane dual to f_j. Then $\bar{h}_j$ separates l_u from P_l (otherwise ω would not be the shortest edge path), hence separates p from P_l. This and $u \in \bigcap_{i=0}^{n} (\mathrm{St}(w_i) \cap K_p)$ imply that $d(\pi(u), V_{f_j}) \le 1$ for each j, where π is the map in (4-1) and V_{f_j} is the label of the edge f_j. It follows that ω is contained in the parallel set P_{l_u}, and hence the intersection $P_{l_u} \cap P_l$ contains some vertex z. Again, since $u \in \mathrm{St}(w_0)$, and since w_0 has the same label as v, we find that the standard geodesic $l'_u \subset P_{l_u}$ that is parallel to l_u and passes through z is contained in P_l. Therefore, $u \in \mathrm{St}(v) \cap K_p$, and (4-4) follows.

By condition (2) of Lemma 4.10, we have

$$(\mathrm{St}(w_0) \cap K_p) \cup \left(\bigcap_{i=1}^{n} (\mathrm{St}(w_i) \cap K_p) \right) \subsetneq K_p. \tag{4-5}$$

Let $A = K_p \setminus (\mathrm{St}(w_0) \cap K_p)$, and let $B = K_p \setminus \big(\bigcap_{i=1}^{n} (\mathrm{St}(w_i) \cap K_p) \big)$. Then (4-5) implies $A \cap B \neq \varnothing$. Thus we have the following Mayer–Vietoris sequence for reduced homology:

$$\cdots \to \widetilde{H}_0(A \cap B) \to \widetilde{H}_0(A) \oplus \widetilde{H}_0(B) \to \widetilde{H}_0(A \cup B) \to 0.$$

Recall that $K_p \setminus (K_p \cap \mathrm{St}(v))$ is disconnected, we deduce that $\widetilde{H}_0(A \cup B)$ is nontrivial from (4-4). Thus $\widetilde{H}_0(A) \oplus \widetilde{H}_0(B)$ is nontrivial, which implies that either $\bigcap_{i=1}^{n} (\mathrm{St}(w_i) \cap K_p)$ or $\mathrm{St}(w_0) \cap K_p$ would separate K_p. Thus we can induct on n to deduce that there exists i_0 such that $\mathrm{St}(w_{i_0}) \cap K_p$ separates K_p. This yields a contradiction to condition (1) of Lemma 4.10. □

There are counterexamples if we only assume (1) in Lemma 4.10. For example, let Γ_1 and Γ_2 be discrete graphs made of two points. Then $\mathcal{P}(\Gamma_1)$ and $\mathcal{P}(\Gamma_2)$ are discrete sets. Now it is not hard to construct a permutation of a discrete set to itself which does not satisfy the conclusion of Lemma 4.10. If we go back to the proof of Lemma 4.10, then the step using the Mayer–Vietoris sequence will fail, since we need $A \cap B \neq \varnothing$ in order to use the reduced version of Mayer–Vietoris sequence.

Corollary 4.11 *Suppose that $G(\Gamma_1)$ and $G(\Gamma_2)$ both satisfy the assumption of Lemma 4.10. Then they are isomorphic if and only if $\mathcal{P}(\Gamma_1)$ and $\mathcal{P}(\Gamma_2)$ are isomorphic as simplicial complexes.*

Proof The "only if" direction follows from the fact that $G(\Gamma_1)$ and $G(\Gamma_2)$ are isomorphic if and only if Γ_1 and Γ_2 are isomorphic; see [23]. It remains to prove the "if" direction. Pick an isomorphism $s\colon \mathcal{P}(\Gamma_1) \to \mathcal{P}(\Gamma_2)$, and let $s'\colon G(\Gamma_1) \to G(\Gamma_2)$ be the map in Lemma 4.10. Pick a vertex $p \in G(\Gamma_1)$ and let $q = s(p)$. We define $(F(\Gamma_1))_p \subset \mathcal{P}(\Gamma_1)$ and $(F(\Gamma_2))_q \subset \mathcal{P}(\Gamma_2)$ as in the first paragraph of the proof of Lemma 4.10. Then (4-3) implies $s((F(\Gamma_1))_p) \subset (F(\Gamma_2))_q$. This induces a graph embedding $\Gamma_1 \hookrightarrow \Gamma_2$. By repeating the previous discussion for s^{-1}, we obtain another graph embedding $\Gamma_2 \hookrightarrow \Gamma_1$. Since both Γ_1 and Γ_2 are finite simplicial graphs, they are isomorphic. Hence $G(\Gamma_1) \cong G(\Gamma_2)$. □

Lemma 4.12 *Let $G(\Gamma)$ be a RAAG such that* $\mathrm{Out}(G(\Gamma))$ *is finite and $G(\Gamma) \ncong \mathbb{Z}$. Then $F(\Gamma)$ satisfies the assumption of Lemma 4.10.*

Proof It is clear that $F(\Gamma)$ should satisfy condition (1) of Lemma 4.10 since no nontrivial partial conjugation is allowed. If $F(\Gamma)$ is contained in a closed star, then Γ is a point. So if (2) is not true, then $F(\Gamma) = \mathrm{St}(v) \cup \mathrm{St}(w)$ for distinct vertices $v, w \in \Gamma$. Since the orthogonal complement $v^{\perp}$ satisfies $v^{\perp} \not\subseteq \mathrm{St}(w)$, there exists $u \in v^{\perp}$ such that $d(u, w) \geq 2$. Pick any edge e such that $u \in e$; then $e \not\subseteq \mathrm{St}(w)$, and so $e \subset \mathrm{St}(v)$. This implies $u^{\perp} \subset \mathrm{St}(v)$; hence $\mathrm{Out}(G(\Gamma))$ is infinite, which yields a contradiction. □

By Lemma 4.5, Lemma 4.12 and Corollary 4.11, we have following result, which in particular establishes Theorem 1.1 of the introduction.

Theorem 4.13 *Let Γ_1 and Γ_2 be two finite simplicial graphs such that* $\mathrm{Out}(G(\Gamma_i))$ *is finite for $i = 1, 2$. Then $G(\Gamma_1)$ and $G(\Gamma_2)$ are quasi-isometric if and only if they are isomorphic. Moreover, for any (L, A)–quasi-isometry $q\colon X(\Gamma_1) \to X(\Gamma_2)$, there exist a bijection $q'\colon G(\Gamma_1) \to G(\Gamma_2)$ and a constant $D = D(L, A, \Gamma_1)$ such that:*

(1) *$d(q(v), q'(v)) < D$ for any $v \in G(\Gamma_1)$.*

(2) *For any standard flat $F_1 \subset X(\Gamma_1)$, there exists a standard flat $F_2 \subset X(\Gamma_2)$ such that q' induces a bijection between $F_1 \cap G(\Gamma_1)$ and $F_2 \cap G(\Gamma_2)$.*

If $G(\Gamma_1) \neq \mathbb{Z}$, then such a q' is unique.

Proof It suffices to look at the case where $G(\Gamma_1) \neq \mathbb{Z}$. Then $G(\Gamma_2) \neq \mathbb{Z}$. In this case, every vertex v in Γ_1 or Γ_2 is the intersection of maximal cliques that contain v (otherwise there exists a vertex w such that $w \neq v$ and $v^{\perp} \subset \mathrm{St}(w)$). It follows that every standard geodesic in $X(\Gamma_1)$ or $X(\Gamma_2)$ is the intersection of finitely many maximal standard flats, and so is every standard flat. Let $q_*\colon \mathcal{P}(\Gamma_1) \to \mathcal{P}(\Gamma_2)$ be the map in Lemma 4.5. We apply Lemma 4.10 to q_* and q_*^{-1} to obtain q' with the required properties. Note that each vertex of $X(\Gamma)$ is the intersection of maximal standard flats that contain it; thus q' is unique. □

4.3 The automorphism groups of extension complexes

Suppose $\mathrm{Out}(G(\Gamma))$ is finite; by Theorem 4.13, each element in the simplicial automorphism group $\mathrm{Aut}(\mathcal{P}(\Gamma))$ of $\mathcal{P}(\Gamma)$ induces a bijection $G(\Gamma) \to G(\Gamma)$. However, this bijection does not extend to an isomorphism from $X(\Gamma)$ to itself in general. We start by looking at the following example which was first pointed out in [9, Section 11] in a slightly different form.

Example 4.14 Let $l \subset X(\Gamma)$ be a standard geodesic, and let $\pi_l\colon X(\Gamma) \to l$ be the CAT(0) projection. We identify the vertex set of l with $\mathbb{Z}$. Let $X^{(0)}(\Gamma)$ be the vertex set of $X(\Gamma)$. Then the above projection induces a map $\pi_l\colon X^{(0)}(\Gamma) \to \mathbb{Z}$.

Recall that each edge of $X(\Gamma)$ is oriented and labeled, and $G(\Gamma)$ acts on $X(\Gamma)$ by transformations that preserve labels and orientations. There is a unique element $\alpha \in G(\Gamma)$ such that α translates l one unit in the positive direction.

We want to define a bijection $q\colon X^{(0)}(\Gamma) \to X^{(0)}(\Gamma)$ which basically flips $\pi_l^{-1}(0)$ and $\pi_l^{-1}(1)$. More precisely,

$$q(x) = \begin{cases} x & \text{if } \pi_l(x) \neq 0, 1, \\ \alpha(x) & \text{if } x \in \pi_l^{-1}(0), \\ \alpha^{-1}(x) & \text{if } x \in \pi_l^{-1}(1). \end{cases}$$

One can check the following:

(1) q is a quasi-isometry.

(2) q does not respect the word metric.

(3) q maps vertices in a standard flat to vertices in another standard flat. Thus q induces an element in $\mathrm{Aut}(\mathcal{P}(\Gamma))$.

The above example implies that, in general, elements in $\mathrm{Aut}(\mathcal{P}(\Gamma))$ do not respect the order along the standard geodesics of $X(\Gamma)$. There is another metric on $G(\Gamma)$ which "forgets about" the ordering. Following [43], we define the *syllable length* of a word ω to be the minimal l such that ω can be written as a product of l elements of the form $v_i^{k_i}$, where v_i is a standard generator and k_i is an integer.

An alternative definition is the following. Let $\{h_i\}_{i=1}^k$ be the collection of hyperplanes separating $\omega \in G(\Gamma)$ and the identity element (recall that we have identified $G(\Gamma)$ with the 0–skeleton of $X(\Gamma)$). For each i, pick a standard geodesic l_i dual to h_i. Then the syllable length of ω is the number of elements in $\{\Delta(l_i)\}_{i=1}^k$. The syllable length induces a left invariant metric on $G(\Gamma)$, which will be denoted by d_r. Note that the map in Example 4.14 is an isometry with respect to d_r.

Denote the word metric on $G(\Gamma)$ with respect to the standard generators by d_w.

Corollary 4.15 *Let Γ be a graph such that $\mathrm{Out}(G(\Gamma))$ is finite, and denote the simplicial automorphism group of $\mathcal{P}(\Gamma)$ by $\mathrm{Aut}(\mathcal{P}(\Gamma))$. Then*

$$\mathrm{Aut}(\mathcal{P}(\Gamma)) \cong \mathrm{Isom}(G(\Gamma), d_r).$$

Proof Let $\mathrm{Perm}(G(\Gamma))$ be the permutation group of elements in $G(\Gamma)$. We have a group homomorphism $h_1\colon \mathrm{Aut}(\mathcal{P}(\Gamma)) \to \mathrm{Perm}(G(\Gamma))$ by Lemma 4.10. Take $\phi \in \mathrm{Aut}(\mathcal{P}(\Gamma))$; by Lemma 4.12, $\varphi = h_1(\phi)$ and $\varphi^{-1} = h_1(\phi^{-1})$ satisfy the conclusion of Lemma 4.10. Since every standard geodesic is the intersection of finitely many maximal standard flats, points in a standard geodesic are mapped to points in a standard geodesic by ϕ, which implies $d_r(\varphi(v_1),\varphi(v_2)) \le d_r(v_1,v_2)$ if $d_r(v_1,v_2)\le 1$. By the triangle inequality, we have $d_r(\varphi(v_1),\varphi(v_2)) \le d_r(v_1,v_2)$ for any $v_1,v_2\in G(\Gamma)$. Similarly, $d_r(\varphi^{-1}(v_1),\varphi^{-1}(v_2)) \le d_r(v_1,v_2)$. Thus $\varphi\in \mathrm{Isom}(G(\Gamma),d_r)$, and we have a homomorphism $h_1\colon \mathrm{Aut}(\mathcal{P}(\Gamma))\to \mathrm{Isom}(G(\Gamma),d_r)$.

Now pick $\varphi\in \mathrm{Isom}(G(\Gamma),d_r)$. Let $v_1,v_2,v_3\in G(\Gamma)$ such that $d_r(v_1,v_i)=1$ for $i=2,3$. We claim

$$\angle_{v_1}(v_2,v_3)=\tfrac{\pi}{2} \iff \angle_{\varphi(v_1)}(\varphi(v_2),\varphi(v_3))=\tfrac{\pi}{2}. \tag{4-6}$$

If $\angle_{v_1}(v_2,v_3)=\frac{\pi}{2}$, then we can find $v_4\in G(\Gamma)$ such that $\{v_i\}_{i=1}^4$ are the vertices of a flat rectangle in $X(\Gamma)$. Note that

$$d_r(v_1,v_4)=d_r(v_2,v_3)=2 \quad\text{and}\quad d_r(v_4,v_2)=d_r(v_4,v_3)=1,$$

so

$$d_r(\varphi(v_1),\varphi(v_4))=d_r(\varphi(v_2),\varphi(v_3))=2 \quad\text{and}\quad d_r(\varphi(v_4),\varphi(v_2))=d_r(\varphi(v_4),\varphi(v_3))=1.$$

Now we consider the 4–gon formed by $\overline{\varphi(v_1)\varphi(v_2)}$, $\overline{\varphi(v_2)\varphi(v_4)}$, $\overline{\varphi(v_4)\varphi(v_3)}$ and $\overline{\varphi(v_3)\varphi(v_1)}$. Then the angles at the four vertices of this 4–gon are bigger or equal to $\frac{\pi}{2}$. It follows from CAT(0) geometry that the angles are exactly $\frac{\pi}{2}$ and the 4–gon actually bounds a flat rectangle. Thus one direction of (4-6) is proved; the other direction is similar.

We need another observation as follows. If three points $v_1,v_2,v_3\in G(\Gamma)$ satisfy $d_r(v_i,v_j)=1$ for $1\le i\ne j\le 3$, then the angle at each vertex of the triangle $\Delta(v_1,v_2,v_3)$ could only be 0 or π; thus $\{v_i\}_{i=1}^3$ are inside a standard geodesic. It follows from this observation that points in a standard geodesic are mapped by φ to points in a standard geodesic.

We define $\phi\colon \mathcal{P}(\Gamma)\to\mathcal{P}(\Gamma)$ as follows. For vertex $w\in\mathcal{P}(\Gamma)$, let l be a standard geodesic such that $\Delta(l)=w$. Suppose $l'\subset X(\Gamma)$ is the standard geodesic such that $\phi(v(l))\subset l'$ ($v(l)$ denotes the vertex set of l). Suppose $w'=\Delta(l')$. We define $w'=\phi(w)$; (4-6) implies w' does not depend on the choice of l, and $\phi(w_1)$ and $\phi(w_2)$ are adjacent if vertices $w_1,w_2\in\mathcal{P}(\Gamma)$ are adjacent. Thus ϕ is a well-defined simplicial map. Note that φ^{-1} also induces a simplicial map from $\mathcal{P}(\Gamma)$ to itself in a similar way, so $\phi\in\mathrm{Aut}(\mathcal{P}(\Gamma))$. We define $\phi=h_2(\varphi)$. One readily verifies that $h_2\colon \mathrm{Isom}(G(\Gamma),d_r)\to\mathrm{Aut}(\mathcal{P}(\Gamma))$ is a group homomorphism, and $h_2\circ h_1=h_1\circ h_2=\mathrm{Id}$. Thus the corollary follows. ☐

Remark 4.16 If we drop the assumption in the above corollary about Γ, then there is still a monomorphism $h\colon \mathrm{Isom}(G(\Gamma), d_r) \to \mathrm{Aut}(\mathcal{P}(\Gamma))$; moreover, any $\varphi \in \mathrm{Isom}(G(\Gamma), d_r)$ maps vertices in a standard flat to vertices in a standard flat of the same dimension. The homomorphism h is surjective if $\mathrm{Out}(G(\Gamma))$ is finite.

Remark 4.17 For any finite simplicial graphs Γ_1 and Γ_2, we have $G(\Gamma_1) \cong G(\Gamma_2)$ if and only if $(G(\Gamma_1), d_r)$ and $(G(\Gamma_2), d_r)$ are isometric as metric spaces. The "only if" direction follows from [23; 49]. For the other direction, let $\varphi\colon (G(\Gamma_1), d_r) \to (G(\Gamma_2), d_r)$ be an isometry. Pick $v \in G(\Gamma_1)$, and let $\{l_i\}_{i=1}^{k}$ be the collection of standard geodesics passing through v. Pick $v_i \in G(\Gamma_1)$ such that $v_i \in l_i \setminus \{v\}$. Then $d_r(v, v_i) = 1$ for $1 \le i \le k$, and $d_r(v_i, v_j) = 2$ for $1 \le i \ne j \le k$. So $d_r(\varphi(v), \varphi(v_i)) = 1$ for $1 \le i \le k$, and $d_r(\varphi(v_i), \varphi(v_j)) = 2$ for $1 \le i \ne j \le k$, and $\angle_v(v_i, v_j) = \frac{\pi}{2}$ if and only if $\angle_{\varphi(v)}(\varphi(v_i), \varphi(v_j)) = \frac{\pi}{2}$ by (4-6). This induces a graph embedding $\Gamma_1 \to \Gamma_2$. By considering φ^{-1}, we obtain another graph embedding $\Gamma_2 \to \Gamma_1$. Hence Γ_1 and Γ_2 are isomorphic.

Corollary 4.18 *If $\mathrm{Out}(G(\Gamma))$ is finite and $\mathrm{QI}(G(\Gamma))$ is the quasi-isometry group of $G(\Gamma)$, then we have the following commutative diagram, where i_1, i_2 and i_3 are injective homomorphisms:*

$$\begin{tikzcd} \mathrm{Isom}(G(\Gamma), d_w) \arrow[r, "i_1"] \arrow[rr, bend right, "i_3"'] & \mathrm{QI}(G(\Gamma)) \arrow[r, "i_2"] & \mathrm{Isom}(G(\Gamma), d_r). \end{tikzcd}$$

Proof The homomorphisms i_1 and i_3 are obvious, and i_2 is given by Lemma 4.5 and Corollary 4.15. It is clear that i_2 is a group homomorphism and $i_3 = i_2 \circ i_1$. Note that i_3 is injective, so i_1 is injective. Pick $\alpha \in \mathrm{QI}(G(\Gamma))$; by Corollary 4.15, we know $i_2(\alpha) = \mathrm{Id}$ implies the image of every standard flat under α is uniformly Hausdorff close to itself; thus α is of bounded distance from the identity map. □

5 Quasi-isometries and special subgroups

Let $G(\Gamma)$ be a RAAG with finite outer automorphism group. In this section, we characterize all other RAAGs quasi-isometric to $G(\Gamma)$.

5.1 Preservation of extension complex

Lemma 5.1 *Let Γ be a finite simplicial graph. Pick a vertex $w \in \Gamma$, and let Γ_w be the minimal stable subgraph containing w. Denote $\Gamma_1 = \mathrm{lk}(w)$ and $\Gamma_2 = \mathrm{lk}(\Gamma_1)$ (see Section 2.1 for the definition of links). Then exactly one of the following is true:*

(1) *Γ_w is a clique. In this case, $\mathrm{St}(w)$ is a stable subgraph.*

(2) *Both Γ_1 and $\Gamma_1 \circ \Gamma_2$ are stable subgraphs of Γ. Moreover, Γ_2 is disconnected.*

Recall that we use $(\Gamma')^{\perp}$ to denote the orthogonal complement of the subgraph $\Gamma' \subset \Gamma$ (see Section 2.1), and we assume $(\varnothing)^{\perp} = \Gamma$.

Proof If $\Gamma_w \subset \mathrm{St}(w)$, then Γ_w is a clique by Lemma 3.23. We also deduce from Lemma 3.23 that each vertex of $\mathrm{St}(w) \setminus \Gamma_w$ is in $\Gamma_w^{\perp}$. Moreover, $\Gamma_w^{\perp} \subset w^{\perp}$ since $w \in \Gamma_w$. Thus $\mathrm{St}(w)$ is the full subgraph spanned by vertices in Γ_w and $\Gamma_w^{\perp}$. So $\mathrm{St}(w)$ is stable by Lemma 3.12.

If $\Gamma_w \not\subseteq \mathrm{St}(w)$, let Γ_{11} be the full subgraph spanned by vertices in $\Gamma_w \cap \mathrm{lk}(w)$, and let Γ_2' be the full subgraph spanned by vertices in $\Gamma_w \setminus \Gamma_{11}$. By Lemma 3.23, $\Gamma_w = \Gamma_{11} \circ \Gamma_2'$ and $\Gamma_2' = \Gamma_2$. Note that Γ_2 is disconnected with isolated point $w \in \Gamma_2$, and Γ_{11} may be empty.

Let $V_w = v(\Gamma_w)$ be the vertex set of Γ_w and let Γ_{12} be the full subgraph spanned by $V_w^{\perp}$. Then $\Gamma_w \circ \Gamma_{12} = \Gamma_{11} \circ \Gamma_2 \circ \Gamma_{12}$ is stable by Lemma 3.12. Pick a vertex $v \in \Gamma_1 \setminus \Gamma_{11}$; then $v \in w^{\perp} \subset \mathrm{St}(u)$ for any vertex $u \in \Gamma_w$ by Lemma 3.23. Thus $v \in \Gamma_{12}$ and $\Gamma_1 \subset \Gamma_{11} \circ \Gamma_{12}$. On the other hand, $w \in \Gamma_2$, so $\Gamma_{11} \circ \Gamma_{12} \subset \Gamma_1$ and $\Gamma_1 = \Gamma_{11} \circ \Gamma_{12}$. Since Γ_2 does not contain any clique factor and $\Gamma_{11} \circ \Gamma_2 \circ \Gamma_{12} = \Gamma_1 \circ \Gamma_2$ is stable, we know Γ_1 is stable in Γ by Theorem 2.9. □

Remark 5.2 In the above proof, Γ_{12} may be empty. But if $\Gamma_{12} \neq \varnothing$, then it does not contain any clique join factor. Thus Γ_{11} is the maximal clique join factor of $\Gamma_{11} \circ \Gamma_2 \circ \Gamma_{12}$.

The next result answers the question at the end of Example 3.22.

Theorem 5.3 *Suppose* $\mathrm{Out}(G(\Gamma))$ *is finite and let* $q\colon X(\Gamma) \to X(\Gamma')$ *be a quasi-isometry. Then* q *induces a simplicial isomorphism* $q_*\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$*; in particular,* $\mathrm{Out}(G(\Gamma'))$ *is transvection-free.*

In the following proof, we identify Γ with the one-skeleton of $F(\Gamma)$, which is the flag complex of Γ. Also recall that there are label-preserving projections $\pi\colon \mathcal{P}(\Gamma) \to F(\Gamma)$ and $\pi\colon \mathcal{P}(\Gamma') \to F(\Gamma')$.

Proof By Lemma 4.5, there is a simplicial embedding $q_*\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$. Note that $q_*(\mathcal{P}(\Gamma))$ is a full subcomplex in $\mathcal{P}(\Gamma')$. To see this, pick a simplex $\Delta' \subset \mathcal{P}(\Gamma')$ with its vertices in $q_*(\mathcal{P}(\Gamma))$. Then each vertex of Δ' comes from a stable standard geodesic line in $X(\Gamma')$. Thus there exists a stable standard flat $F' \subset X(\Gamma')$ such that $\Delta(F') = \Delta'$ by Lemma 3.18. By considering the quasi-inverse of q, we know F' is Hausdorff close to the q–image of a stable standard flat in $X(\Gamma)$. Thus $\Delta(F') = \Delta' \subset q_*(\mathcal{P}(\Gamma))$.

Pick a vertex $p \in X(\Gamma)$, and let $\{\Delta_i\}_{i=1}^k$, $\{F_i\}_{i=1}^k$, $\{\Delta'_i\}_{i=1}^k$ and $\{F'_i\}_{i=1}^k$ be as in the proof of Lemma 4.10. We claim

$$\bigcap_{i=1}^{k} F'_i \neq \varnothing. \tag{5-1}$$

Suppose (5-1) is not true. Then there exist $1 \le i_1 \neq i_2 \le k$ and hyperplane $h' \subset X(\Gamma)$ such that h' separates F'_{i_1} and F'_{i_2}. Let l' be a standard geodesic that intersects h' transversely, and let $v' = \Delta(l')$. By the discussion in Lemma 4.9, we can find vertices $v'_1 \in \Delta'_{i_1}$ and $v'_2 \in \Delta'_{i_2}$ such that v'_1 and v'_2 are separated by $\mathrm{St}(v')$. If there exists i_0 such that $F'_{i_0} \cap h \neq \varnothing$, then $v' \in q_*(\mathcal{P}(\Gamma'))$, and we can prove (5-1) as in Lemma 4.10. Now we assume $F'_i \cap h' = \varnothing$ for any i. Let $w' = \pi(v') \in \Gamma'$, and let $\Gamma_{w'}$ be the minimal stable subgraph of Γ' that contains w'.

We apply Lemma 5.1 to $w' \in \Gamma'$; if case (1) is true, let F' be the standard flat in $X(\Gamma')$ such that $l' \subset F'$ and $\Gamma_{F'} = \Gamma_{w'}$. Since $\Gamma_{w'}$ is stable, $\Delta(F') \subset q_*(\mathcal{P}(\Gamma'))$; in particular, $v' \in q_*(\mathcal{P}(\Gamma'))$, and we can prove (5-1) as in Lemma 4.10.

If case (2) is true, let $\Gamma'_1 = \mathrm{lk}(w')$ and let $\Gamma'_2 = \mathrm{lk}(\Gamma'_1)$. Take K'_1 and K' to be the standard subcomplexes in $X(\Gamma')$ such that: (a) the defining graphs $\Gamma_{K'_1}$ and $\Gamma_{K'}$ of K'_1 and K' satisfy $\Gamma_{K'_1} = \Gamma'_1$ and $\Gamma_{K'} = \Gamma'_1 \circ \Gamma'_2$; (b) $l' \subset K'$ and $K'_1 \subset K'$. Set $M'_1 = \Delta(K'_1)$ and $M' = \Delta(K')$. Let K'_2 be an orthogonal complement of K'_1 in K'; ie K'_2 is a standard subcomplex such that $\Gamma_{K'_2} = \Gamma'_2$ and $K' = K'_1 \times K'_2$. It follows that $M' = M'_1 * M'_2$ for $M'_2 = \Delta(K'_2)$. By construction, $v' \in M'$ and $\mathrm{lk}(v') = M'_1$.

Since K' and K'_1 are stable, there exist stable standard subcomplexes K and K_1 in $X(\Gamma)$ such that $q(K) \overset{\infty}{=} K'$ and $q(K_1) \overset{\infty}{=} K'_1$. Moreover, by applying Theorem 2.9 to the quasi-isometry between K and K', there exists a standard subcomplex $K_2 \subset K$ such that $K = K_1 \times K_2$, and K_2 is quasi-isometric to K'_2. Thus Γ_{K_2} is also disconnected. Let $M_i = \Delta(K_i) \subset \mathcal{P}(\Gamma)$ for $i = 1, 2$, and let $M = M_1 * M_2 = \Delta(K)$. Then $q_*(M_1) \subset M'_1$ (at this stage we may not know $q_*(M_1) = M'_1$), and

$$q_*^{-1}(M'_1) = M_1. \tag{5-2}$$

To see this, pick a simplex $\Delta \subset \mathcal{P}(\Gamma)$ with $q_*(\Delta) \subset M'_1$. Suppose $\Delta = \Delta(F)$ for a stable standard flat $F \subset X(\Gamma)$. Then $q(F) \subset_\infty K'_1$; hence $F \subset_\infty K_1$ and $\Delta \subset M_1$.

Let $L = \bigcup_{i=1}^k \Delta_i$ and $L' = \bigcup_{i=1}^k \Delta'_i$. By the proof of Lemma 4.10, $L' \setminus (\mathrm{St}(v') \cap L')$ is disconnected; thus $L \setminus q_*^{-1}(\mathrm{St}(v') \cap L')$ is disconnected. Recall that $\mathrm{lk}(v') = M'_1$, and we are assuming $v' \notin L'$. Thus $(\mathrm{St}(v') \cap L') \subset M'_1$. Then $q_*^{-1}(\mathrm{St}(v') \cap L') \subset q_*^{-1}(M'_1)$; hence $q_*^{-1}(\mathrm{St}(v') \cap L') \subset M_1$ by (5-2).

Let $N = \pi(q_*^{-1}(\mathrm{St}(v') \cap L'))$, and let $N_i = \pi(M_i)$ for $i = 1, 2$. Then N separates $F(\Gamma)$, $N \subset N_1$ and N_2 is disconnected. Pick vertices u_1, u_2 in different connected components of N_2; then $d(u_1,u_2) \geq 2$ (since N_2 is the full subcomplex spanned by Γ_{K_2}). Since $\pi(M) = N_1 * N_2 \subset F(\Gamma)$, we have $N \subset \mathrm{St}(u_i) \setminus \{u_i\}$ for $i = 1,2$. Let $\{C_j\}_{j=1}^{d}$ be the connected components of $F(\Gamma) \setminus N$. Then at most one of C_j is contained in $\mathrm{St}(u_1)$. If $d \geq 3$, then $\mathrm{St}(u_1)$ would separate $F(\Gamma)$, which is a contradiction. Now we suppose $d = 2$. Note that for $i = 1,2$, there must exist j such that $C_j \subset \mathrm{St}(u_i)$: otherwise, $\mathrm{St}(u_i)$ would separate $F(\Gamma)$. Moreover, if $C_j \subset \mathrm{St}(u_i)$, then $u_i \in C_j$. So we can assume without loss of generality that $C_1 \subset \mathrm{St}(u_1)$ and $C_2 \subset \mathrm{St}(u_2)$, which implies $F(\Gamma) = \mathrm{St}(u_1) \cup \mathrm{St}(u_2)$, and again we have a contradiction by Lemma 4.12. Thus case (2) is impossible, and (5-1) is true.

Let $\{F_\lambda\}_{\lambda\in\Lambda}$ be the collection of maximal standard flats in $X(\Gamma)$. Then $X(\Gamma) = \bigcup_{\lambda\in\Lambda} F_\lambda$. For each λ, let F'_λ be the unique maximal standard flat in $X(\Gamma')$ such that $q(F_\lambda) \overset{\infty}{=} F'_\lambda$. Then

$$X(\Gamma') \overset{\infty}{=} \bigcup_{\lambda\in\Lambda} F'_\lambda. \tag{5-3}$$

Let $h \subset X(\Gamma')$ be an arbitrary hyperplane. Then $h \cap \left(\bigcup_{\lambda\in\Lambda} F'_\lambda\right) \neq \varnothing$: otherwise, $\bigcup_{\lambda\in\Lambda} F'_\lambda$ would stay on one side of the hyperplane since it is a connected set by (5-1), and this contradicts (5-3). Pick any standard geodesic $r \subset X(\Gamma')$, and let h_r be a hyperplane dual to r. Then there exists $\lambda \in \Lambda$ such that $F'_\lambda \cap h_r \neq \varnothing$. It follows that $r \subset_\infty F'_\lambda$. So $\Delta(r) \in \Delta(F'_\lambda) \subset q_*(\mathcal{P}(\Gamma))$, which implies q_* is surjective on the vertices. However, $q_*(\mathcal{P}(\Gamma))$ is a full subcomplex in $\mathcal{P}(\Gamma')$, so q_* is surjective. □

5.2 Coherent ordering and coherent labeling

Throughout this section, we assume that $\mathrm{Out}(G(\Gamma))$ is finite and $G(\Gamma) \ncong \mathbb{Z}$. If $q\colon G(\Gamma) \to G(\Gamma')$ is a quasi-isometry, then $G(\Gamma')$ has a quasi-action (see [46, Definition 2.2]) on $G(\Gamma)$, which induces a group homomorphism

$$H\colon G(\Gamma') \to \mathrm{QI}(G(\Gamma)).$$

On the other hand, since $G(\Gamma)$ acts by isometries on $X(\Gamma)$, we can identify $G(\Gamma)$ as a subgroup of $\mathrm{QI}(G(\Gamma))$ (more precisely, we embed $G(\Gamma)$ into $\mathrm{Isom}(G(\Gamma), d_w)$ and embed $\mathrm{Isom}(G(\Gamma), d_w)$ into $\mathrm{QI}(G(\Gamma))$ by Corollary 4.18). In this subsection, we will seek to answer the following question:

$$\text{Does there exist } g \in \mathrm{QI}(G(\Gamma)) \text{ such that } g \cdot H(G(\Gamma')) \cdot g^{-1} \subset G(\Gamma)?$$

Recall that we have picked an identification between $G(\Gamma)$ and the 0–skeleton of $X(\Gamma)$. Each circle in the 1–skeleton of the Salvetti complex of $G(\Gamma)$ is labeled by an element in the standard generating set S of $G(\Gamma)$. Moreover, we have chosen an orientation for each such circle. By pulling back the labeling and orientation of edges to the universal cover $X(\Gamma)$, we obtain a $G(\Gamma)$–invariant directed labeling of edges in $X(\Gamma)$. Moreover, both the labeling and orientation of edges in $X(\Gamma)$ are compatible with parallelism between edges. This also induces an associated $G(\Gamma)$–invariant labeling of vertices in $\mathcal{P}(\Gamma)$.

Let $\{l_\lambda\}_{\lambda\in\Lambda}$ be the collection of standard geodesics in $X(\Gamma)$, and let $V_\lambda = v(l_\lambda)$ be the vertex set of l_λ. A *coherent ordering* of $G(\Gamma)$ is obtained by assigning a collection of bijections $f_\lambda\colon V_\lambda \to \mathbb{Z}$ for each $\lambda \in \Lambda$ such that if l_{λ_1} and l_{λ_2} are parallel, then the $f_{\lambda_2} \circ p \circ f_{\lambda_1}^{-1}\colon \mathbb{Z} \to \mathbb{Z}$ is a translation, where $p\colon V_{\lambda_1} \to V_{\lambda_2}$ is the map induced by parallelism. The map f_λ pulls back the total order on $\mathbb{Z}$ to V_λ, which we denote by $\leq_\lambda$. Then $p\colon V_{\lambda_1} \to V_{\lambda_2}$ is order preserving.

Two coherent orderings Ω_1 and Ω_2 are *equivalent*, denoted by $\Omega_1 = \Omega_2$, if their collections of bijections agree up to a translation of $\mathbb{Z}$. Recall that we have a $G(\Gamma)$–invariant orientation of edges in $X(\Gamma)$ which is compatible with parallelism between edges. This induces a unique coherent ordering Ω of $G(\Gamma)$ up to the equivalence relation defined before. Moreover, for any element $g \in G(\Gamma)$, the pull-back $g^*(\Omega)$ is also a coherent ordering; additionally, $g^*(\Omega) = \Omega$.

Recall that for any vertex $v \in X(\Gamma)$, there is a label-preserving simplicial embedding $i_v\colon F(\Gamma) \to \mathcal{P}(\Gamma)$ by considering the standard geodesics passing through v. A *coherent labeling* of $G(\Gamma)$ is a simplicial map $a\colon \mathcal{P}(\Gamma) \to F(\Gamma)$ such that $a \circ i_v\colon F(\Gamma) \to F(\Gamma)$ is a simplicial isomorphism for every vertex $v \in X(\Gamma)$.

The label-preserving projection $L\colon \mathcal{P}(\Gamma) \to F(\Gamma)$ gives rise to a coherent labeling of $G(\Gamma)$. Recall that $G(\Gamma)$ acts on $\mathcal{P}(\Gamma)$ by simplicial automorphisms, and the labeling of vertices in $\mathcal{P}(\Gamma)$ is $G(\Gamma)$–invariant. Thus for any element $g \in G(\Gamma)$, the pull-back $g^*(L)$ is also a coherent labeling and $g^*(L) = L$.

We have the following alternative characterization of elements in $\mathrm{Isom}(G(\Gamma), d_r)$.

Lemma 5.4 *There is a one-to-one correspondence which associates each element of* $\mathrm{Isom}(G(\Gamma), d_r)$ *to a triple consisting of*

(1) *a point $v \in G(\Gamma)$,*

(2) *a coherent ordering of $G(\Gamma)$ (up to the equivalence relation defined above),*

(3) *a coherent labeling of $G(\Gamma)$.*

Proof Pick $\phi \in \mathrm{Isom}(G(\Gamma), d_r)$ and let $\varphi = h(\phi)\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma)$, where h is the monomorphism in Remark 4.16. Then $\varphi^* L = L \circ \varphi\colon \mathcal{P}(\Gamma) \to F(\Gamma)$ is a coherent labeling of $G(\Gamma)$. Pick a standard geodesic $l_1 \subset X(\Gamma)$. Then the parallel set P_{l_1}

admits a splitting $P_{l_1} = l_1 \times l_1^{\perp}$. Since ϕ maps vertices in a standard flat bijectively to vertices in a standard flat, there exists a standard geodesic $l_2 \subset X(\Gamma)$ such that $\phi(v(l_1)) = v(l_2)$ and $\phi(v(P_{l_1})) = P_{l_2}$; moreover, ϕ respects the product structure on P_{l_1}. Thus the pull-back $\phi^*\Omega$ is a coherent ordering of $G(\Gamma)$. Now we can set up the correspondence in one direction:

$$\phi \rightsquigarrow (\phi(\mathrm{id}), \phi^*\Omega, \varphi^* L).$$

Here, id denotes the identity element of $G(\Gamma)$.

Conversely, given a point $v \in G(\Gamma)$, a coherent ordering Ω' and a coherent labeling L', we can construct a map ϕ as follows. Set $\phi(\mathrm{id}) = v$. For $u \in G(\Gamma)$, pick a word $w_u = a_1 a_2 \cdots a_n$ representing u. Let u_i be the point in $G(\Gamma)$ represented by the word $a_1 a_2 \cdots a_i$ for $1 \le i \le n$, and let $u_0 = \mathrm{id}$. We define $q_i = \phi(a_1 a_2 \cdots a_i) \in G(\Gamma)$ inductively as follows. Set $q_0 = v$, and suppose q_{i-1} is already defined. Denote the standard geodesic containing u_{i-1} and u_i by l_i. Let $v_i = L'(\Delta(l_i))$, which is a vertex of Γ, and let l_i' be the standard line that contains q_{i-1} and is labeled by v_i. Denote the vertex set of l_i with the order from Ω' by $(v(l_i), \le_{\Omega'})$. Suppose that $k\colon (v(l_i), \le_{\Omega'}) \to (v(l_i'), \le_{\Omega})$ is the unique order-preserving bijection such that $k(u_{i-1}) = q_{i-1}$. Then we define $q_i = k(u_i)$.

We claim that for any other word w_u' representing u, we have $\phi(w_u) = \phi(w_u')$, and hence there is a well-defined map $\phi\colon G(\Gamma) \to G(\Gamma)$. To see this, recall that one can obtain w_u from w_u' by performing the following two basic moves:

(1) $w_1 a a^{-1} w_2 \to w_1 w_2$,

(2) $w_1 a b w_2 \to w_1 b a w_2$ when a and b commute.

It is clear that $\phi(w_1 a a^{-1} w_2) = \phi(w_1 w_2)$. For the second move, let u_{i-1}, u_i, u_i' and u_{i+1} be points in $G(\Gamma)$ represented by $w_1, w_1 a, w_1 b$ and $w_1 ab = w_1 ba$, respectively. Define $q_{i-1} = \phi(w_1)$, $q_i = \phi(w_1 a)$, $q_i' = \phi(w_1 b)$, $q_{i+1} = \phi(w_1 ab)$ and $q_{i+1}' = \phi(w_1 ba)$. Since L' is a coherent labeling, $\angle_{q_i}(q_{i+1}, q_{i-1}) = \angle_{q_{i-1}}(q_i, q_i') = \angle_{q_i'}(q_{i-1}, q_{i+1}') = \frac{\pi}{2}$; moreover, the standard geodesic containing q_i and q_{i+1} is parallel to the standard geodesic containing q_{i-1} and q_i'. Since Ω' is a coherent ordering, $d(q_i, q_{i+1}) = d(q_{i-1}, q_i')$; thus $\overline{q_i q_{i+1}}$ and $\overline{q_{i-1} q_i'}$ are parallel. Similarly, $\overline{q_{i-1} q_i}$ and $\overline{q_i' q_{i+1}'}$ are parallel; thus $q_{i+1} = q_{i+1}'$.

Now we define another map $\phi'\colon G(\Gamma) \to G(\Gamma)$, which serves as the inverse of ϕ. Set $\phi'(v) = \mathrm{id}$ and pick a word $w = a_1 a_2 \cdots a_n$. Let r_i be the point in $G(\Gamma)$ represented by $v a_1 a_2 \cdots a_i$ for $1 \le i \le n$, and let $r_0 = v$. We define $p_i = \phi'(v a_1 a_2 \cdots a_i)$ inductively as follows. Put $p_0 = \mathrm{id}$, and suppose p_{i-1} is already defined. Since L' is a coherent labeling, there exists a unique standard geodesic l_i containing p_{i-1} such that $L'(\Delta(l_i))$ and the edge $\overline{r_{i-1}r_i}$ share the same label. Let l_i' be the unique standard geodesic containing r_{i-1} and r_i, and let $k'\colon (v(l_i'), \leq_\Omega) \to (v(l_i), \leq_{\Omega'})$ be the unique order-preserving bijection such that $k'(r_{i-1}) = p_{i-1}$. Put $p_i = k'(r_i)$. By a similar argument as above, $\phi'\colon G(\Gamma) \to G(\Gamma)$ is well defined. It is not hard to deduce the following properties from our construction:

(1) $\phi' \circ \phi = \phi \circ \phi' = \mathrm{Id}$.

(2) $d_r(\phi(v_1), \phi(v_2)) \leq d_r(v_1, v_2)$ and $d_r(\phi'(v_1), \phi'(v_2)) \leq d_r(v_1, v_2)$ for any vertices $v_1, v_2 \in G(\Gamma)$.

(3) If $L' = L$ and $\Omega' = \Omega$, then ϕ is a left translation. If, in addition, $v = \mathrm{id}$, then $\phi = \mathrm{Id}$.

It follows from (1) and (2) that $\phi \in \mathrm{Isom}(G(\Gamma), d_r)$. Moreover, $v = \phi(\mathrm{id})$, $L' = \varphi^* L$ ($\varphi = h(\phi)$, where h is the monomorphism in Remark 4.16) and $\Omega' = \phi^*\Omega$; thus we have established the required one-to-one correspondence. □

Pick finite simplicial graphs Γ and Γ' such that: (1) $\mathrm{Out}(G(\Gamma))$ is finite; (2) there exists a simplicial isomorphism $s\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$. By Lemma 4.10, s induces a map $\phi\colon G(\Gamma) \to G(\Gamma')$. For every $g' \in G(\Gamma')$, there is a left translation

$$\bar{\phi}_{g'}\colon G(\Gamma') \to G(\Gamma'),$$

which gives rise to a simplicial isomorphism $\bar{s}_{g'}\colon \mathcal{P}(\Gamma') \to \mathcal{P}(\Gamma')$. Let $s_{g'} = s^{-1} \circ \bar{s}_{g'} \circ s$. Then $s_{g'}$ gives rise to a map $\phi_{g'} \in \mathrm{Isom}(G(\Gamma), d_r)$ by Corollary 4.15; moreover, by Lemma 4.10,

$$\bar{\phi}_{g'} \circ \phi = \phi \circ \phi_{g'} \tag{5-4}$$

for any $g' \in G(\Gamma')$. So $G(\Gamma')$ acts on $G(\Gamma)$, and we can define a homomorphism $\Phi\colon G(\Gamma') \to \mathrm{Isom}(G(\Gamma), d_r)$ by sending g' to $\phi_{g'}$. Φ is injective since each step in defining Φ is injective.

Lemma 5.5 *In the above setting, there exists an element $\phi_1 \in \mathrm{Isom}(G(\Gamma), d_r)$ such that it conjugates the image of Φ to a finite-index subgroup of $G(\Gamma)$.*

We identify $G(\Gamma)$ as a subgroup of $\mathrm{Isom}(G(\Gamma), d_r)$ via the left action of $G(\Gamma)$ on itself.

Proof Pick a reference point $q \in \mathrm{Im}\,\phi$, and let $K_q = (F(\Gamma'))_q$. Denote the points in $\phi^{-1}(q)$ by $\{p_\lambda\}_{\lambda\in\Lambda}$, and let $K_{p_\lambda} = (F(\Gamma))_{p_\lambda}$. Since the $\{\phi(K_{p_\lambda})\}_{\lambda\in\Lambda}$ are distinct subcomplexes of K_q, the set Λ must be finite.

Let $L\colon \mathcal{P}(\Gamma) \to F(\Gamma)$ and Ω be the coherent labeling and coherent ordering induced by the $G(\Gamma)$–invariant labeling of $X(\Gamma)$ and $\mathcal{P}(\Gamma)$. In a similar fashion, we can obtain a coherent labeling $L'\colon \mathcal{P}(\Gamma') \to F(\Gamma')$ and a coherent ordering Ω' for $G(\Gamma')$ which are invariant under the $G(\Gamma')$–action, ie

$$(\bar{s}_{g'})^* L' = L' \quad \text{and} \quad (\bar{\phi}_{g'})^* \Omega' = \Omega'. \tag{5-5}$$

Our goal is to find a coherent labeling L_1 and a coherent ordering Ω_1 of $G(\Gamma)$ such that $(s_{g'})^* L_1 = L_1$ and $(\phi_{g'})^* \Omega_1 = \Omega_1$ for any $g' \in G(\Gamma')$.

Let $i_q\colon F(\Gamma') \to \mathcal{P}(\Gamma')$ be the canonical embedding, and let

$$L_1 = L \circ s^{-1} \circ i_q \circ L' \circ s$$

be the simplicial map from $\mathcal{P}(\Gamma)$ to $F(\Gamma)$. Pick an arbitrary $p \in G(\Gamma)$, and let $i_p\colon F(\Gamma) \to \mathcal{P}(\Gamma)$ be the canonical embedding. We need to show $L_1 \circ i_p$ is a simplicial isomorphism. Let $K_p = i_p(F(\Gamma))$, and let $g_1' \in G(\Gamma')$ such that $g_1' \cdot \phi(p) = q$. Then $i_q \circ L'|_{s(K_p)} = \bar{s}_{g_1'}|_{s(K_p)}$. Thus

$$L_1 \circ i_p = L \circ s^{-1} \circ i_q \circ L' \circ s \circ i_p = L \circ s^{-1} \circ \bar{s}_{g_1'} \circ s \circ i_p = L \circ s_{g_1'} \circ i_p,$$

which is a simplicial isomorphism by Lemma 4.10. It follows that L_1 is a coherent labeling; moreover,

$$\begin{aligned}(s_{g'})^* L_1 &= (L \circ s^{-1} \circ i_q \circ L' \circ s) \circ (s^{-1} \circ \bar{s}_{g'} \circ s) = L \circ s^{-1} \circ i_q \circ L' \circ \bar{s}_{g'} \circ s \\ &= L \circ s^{-1} \circ i_q \circ L' \circ s = L_1\end{aligned}$$

for any $g' \in G(\Gamma')$, where the third equality follows from (5-5). So L_1 is the required coherent labeling.

To simplify notation, we will write $x <_\Omega y$ if $x < y$ under the ordering Ω. We define Ω_1 as follows. Let $p_1, p_2 \in G(\Gamma)$ be two distinct points in a standard geodesic line. If $\phi(p_1) \neq \phi(p_2)$, then we set $p_1 <_{\Omega_1} p_2$ if and only if $\phi(p_1) <_{\Omega'} \phi(p_2)$. If $\phi(p_1) = \phi(p_2)$, then by (5-4), there exists a unique $g' \in G(\Gamma')$ such that $\phi_{g'}(p_i) \in \phi^{-1}(q)$ for $i = 1, 2$, and we set $p_1 <_{\Omega_1} p_2$ if and only if $\phi_{g'}(p_1) <_\Omega \phi_{g'}(p_2)$. It follows from (5-5), (5-4) and our construction that $p_1 <_{\Omega_1} p_2$ if and only if $\phi_{g'}(p_1) <_{\Omega_1} \phi_{g'}(p_2)$ for any p_1, p_2 in the same standard geodesic line and any $g' \in G(\Gamma')$; thus $(\phi_{g'})^* \Omega_1 = \Omega_1$.

To verify Ω_1 is coherent, pick parallel standard geodesics l_1 and l_2 in $X(\Gamma)$, and pick distinct vertices $p_{11}, p_{12} \in l$. Let p_{21}, p_{22} be the corresponding vertices in l_2 via parallelism. We assume $p_{11} <_{\Omega_1} p_{12}$; it suffices to prove $p_{21} <_{\Omega_1} p_{22}$.

Case 1 We assume $\phi(p_{11}) \neq \phi(p_{12})$. Recall that l_1 can be realized as an intersection of finitely many maximal standard flats, so by Lemma 4.10, there exists a standard geodesic line $l_1' \subset X(\Gamma')$ such that $\phi(v(l_1)) \subset v(l_1')$ and $\phi(v(P_{l_1})) \subset v(P_{l_1'})$; moreover, ϕ respects the product structures of P_{l_1} and $P_{l_1'}$. Thus $\overline{\phi(p_{11})\phi(p_{21})}$ and $\overline{\phi(p_{21})\phi(p_{22})}$

are the opposite sides of a flat rectangle in $X(\Gamma')$. Now $p_{21} <_{\Omega_1} p_{22}$ follows since Ω' is coherent.

Case 2 We assume $\phi(p_{11}) = \phi(p_{12}) \neq \phi(p_{21})$. In this case, we can assume without loss of generality that $\phi(p_{11}) = \phi(p_{12}) = q$ (since $(\phi_{g'})^*\Omega_1 = \Omega_1$), and the points p_{11} and p_{21} stay in the same standard geodesic. For $i = 1, 2$, let r_i be the standard geodesic passing p_{1i} and p_{2i}. Take $r'_i \subset X(\Gamma')$ and $l'_i \subset X(\Gamma')$ to be the standard geodesics such that $\phi(v(r_i)) \subset v(r'_i)$ and $\phi(v(l_i)) \subset v(l'_i)$, respectively. Let $q' = \phi(p_{21})$. Since ϕ restricted to $v(P_{l_1})$ respects the product structure, $\phi(p_{21}) = \phi(p_{22}) = q'$ and $r'_1 = r'_2$.

Let $\overline{\phi}_{g'}$ be the left translation such that $\overline{\phi}_{g'}(q') = q$. Since $q' \in r'_1$ and $q \in r'_1$, we have that $\overline{\phi}_{g'}$ is a translation along r'_1, and $\overline{s}_{g'}$ fixes every point in $\mathrm{St}(\Delta(r'_1))$; hence $s_{g'}$ fixes every point in $s^{-1}(\mathrm{St}(\Delta(r'_1))) = \mathrm{St}(\Delta(r_1))$, and

$$\phi_{g'}(r_i) = r_i \tag{5-6}$$

for $i = 1, 2$. Let $l_3 = \phi_{g'}(l_2)$. Then l_3 is parallel to l_1 (or l_2). To see this, note that $\Delta(l_1) \in \mathrm{St}(\Delta(r_1))$; hence $\Delta(l_1)$ is fixed by $s_{g'}$. Put $p_{3i} = \phi_{g'}(p_{2i})$ for $i = 1, 2$. Then $p_{3i} \in r_i$ by (5-6); hence $\overline{p_{11}p_{12}}$ and $\overline{p_{31}p_{32}}$ are the opposite sides of a flat rectangle. Moreover, $p_{3i} \in \phi^{-1}(q)$ for $i = 1, 2$ by (5-4), so $p_{31} <_{\Omega_1} p_{32}$ since Ω is coherent, and $\Omega = \Omega_1$ while restricted on $\phi^{-1}(q)$. Now the $G(\Gamma')$–invariance of Ω_1 implies $p_{21} <_{\Omega_1} p_{22}$.

Case 3 If $\phi(p_{11}) = \phi(p_{12}) = \phi(p_{21})$, then we can assume without loss of generality that they all equal to q. It follows that $\phi(p_{22}) = q$ since ϕ respects the product structure while restricted to $v(P_{l_1})$. Thus $p_{21} <_{\Omega_1} p_{22}$ by definition.

By Lemma 5.4, there exists $\phi_1 \in \mathrm{Isom}(G(\Gamma), d_r)$ such that $\phi_1^*\Omega = \Omega_1$ and $s_1^*L = L_1$ ($s_1 = h(\phi_1)$ where h is the monomorphism in Remark 4.16). Thus

$$\begin{aligned}(\phi_1 \circ \phi_{g'} \circ \phi_1^{-1})^*\Omega &= (\phi_1^{-1})^* \circ (\phi_{g'})^* \circ (\phi_1^*\Omega) = (\phi_1^{-1})^* \circ (\phi_{g'})^*\Omega_1 \\ &= (\phi_1^{-1})^*\Omega_1 = \Omega\end{aligned}$$

for any $g' \in G(\Gamma')$. Similarly, $(s_1 \circ s_{g'} \circ s_1^{-1})^*L = L$ for any $g' \in G(\Gamma')$. Note that $s_1 \circ s_{g'} \circ s_1^{-1} = h(\phi_1 \circ \phi_{g'} \circ \phi_1^{-1})$; thus by Lemma 5.4, $G(\Gamma')$ acts on $G(\Gamma)$ by left translations via $g' \to \phi_1 \circ \phi_{g'} \circ \phi_1^{-1}$. This induces a monomorphism $G(\Gamma') \to G(\Gamma)$. Moreover, by (5-4) and the fact that $\phi^{-1}(q)$ is finite, this action has finite quotient; thus we can realize $G(\Gamma')$ as a finite-index subgroup of $G(\Gamma)$. □

The next result basically says under suitable conditions, if there exists a quasi-isometry $q\colon G(\Gamma) \to G(\Gamma')$, then there exists a very "nice" quasi-isometry $q'\colon G(\Gamma) \to G(\Gamma')$. However, we do not insist that q' is of bounded distance away from q (compared to Theorem 4.13).

Theorem 5.6 *Let Γ and Γ' be finite simplicial graphs such that $\mathrm{Out}(G(\Gamma))$ is finite and $G(\Gamma')$ is quasi-isometric to $G(\Gamma)$. Then there exists a cubical map (see Definition 2.1) $\varphi\colon X(\Gamma) \to X(\Gamma')$ such that:*

(1) *The map φ is onto, and φ maps any standard flat in $X(\Gamma)$ onto a standard flat in $X(\Gamma')$ of the same dimension.*

(2) *The map φ maps combinatorial geodesics in the 1–skeleton of $X(\Gamma)$ to combinatorial geodesics in the 1–skeleton of $X(\Gamma')$.*

(3) *The map φ is a quasi-isometry.*

Proof Let $f\colon G(\Gamma) \to G(\Gamma')$ be a quasi-isometry. By Theorem 5.3, f induces a simplicial isomorphism $s\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$. By Lemma 4.10, s induces a map $\phi\colon G(\Gamma) \to G(\Gamma')$ such that $d_w(f(x), \phi(x)) < D$ for any $x \in G(\Gamma)$. Let ϕ_1 be the map in Lemma 5.5 and let $\varphi = \phi \circ \phi_1^{-1}$. We will use the same notation as in the proof of Lemma 5.5.

We claim that if $F = \bigcap_{i=1}^{h} F_i$, where each F_i is a maximal standard flat, then there exists a unique standard flat $F' \subset G(\Gamma')$ such that $\phi(v(F)) = v(F')$. To see this, let F'_i be the maximal standard flat in $X(\Gamma')$ such that $\Delta(F'_i) = s(\Delta(F_i))$ for $1 \le i \le h$, and let $F' = \bigcap_{i=1}^{h} F'_i$. Then it follows from Lemma 4.10 that $\phi(v(F)) \subset v(F')$. Recall that $G(\Gamma')$ acts on $G(\Gamma')$, $\mathcal{P}(\Gamma')$, $G(\Gamma)$ and $\mathcal{P}(\Gamma)$. The stabilizer $\mathrm{Stab}(v(F'))$ fixes $\Delta(F'_i)$ for all i; hence it fixes Δ_i for all i, and $\mathrm{Stab}(v(F')) \subset \mathrm{Stab}(v(F))$. Since $\mathrm{Stab}(v(F'))$ acts on $v(F')$ transitively, (5-4) implies $\phi(v(F)) = v(F')$ and $|\phi^{-1}(y) \cap F| = |\phi^{-1}(y') \cap F|$ for any $y, y' \in v(F')$. It also follows that $\mathrm{Stab}(v(F)) \subset \mathrm{Stab}(v(F'))$; thus $\mathrm{Stab}(v(F')) = \mathrm{Stab}(v(F))$.

Note that the above claim is also true for φ, and any standard geodesic satisfies the assumption of the claim. Moreover, φ is surjective since ϕ_1 is surjective by (5-4). Pick standard geodesics $l \subset X(\Gamma)$ and $l' \subset X(\Gamma')$ such that $v(l') = \varphi(v(l))$, and we identify $v(l)$ and $v(l')$ with $\mathbb{Z}$ in an order-preserving way. Then the above claim and the construction of ϕ_1 imply that $\varphi|_{v(l)}$ is of the form

$$\varphi(a) = \lfloor a/d \rfloor + r \tag{5-7}$$

for some integers r and d (with $d \ge 1$). In particular, φ can be extended to a simplicial map from the Cayley graph $C(\Gamma)$ of $G(\Gamma)$ to $C(\Gamma')$.

Pick a combinatorial geodesic $\omega \subset C(\Gamma)$ connecting vertices x and y; we claim that $\omega' = \phi(\omega)$ is also a geodesic in $C(\Gamma')$ (it could be a point). Let $\{v_i\}_{i=0}^{n}$ be vertices in ω such that for $0 \le i \le n-1$, we have that $[v_i, v_{i+1}]$ is a maximal subsegment of ω that is contained in a standard geodesic ($v_0 = x$ and $v_n = y$). Denote the corresponding standard geodesic by l_i. For $0 \le i \le n-1$, let $l'_i \subset X(\Gamma')$ be the standard geodesic such that $v(l_i') = \varphi(v(l_i))$, and let $\omega_i' = \phi([v_i, v_{i+1}])$. Then ω_i' is a (possibly degenerate) segment in l_i' by (5-7). Since ω is a geodesic, no two geodesics in $\{l_i\}_{i=0}^{n-1}$ are parallel. Note that φ is induced by a simplicial isomorphism between $\mathcal{P}(\Gamma)$ and $\mathcal{P}(\Gamma')$; thus the same property is true for the collection of geodesics $\{l_i'\}_{i=0}^{n-1}$. It follows that no hyperplane in $X(\Gamma')$ could intersect ω' at more than one point; hence ω' is a combinatorial geodesic.

Let $u_i = \varphi(v_i)$. Then $d_w(u_i, u_{i+1}) \le d_w(v_i, v_{i+1})$ by (5-7) (recall that d_w denotes the word metric on the corresponding group). Thus

$$d_w(\varphi(x), \varphi(y)) = \sum_{i=0}^{n-1} d_w(u_i, u_{i+1}) \le \sum_{i=0}^{n-1} d_w(v_i, v_{i+1}) = d_w(x, y) \tag{5-8}$$

for any $x, y \in G(\Gamma)$.

Pick $p \in G(\Gamma')$ and let $k = |\varphi^{-1}(p)|$. Then k does not depend on p by (5-4). It follows that $d_w(\varphi(x), \varphi(y)) \ge 1$ whenever $d_w(x, y) \ge k+1$. Now we can cut ω into pieces of length $k+1$. Since $\varphi(\omega)$ is a combinatorial geodesic,

$$d_w(\varphi(x), \varphi(y)) \ge \frac{d_w(x, y)}{k+1} - 1.$$

Note that φ naturally extends to a cubical map from $X(\Gamma)$ to $X(\Gamma')$, which satisfies all the required properties. □

Theorem 5.7 *If Γ and Γ' are finite simplicial graphs such that* $\mathrm{Out}(G(\Gamma))$ *is finite, then the following are equivalent:*

1. *$G(\Gamma')$ is quasi-isometric to $G(\Gamma)$.*
2. *$\mathcal{P}(\Gamma')$ is isomorphic to $\mathcal{P}(\Gamma)$ as simplicial complexes.*
3. *$G(\Gamma')$ is isomorphic to a subgroup of finite index in $G(\Gamma)$.*

Proof $(1) \Longrightarrow (2)$ follows from Theorem 5.3. $(2) \Longrightarrow (3)$ follows from Lemma 5.5. $(3) \Longrightarrow (1)$ is trivial. □

This establishes Theorem 1.2 in the introduction.

6 The geometry of finite-index RAAG subgroups

Throughout this section, we assume $G(\Gamma) \ncong \mathbb{Z}$, since the main results of this section (Theorems 6.13 and 6.19) are trivial when $G(\Gamma) \cong \mathbb{Z}$.

6.1 Constructing finite-index RAAG subgroups

A *right-angled Artin subgroup* is a subgroup which is also a right-angled Artin group. In this section, we introduce a process to obtain finite-index RAAG subgroups of an arbitrary RAAG.

Lemma 6.1 *Let X be a* CAT(0) *cube complex, let $l \subset X$ be a geodesic in the* 1–*skeleton and let $\{h_i\}_{i\in\mathbb{Z}}$ be consecutive hyperplanes dual to l. Let $\pi_l\colon X \to l$ be the* CAT(0) *projection. Then:*

1. *For every edge $e \subset X$, if $e \cap h_i = \varnothing$ for all i, then $\pi_l(e)$ is a vertex in l, and if $e \cap h_i \neq \varnothing$ for some i, then $\pi_l(e)$ is an edge in l.*
2. *If K is any connected subcomplex such that $e \cap h_i = \varnothing$ for all i, then $\pi_l(K)$ is a vertex in l; moreover, if K stays between h_i and h_{i+1}, then $\pi_l(K)$ is the vertex in l that stays between h_i and h_{i+1}.*
3. *For every interval $[a,b] \subset l$, we have that $\pi_l^{-1}([a,b])$ is a convex set in X. In particular, if $x \in l$ is a vertex, then $\pi_l^{-1}(x)$ is a convex subcomplex of X.*
4. *If K is a convex subcomplex such that $K \cap l \neq \varnothing$, then $\pi_l(K) = K \cap l$.*

Proof Here (1) and (3) follow from the fact the every hyperplane has a carrier, and (2) follows from (1). To see (4), it suffices to show that for every i such that $h_i \cap l \neq \varnothing$ and $h_i \cap K \neq \varnothing$, we have $e_i \subset K$ (e_i is the edge in l dual to h_i). Let N_{h_i} be the carrier of h_i. By Lemma 2.3, $d(x, N_{h_i} \cap K) \equiv c$ for any $x \in e_i$. Moreover, $d(x, N_{h_i} \cap K) = d(x, K)$ for x in the interior of e_i, so we must have $c = 0$: otherwise, the convexity of $d(\cdot, K)$ would imply $K \cap l = \varnothing$. □

Recall that $v(\mathcal{P}(\Gamma) \setminus \mathrm{St}(\Delta(l)))$ is the collection of vertices in $\mathcal{P}(\Gamma) \setminus \mathrm{St}(\Delta(l))$.

Lemma 6.2 *Let $l \subset X(\Gamma)$ be a standard geodesic. Then there is a map*

$$\pi_{\Delta(l)}\colon v(\mathcal{P}(\Gamma) \setminus \mathrm{St}(\Delta(l))) \to v(l)$$

such that if v_1 and v_2 are in the same connected component of $\mathcal{P}(\Gamma) \setminus \mathrm{St}(\Delta(l))$, then $\pi_{\Delta(l)}(v_1) = \pi_{\Delta(l)}(v_2)$.

Proof Let $\pi_l\colon X(\Gamma) \to l$ be the CAT(0) projection and let $l_1 \subset X(\Gamma)$ be a standard geodesic such that $d(\Delta(l_1), \Delta(l)) \geq 2$. Then $\pi_l(l_1)$ is a vertex in l by Lemma 3.1 and Corollary 3.2. Moreover, we claim $\pi_l(l_1) = \pi_l(l_2)$ if l_2 is a standard geodesic parallel to l_1. It suffices to prove the case when there is a unique hyperplane h separating l_1 from l_2. Note that $d(\Delta(l_1), \Delta(l)) \geq 2$ yields $h \cap l = \varnothing$, so l_1 and l_2 are pinched by two hyperplanes dual to l; then the claim follows from Lemma 6.1. Thus π_l

induces a well-defined map $\pi_{\Delta(l)}\colon v(\mathcal{P}(\Gamma)\setminus \mathrm{St}(\Delta(l))) \to v(l)$. If $\Delta(l_1)$ and $\Delta(l_2)$ are connected by an edge, then there exist standard geodesics l_1' and l_2' such that $l_1' \cap l_2' \neq \varnothing$ and l_i' is parallel to l_i for $i = 1, 2$. Thus $\pi_l(l_1) = \pi_l(l_1') = \pi_l(l_2') = \pi_l(l_2)$, and $\pi_{\Delta(l)}(\Delta(l_1)) = \pi_{\Delta(l)}(\Delta(l_2))$. $\square$

Pick a standard generating set S of $G(\Gamma)$, and let $C(\Gamma, S)$ be the Cayley graph. We identify $G(\Gamma)$ as a subset of $C(\Gamma, S)$ and attach higher-dimensional cubes to $C(\Gamma, S)$ to obtain a CAT(0) cube complex $X(\Gamma, S)$, which is basically the universal cover of the Salvetti complex. Here we would like to think of $G(\Gamma)$ as a fixed set and of $C(\Gamma, S)$ and $X(\Gamma, S)$ as objects formed by adding edges and cubes to $G(\Gamma)$ in a particular way determined by S, so we write S explicitly. We will choose a $G(\Gamma)$–equivariant orientation for edges in $X(\Gamma, S)$ as before.

An *S–flat* (or an *S–geodesic*) in $G(\Gamma)$ is defined to be the vertex set of a standard flat (or geodesic) in $X(\Gamma, S)$. We define $\mathcal{P}(\Gamma, S)$ as before such that its vertices correspond to coarse equivalence classes of S–geodesics.

We define an isometric embedding $I\colon G(\Gamma) \to \ell^1(v(\mathcal{P}(\Gamma, S)))$ which depends on S and the orientation of edges in $X(\Gamma, S)$. Pick a standard geodesic $l \subset X(\Gamma, S)$, and let $\pi_l\colon X(\Gamma, S) \to l$ be the CAT(0) projection. We identify $v(l)$ with $\mathbb{Z}^{\Delta(l)}$ in an orientation-preserving way such that $\pi_l(\mathrm{id}) = 0$ (id is the identity element in $G(\Gamma)$). Then π_l induces a coordinate function $I_{\Delta(l)}\colon G(\Gamma) \to \mathbb{Z}^{\Delta(l)}$. If we change l to a standard geodesic l_1 parallel to l, then $I_{\Delta(l)}$ and $I_{\Delta(l_1)}$ are identical by Lemma 6.1. Thus for every vertex $v \in \mathcal{P}(\Gamma)$, there is a well-defined coordinate function $I_v\colon G(\Gamma) \to \mathbb{Z}^v$. These coordinate functions induce a map $I\colon G(\Gamma) \to \mathbb{Z}^{(v(\mathcal{P}(\Gamma)))}$.

The map I is an embedding since every two points in $G(\Gamma)$ are separated by some hyperplane. $I(G(\Gamma)) \subset \ell^1(v(\mathcal{P}(\Gamma)))$ since for any $g \in G(\Gamma)$, there are only finitely many hyperplanes separating id and g. I naturally extends to a map $I\colon X(\Gamma, S) \to \ell^1(v(\mathcal{P}(\Gamma)))$, and it maps combinatorial geodesics to geodesics by the argument in Theorem 5.6. Thus I is an isometric embedding with respect to the ℓ^1 metric on $X(\Gamma, S)$. We say a convex subcomplex $K \subset X(\Gamma, S)$ is *nonnegative* if each point in $I(K)$ has nonnegative coordinates (this notion depends on the orientation of edges in $X(\Gamma, S)$). Let $\mathrm{CN}(\Gamma, S)$ be the collection of compact, convex, nonnegative subcomplexes of $X(\Gamma, S)$ that contain the identity.

For any $K \in \mathrm{CN}(\Gamma, S)$, we find a maximal collection of standard geodesics $\{c_i\}_{i=1}^{s}$ such that $c_i \cap K \neq \varnothing$ for all i and $\Delta(c_i) \neq \Delta(c_j)$ for any $i \neq j$. Let $g_i \in S$ be the label of edges in c_i and let $\alpha_i = \pi_{c_i}(\mathrm{id})$. Put $n_i = |v(K \cap c_i)|$ and $v_i = \alpha_i g_i^{n_i} \alpha_i^{-1}$. Let G' be the subgroup generated by $\{v_i\}_{i=1}^{s}$. It follows from the convexity of K that if a standard geodesic c is parallel to c_i and $c \cap K \neq \varnothing$, then $|v(K \cap c_i)| = |v(K \cap c)|$. Thus $\{v_i\}_{i=1}^{s}$ and G' do not depend on the choice of the c_i.

Lemma 6.3 *G' is a finite-index subgroup of $G(\Gamma)$.*

Proof We prove this by showing $G' \cdot v(K) = G(\Gamma)$. Let d_r be the syllable metric on $G(\Gamma)$ defined in Section 4.3. Pick a word $\alpha \in G(\Gamma)$ and assume $\alpha \in G' \cdot v(K)$ when $d_r(\alpha, \mathrm{id}) \le k-1$. If $d_r(\alpha, \mathrm{id}) = k$, then there exists $\beta \in G(\Gamma)$ such that $d_r(\mathrm{id}, \beta) = k-1$ and $d_r(\beta, \alpha) = 1$. Let $\beta = \beta_1\beta_2$ for $\beta_1 \in G'$ and $\beta_2 \in v(K)$. Then $d_r(\beta_2, \beta_1^{-1}\alpha) = 1$. Suppose c is the standard geodesic containing β_2 and $\beta_1^{-1}\alpha$. Then there exists i such that c_i and c are parallel. Note that $P_c \cap K$ is a convex set in the parallel set P_c, hence respects the natural splitting $P_c = c \times c^{\perp}$; moreover, the left action of v_i translates the c factor by n_i units and fixes the other factor. Thus there exists $d \in \mathbb{Z}$ and $\beta_2' \in K \cap c$ such that $v_i^d \beta_2' = \beta_1^{-1}\alpha$, which implies $\alpha = \beta_1 v_i^d \beta_2' \in G' \cdot v(K)$. □

Let Γ' be the full subgraph of $\mathcal{P}(\Gamma)$ spanned by points $\{\Delta(c_i)\}_{i=1}^s$. Then there is a natural homomorphism $G(\Gamma') \to G'$.

Lemma 6.4 *The homomorphism $G(\Gamma') \to G'$ is actually an isomorphism. Hence G' is a finite-index RAAG subgroup of $G(\Gamma)$.*

We will follow the strategy in [47], where the following version of the ping-pong lemma for right-angled Artin groups was used.

Theorem 6.5 [47, Theorem 4.1] *Let $G = G(\Gamma)$ and let X be a set with a G–action. Suppose the following hold:*

(1) *For each vertex v_i of Γ, there exists a subset $X_i \subset X$ such that the union of all the X_i is properly contained in X.*
(2) *For each nonzero $k \in \mathbb{Z}$ and vertices v_i, v_j joined by en edge, $v_i^k(X_j) \subset X_j$.*
(3) *For each nonzero $k \in \mathbb{Z}$ and vertices v_i, v_j not joined by en edge, $v_i^k(X_j) \subset X_i$.*
(4) *There exists $x_0 \in X \setminus \bigcup_{i \in V} X_i$ (V is the vertex set of Γ) such that $v_i^k(x_0) \in X_i$ for each nonzero $k \in \mathbb{Z}$.*

Then the G–action is faithful.

Proof of Lemma 6.4 We will apply Theorem 6.5 with $X = X(\Gamma, S)$ and $G = G(\Gamma')$. For $1 \le i \le s$, we identify c_i and $\mathbb{R}$ in an orientation-preserving way such that $\pi_{c_i}(\mathrm{id})$ corresponds to $0 \in \mathbb{R}$. Define $X_i^+ = \pi_{c_i}^{-1}\big([n_i - \frac{1}{2}, \infty)\big)$, $X_i^- = \pi_{c_i}^{-1}\big((-\infty, -\frac{1}{2}]\big)$ and $X_i = X_i^+ \cup X_i^-$. It clear that the identity element id does not lie in X_i for any i, so Theorem 6.5(1) is true. Each $v_i = \alpha_i g_i^{n_i} \alpha_i^{-1}$ translates c_i by n_i units, so (4) is also true with $x_0 = \mathrm{id}$.

If $\Delta(c_i)$ and $\Delta(c_j)$ are connected by an edge in $\mathcal{P}(\Gamma)$, then v_i stabilizes every hyperplane dual to v_j; thus $v_i^k(X_j) = X_j$, and (2) is true. If

$$d(\Delta(c_i), \Delta(c_j)) \geq 2, \tag{6-1}$$

then $\pi_{c_j}(c_i)$ is a point. Lemma 6.1 and $c_i \cap K \neq \varnothing$ yield that $\pi_{c_j}(c_i) \subset \pi_{c_j}(K) = c_j \cap K = [0, n_j - 1]$; thus

$$c_i \cap X_j = \varnothing. \tag{6-2}$$

Similarly, $c_i \cap X_j = \varnothing$. Let $h = \pi_{c_j}^{-1}\bigl(-\frac{1}{2}\bigr)$ be the boundary of X_j^-, and let N_h be the carrier of h. Then (6-1) implies that h has empty intersection with any hyperplane dual to c_i, and so does N_h. It follows from Lemma 6.1 that $\pi_{c_i}(h) = \pi_{c_i}(N_h) = p$ is a vertex in c_i. If $h_1 = \pi_{c_i}^{-1}\bigl(p - \frac{1}{2}\bigr)$ and $h_2 = \pi_{c_i}^{-1}\bigl(p + \frac{1}{2}\bigr)$ are two hyperplanes that pinch p, then $h \cap h_k = \varnothing$ for $k = 1, 2$. This and (6-2) yield $X_j^- \cap h_k = \varnothing$; hence $\pi_{c_i}(X_j^-) = p$ by Lemma 6.1. Similarly, $\pi_{c_i}(X_j^+) = p$, so

$$p = \pi_{c_i}(X_j) = \pi_{c_i}(c_j) \subset \pi_{c_i}(K) = c_i \cap K = [0, n_i - 1].$$

Note $(\pi_{c_i} \circ v_i^k)(X_j) = (v_i^k \circ \pi_{c_i})(X_j) = v_i^k(p) = p + k n_i$, so $v_i^k(X_j) \subset X_i$ for $k \neq 0$. □

The discussion in this subsection yields a well-defined map

$$\Theta_S \colon \mathrm{CN}(\Gamma, S) \to \{\text{finite-index RAAG subgroups of } G(\Gamma)\}.$$

The images of Θ_S are called *S–special subgroups of* $G(\Gamma)$. A subgroup of $G(\Gamma)$ is *special* if it is S–special for some standard generating set S of $G(\Gamma)$.

6.2 Rigidity of RAAG subgroups

In this subsection, we will assume $G(\Gamma')$ is a finite-index RAAG subgroup in $G(\Gamma)$ and $\mathrm{Out}(G(\Gamma))$ is finite. We will show that under such conditions, $G(\Gamma')$ must arise from the process described in the previous subsection. We will prove this in three steps. First we produce a convex subcomplex of $X(\Gamma, S)$ from $G(\Gamma')$. Then we will modify this convex subcomplex such that it is an element in $\mathrm{CN}(\Gamma, S)$. Thus we have defined a map from finite-index RAAG subgroups of $G(\Gamma)$ to elements in $\mathrm{CN}(\Gamma, S)$. In the last step, we show the map defined in Step 2 is an inverse to the map Θ_S defined in Section 6.1.

Also near the end of this subsection, we will leave several relatively long remarks which discuss relevant material in the literature. The reader can skip these remarks at first reading.

Recall that $\mathrm{Out}(G(\Gamma))$ is finite and $\mathrm{Out}(G(\Gamma'))$ is transvection-free (Theorem 5.3), so any two standard generating sets of $G(\Gamma)$ (or $G(\Gamma')$) differ by a sequence of conjugations or partial conjugations. Then given any two standard generating sets S and S_1 for $G(\Gamma)$, there is a canonical way to identify $\mathcal{P}(\Gamma, S)$ and $\mathcal{P}(\Gamma, S_1)$ (every S–geodesic is Hausdorff close to an S_1–geodesic). Thus we will write $\mathcal{P}(\Gamma)$ and $\mathcal{P}(\Gamma')$ and omit the generating set.

Lemma 6.6 *Let ϕ and s be as in the discussion before Lemma 5.5. Let $l \subset X(\Gamma)$ and $l' \subset X(\Gamma')$ be standard geodesics such that $\phi(v(l)) = v(l')$. Then $\phi \circ \pi_{\Delta(l)} = \pi_{\Delta(l')} \circ s$.*

Proof Pick standard geodesics $r \subset X(\Gamma)$ and $r' \subset X(\Gamma')$ such that $\phi(v(r)) = v(r')$; then $s(\Delta(r)) = \Delta(r')$ by Lemma 4.10 (recall that r is the intersection of maximal standard flats). Therefore, by the definition of $\pi_{\Delta(l)}$, it suffices to show $\phi \circ \pi_l(x) = \pi_{l'} \circ \phi(x)$ for any vertex $x \in X(\Gamma)$. Let y be a vertex such that $y \notin l$, and let $x = \pi_l(y)$. By Lemma 6.1, we can approximate $\overline{xy}$ by a combinatorial geodesic ω in the 1–skeleton of $\pi_l^{-1}(y)$; then no hyperplane could intersect both l and ω. Let $\{v_i\}_{i=0}^n$ be vertices in ω such that for $0 \le i \le n-1$, we have that each $[v_i, v_{i+1}]$ is a maximal subsegment of ω that is contained in a standard geodesic ($v_0 = x$ and $v_n = y$). Denote the corresponding standard geodesic by l_i. Then $\Delta(l) \neq \Delta(l_i)$ for all i. Let $u_i = \phi(v_i)$ and let l'_i be the standard geodesic such that $\phi(v(l_i)) = v(l'_i)$. Then $\overline{u_i u_{i+1}} \subset l'_i$ and $\Delta(l') \neq \Delta(l'_i)$ for all i; thus $\pi_{l'}(l'_i)$ is a point by Corollary 3.2, and $\pi_{l'}(u_i) = \pi_{l'}(u_j)$ for all $1 \le i, j \le n$. $\square$

Step 1 We produce a convex subcomplex of $X(\Gamma, S)$ from $G(\Gamma')$.

The left action $G(\Gamma) \curvearrowright G(\Gamma)$ induces $G(\Gamma') \curvearrowright G(\Gamma)$ and $G(\Gamma') \curvearrowright X(\Gamma, S)$. By choosing a standard generating set S' of $G(\Gamma')$, we have left action $G(\Gamma') \curvearrowright X(\Gamma', S')$. For $h \in G(\Gamma')$, we use ϕ_h, $\overline{\phi}_h$, s_h and $\overline{s}_h$ to denote the action of h on $G(\Gamma)$, $G(\Gamma')$, $\mathcal{P}(\Gamma)$ and $\mathcal{P}(\Gamma')$ respectively. Pick a $G(\Gamma')$–equivariant quasi-isometry $q\colon X(\Gamma, S) \to X(\Gamma', S')$ such that $q|_{G(\Gamma')} = \mathrm{Id}$. By Theorem 5.3 and Lemma 4.10, q induces surjective $G(\Gamma')$–equivariant maps $\phi\colon G(\Gamma) \to G(\Gamma')$ and $s\colon \mathcal{P}(\Gamma) \to \mathcal{P}(\Gamma')$. Note that ϕ depends on the choice of generating set S and S', and this flexibility comes from the automorphism groups of $G(\Gamma)$ and $G(\Gamma')$.

The key of Step 1 is to choose a "nice" standard generating set S' of $G(\Gamma')$ such that ϕ behaves like φ in Theorem 5.6.

Lemma 6.7 *By choosing a possibly different standard generating set S' for $G(\Gamma')$, we can assume the map ϕ satisfies $\phi(\mathrm{id}) = \mathrm{id}$, where* id *denotes the identity element in the corresponding group.*

Proof Assume $\phi(\mathrm{id}) = a \neq \mathrm{id}$; we claim if we change the generating set from S' to $aS'a^{-1}$, then the resulting ϕ will satisfy our requirement. By the construction of ϕ, it suffices to show for any maximal S'–flat F_1' such that $a \in F_1'$, there exists a maximal $aS'a^{-1}$–flat F_2' such that $\mathrm{id} \in F_2'$ and $d_H(F_1', F_2') < \infty$. Let us assume $F_1' = \{ag^k\}_{k\in\mathbb{Z}}$ for some $g \in S'$. Then $F_2' = \{(aga^{-1})^k\}_{k\in\mathbb{Z}}$ would satisfy the required condition. We can prove the general case in a similar way. □

Pick a standard geodesic $l \subset X(\Gamma, S)$; we want to flip the order of points of l in a $G(\Gamma')$–equivariant way such that (5-7) is true. We choose an order-preserving identification of $v(l)$ and $\mathbb{Z}$. Let $d = |\phi^{-1}(\phi(p)) \cap v(l)|$ where p is a vertex in $v(l)$. Let $\mathrm{Stab}(v(l))$ be the stabilizer of $v(l)$ under the action $G(\Gamma') \curvearrowright G(\Gamma)$. By the second paragraph of the proof of Theorem 5.6, d does not depend on the choice of p in $v(l)$, and $\mathrm{Stab}(v(l))$ acts on $v(l)$ in the same way as $d\mathbb{Z}$ acts on $\mathbb{Z}$ (recall that ϕ is $G(\Gamma')$–equivariant and the action of $G(\Gamma')$ on $G(\Gamma)$ is induced from the left action of $G(\Gamma)$ on itself).

We will write $\chi(l) = d$. If $\bar{l}$ and l are parallel, then $\chi(l) = \chi(\bar{l})$. Thus $\chi\colon \mathcal{P}(\Gamma) \to \mathbb{Z}$ is well defined. Since $\chi(l)$ only depends on how $\mathrm{Stab}(v(l))$ acts on $v(l)$, we see that χ does not depend on the standard generating set S'. However, χ descends to $\chi\colon S' \to \mathbb{Z}$ for any choice of S' by the $G(\Gamma')$–equivariance of ϕ.

Let $\phi(0) = a$. Then $\mathrm{Stab}(v(l))$ is generated by aha^{-1} for some $h \in S'$. By the same reasoning as in Lemma 6.7, we can assume $a = \mathrm{id}$. Let $S' = \{h_\lambda\}_{\lambda\in\Lambda}$. For each $h_\lambda \in S'$, we associated an integer n_λ as follows. If $h_\lambda h = hh_\lambda$, we set $n_\lambda = 0$. Now we consider the case where $h_\lambda h \neq hh_\lambda$. Let $l_\lambda' \subset X(\Gamma', S')$ be the standard geodesic that contains all powers of h_λ, and let $b_\lambda = \pi_{\Delta(l)} \circ s^{-1}(\Delta(l_\lambda'))$ ($\pi_{\Delta(l)}$ is the map in Lemma 6.2). Then n_λ is defined to be the unique integer such that $b_\lambda + n_\lambda d \in [0, d-1]$ (recall that $d = \chi(l)$). Define $f\colon S' \to G(\Gamma')$ by sending h_λ to $h^{n_\lambda} h_\lambda h^{-n_\lambda}$; then f extends to an automorphism of $G(\Gamma')$, and $S'' = \{f(h_\lambda)\}_{\lambda\in\Lambda}$ is also a standard generating set. Indeed, if $\Delta(l_{\lambda_1}')$ and $\Delta(l_{\lambda_2}')$ stay in the same connected component of $\mathcal{P}(\Gamma') \setminus \mathrm{St}(\Delta(l'))$, then $b_{\lambda_1} = b_{\lambda_2}$ by Lemma 6.2; hence $n_{\lambda_1} = n_{\lambda_2}$. It follows that f can be realized as a composition of partial conjugations.

Lemma 6.8 *Define ϕ_1 by replacing S' by S'' in the definition of ϕ. Then $\phi_1|_{v(l_1)}$ satisfies (5-7) for any standard geodesic $l_1 \subset X(\Gamma, S)$ with $\Delta(l_1) \in \{s_h(\Delta(l))\}_{h\in G(\Gamma')}$.*

Recall that for any $h \in G(\Gamma')$, we use s_h to denote the action of h on $\mathcal{P}(\Gamma)$.

Proof It suffices to show $\phi_1|_{v(l)}$ satisfies (5-7), and the rest follows from the $G(\Gamma')$–equivariance of ϕ_1. To show this, we only need to prove $\phi_1(i) = \mathrm{id}$ for any $i \in [0, d-1]$. Let Λ, b_λ and n_λ be as above.

We pick $i \in [0, d-1]$. Then there exists $\lambda \in \Lambda$ such that $b_\lambda + n_\lambda d = i$. By Lemma 6.6, $\phi(b_\lambda) = \mathrm{id}$; hence $\phi(i) = h^{n_\lambda}$. Let l_i be a standard geodesic such that $b_\lambda \in l_i$ and $d(\Delta(l_i), \Delta(l)) \geq 2$. Then there exists $h_{\lambda'} \in S'$ with $b_{\lambda'} = b_\lambda$ such that $\phi(v(l_i)) = \{h^k_{\lambda'}\}_{k\in\mathbb{Z}}$. Then $(\phi_h)^{n_\lambda}(v(l_i))$ is an S–geodesic passing through i, and $(\phi \circ (\phi_h)^{n_\lambda})(v(l_i)) = ((\bar{\phi}_h)^{n_\lambda} \circ \phi)(v(l_i)) = \{h^{n_\lambda} h^k_{\lambda'}\}_{k\in\mathbb{Z}}$. Note that

$$d_H\big(\{h^{n_\lambda} h^k_{\lambda'}\}_{k\in\mathbb{Z}}, \{(f(h_{\lambda'}))^k\}_{k\in\mathbb{Z}}\big) < \infty. \tag{6-3}$$

Now we look at the new map ϕ_1. Note that $\phi_1(0) = \mathrm{id}$ is still true. Moreover, (6-3) and Lemma 6.6 imply $\phi_1(i) = \mathrm{id}$. Thus the lemma follows. □

The next lemma basically says the above change-of-basis process does not significantly affect other geodesics.

Lemma 6.9 *Let r be a standard geodesic in $X(\Gamma, S)$ which satisfies the condition that $\Delta(r) \notin \{s_h(\Delta(l))\}_{h\in G(\Gamma')}$. Pick two different vertices $x_1, x_2 \in r$. If $\phi(x) = \phi(y)$, then $\phi_1(x) = \phi_1(y)$.*

Proof For $i = 1, 2$, let $r_i \subset X(\Gamma, S)$ be a standard geodesic containing x_i such that $d(\Delta(r_i), \Delta(r)) \geq 2$ for $i = 1, 2$. Let r' (resp. r'') be an S'–geodesic (resp. S''–geodesic) such that $\phi(v(r)) = v(r')$ (resp. $\phi_1(v(r)) = v(r'')$). Let $\alpha = \phi(x) = \phi(y)$. Then there exist elements h_λ, h_{λ_1} and h_{λ_2} in S' such that $\phi(v(r_i)) = \{\alpha h^k_{\lambda_i}\}_{k\in\mathbb{Z}}$ for $i = 1, 2$, and $r' = \{\alpha h^k_\lambda\}_{k\in\mathbb{Z}}$. Note that

$$h \neq h_\lambda, \quad h_{\lambda_1} \neq h_\lambda \quad \text{and} \quad h_{\lambda_2} \neq h_\lambda. \tag{6-4}$$

Recall that h is the generator of $\mathrm{Stab}(v(l))$. The first inequality of (6-4) follows from $\Delta(r) \notin \{s_h(\Delta(l))\}_{h\in G(\Gamma')}$.

It suffices to show there exist S''–geodesics r''_1 and r''_2 such that

$$d_H\big(\phi(v(r_i)), r''_i\big) < \infty \tag{6-5}$$

for $i = 1, 2$, and

$$\pi_{\Delta(r'')}(\Delta(r''_1)) = \pi_{\Delta(r'')}(\Delta(r''_2)), \tag{6-6}$$

then $\phi_1(x) = \phi_1(y)$ follows from Lemma 6.6. Define $r''_i = \{\alpha h^{-n_{\lambda_i}} (f(h_{\lambda_i}))^k\}_{k\in\mathbb{Z}}$; then (6-5) is immediate. Note that for any $a \in r'_1$ and $b \in r'_2$, we have

$$b = a \cdot (f(h_{\lambda_1}))^{k_1} \cdot h^{n_{\lambda_1} - n_{\lambda_2}} \cdot (f(h_{\lambda_2}))^{k_2}$$

for some $k_1, k_2 \in \mathbb{Z}$; then (6-6) follows from (6-4) and the definition of $\pi_{\Delta(r'')}$. □

Similarly, we can prove that if we change ϕ with respect to the conjugation $S' \to aS'a^{-1}$, then Lemma 6.9 is still true with r being an arbitrary standard geodesic.

By Lemma 6.8 and Lemma 6.9, we can apply the above change-of-basis procedure finitely many times to find an appropriate standard generating set S' of $G(\Gamma')$ such that the corresponding map ϕ satisfies (5-7) when restricted to any standard geodesic in $X(\Gamma, S)$. By the proof of Theorem 5.6, we can extend ϕ to a cubical map $\phi\colon X(\Gamma, S) \to X(\Gamma', S')$ such that combinatorial geodesics in $C(\Gamma, S)$ are mapped to combinatorial geodesics in $C(\Gamma', S')$. Thus $\phi^{-1}(\mathrm{id})$ is a combinatorially convex subcomplex. The subcomplex $\phi^{-1}(\mathrm{id})$ is also compact since $\phi^{-1}(\mathrm{id})$ contains finitely many vertices. Recall that combinatorial convexity in ℓ^1 metric and convexity in CAT(0) metric are the same for subcomplexes of CAT(0) cube complexes [33], so we have constructed a compact convex subcomplex $\phi^{-1}(\mathrm{id}) \subset X(\Gamma, S)$ from a given finite-index RAAG subgroup $G(\Gamma') \le G(\Gamma)$.

Step 2 We show $\phi^{-1}(\mathrm{id})$ can be assumed to be an element in $\mathrm{CN}(\Gamma, S)$.

For $K \subset G(\Gamma)$, denote the union of all standard geodesics in $X(\Gamma, S)$ that have nontrivial intersection with K by K^*. K is S*–convex* if and only if K is the vertex set of some convex subcomplex in $X(\Gamma, S)$. Now we return to ϕ. By Step 1, we can assume $\phi(\mathrm{id}) = \mathrm{id}$, and $\phi^{-1}(y)$ is S–convex for any $y \in G(\Gamma')$.

Step 2.1 Let $\{l_i\}_{i=1}^{q}$ be the collection of standard geodesics passing through id, and let $\Lambda_1 = \{\mathrm{id}\}$. Let $I\colon G(\Gamma) \to \ell^1(v(\mathcal{P}(\Gamma, S)))$ and $I_{\Delta(l)}\colon G(\Gamma) \to \mathbb{Z}^{\Delta(l)}$ be the maps defined in Section 6.1. Since $v(l_i)$ and $v(l_j)$ are in different $G(\Gamma')$–orbits for $i \neq j$, by Lemma 6.8 and Lemma 6.9, we can apply the change-of-basis procedure in Step 1 to find a standard generating set S' for $G(\Gamma')$ such that for each $1 \le i \le q$,

$$I^{-1}_{\Delta(l_i)}\big([0, \chi(l_i) - 1]\big) \cap v(l_i) \subset \phi^{-1}(\mathrm{id}). \tag{6-7}$$

Step 2.2 Let $\Lambda_2 = \Lambda_1^* \cap \phi^{-1}(\mathrm{id})$. Pick a vertex $x \in \Lambda_2 \setminus \Lambda_1$ (if such x does not exist, then our process terminates). Let l be a standard geodesic such that $x \in l$. If l is parallel to some l_i in Step 2.1, then (6-7) with l_i replaced by l is automatically true without any modification on S', because both I and ϕ respect the product structure of P_{l_i}. If l is not parallel to any l_i, then $I_{\Delta(l)}(x) = 0$. Moreover, $\Delta(l)$ is not in the $G(\Gamma')$–orbits of the $\Delta(l_i)$, so we can modify S' as before such that both (6-7) and $I^{-1}_{\Delta(l)}([0, \chi(l) - 1]) \cap v(l) \subset \phi^{-1}(\mathrm{id})$ are true. We deal with other standard geodesics passing through x and other points in $\Lambda_2 \setminus \Lambda_1$ in a similar way.

Step 2.3 Let $\Lambda_3 = \Lambda_2^* \cap \phi^{-1}(\mathrm{id})$. For each vertex in $\Lambda_3 \setminus \Lambda_2$, we repeat the procedure in Step 2.2. Then we can define $\Lambda_4, \Lambda_5, \dots$. Since $|\phi^{-1}(\mathrm{id})|$ is finite and this number does not change after adjusting S', our procedure must terminate after finitely many steps. Since $\phi^{-1}(\mathrm{id})$ remains connected in each step, once the procedure terminates, we must have already dealt with each point in $\phi^{-1}(\mathrm{id})$ and each standard geodesic passing through each point in $\phi^{-1}(\mathrm{id})$. By construction, the resulting ϕ satisfies $\mathrm{id} \in \phi^{-1}(\mathrm{id})$ and $I^{-1}_{\Delta(l)}([0, \chi(l)-1]) \cap v(l) \subset \phi^{-1}(\mathrm{id})$ for each standard geodesic l which intersects $\phi^{-1}(\mathrm{id})$. Thus $\phi^{-1}(\mathrm{id})$ is nonnegative.

Note that the sets Λ_i actually do not depend on the map ϕ from step $i-1$. They only depend on the map $\chi\colon v(\mathcal{P}(\Gamma)) \to \mathbb{Z}$. Thus the nonnegative subset $\phi^{-1}(\mathrm{id}) \subset G(\Gamma)$ produced above depends only on S and the subgroup $G(\Gamma') \le G(\Gamma)$. Then we have a well-defined map

$$\Xi_S\colon \{\text{Finite-index RAAG subgroups of } G(\Gamma)\} \to \mathrm{CN}(\Gamma, S).$$

Step 3 We show Ξ_S is an inverse to the map Θ_S defined in Section 5.2.

First we prove $\Theta_S \circ \Xi_S = \mathrm{Id}$. Let $K = \Xi_S(G(\Gamma'))$. Let S' be the corresponding standard generating set for $G(\Gamma')$ and let $\phi\colon G(\Gamma) \to G(\Gamma')$ be the corresponding map. We find a maximal collection of standard geodesics $\{c_i\}_{i=1}^s$ such that $c_i \cap K \neq \varnothing$ for all i and $\Delta(c_i) \neq \Delta(c_j)$ for any $i \neq j$. Let $n_i = \chi(c_i)$, and let $g_i \in S$ be the label of edges in c_i. Suppose $\alpha_i = \pi_{c_i}(\mathrm{id})$ where $\pi_{c_i}\colon X(\Gamma, S) \to c_i$ is the CAT(0) projection. Then it suffices to prove the following lemma.

Lemma 6.10 $$S' = \{\alpha_i g_i^{n_i} \alpha_i^{-1}\}_{i=1}^s.$$

Proof Pick $h \in S'$ and let $c_h \subset X(\Gamma', S')$ be the standard geodesic containing id and h. Then there exists a unique i such that $\phi(v(c_i)) = c_h$. To see this, let c be a standard geodesic in $X(\Gamma, S)$ such that $s(\Delta(c)) = \Delta(c_h)$. Then $\phi(v(c))$ and c_h are parallel and there exists $u \in G(\Gamma')$ which sends $\phi(v(c))$ to $v(c_h)$. Thus $\phi \circ \phi_u(v(c)) = v(c_h)$ by (5-4), where ϕ_u is defined in the beginning of Step 1. Note that $\phi_u(v(c))$ has nontrivial intersection with K. We choose c_i to be the geodesic parallel to $\phi_u(v(c))$. Then $\phi(v(c_i)) = v(c_h)$.

For any standard geodesic c_i' parallel to c_i, we have that $\phi(c_i')$ is parallel to c_h, so $h \in \mathrm{Stab}(v(\phi(c_i'))) = \mathrm{Stab}(v(c_i'))$. It follows that ϕ_h stabilizes the parallel set P_{c_i} and acts by translation along the c_i–direction. Note that $(I_{\Delta(c_i)} \circ \phi_h)(x) = I_{\Delta(c_i)}(x) + \chi(c_i)$ for any $x \in v(P_{c_i})$, so $h = \phi_h(\mathrm{id}) = \alpha_i g_i^{n_i} \alpha_i^{-1}$ and the claim follows. □

It remains to show $\Xi_S \circ \Theta_S = \mathrm{Id}$. The following result implies this.

Lemma 6.11 *Let Γ be an arbitrary finite simplicial graph. Pick a standard generating set S for $G(\Gamma)$ and $K \in \mathrm{CN}(\Gamma, S)$. Let $G(\Gamma') = \Theta_S(K)$ and let S' be the corresponding generating set. Suppose $q\colon G(\Gamma) \to G(\Gamma')$ is a $G(\Gamma')$–equivariant quasi-isometry such that $q|_{G(\Gamma')}$ is the identity map. Then:*

(1) *q induces a simplicial isomorphism $q_*\colon \mathcal{P}(\Gamma, S) \to \mathcal{P}(\Gamma', S')$.*
(2) *q_* induces a $G(\Gamma')$–equivariant retraction $r\colon G(\Gamma) \to G(\Gamma')$ such that r sends every S–flat to an S'–flat.*
(3) *r extends to a surjective cubical map $r\colon X(\Gamma, S) \to X(\Gamma', S')$ with $r^{-1}(\mathrm{id}) = K$. In particular, the vertex set of K is the strict fundamental domain for the left action $G(\Gamma') \curvearrowright G(\Gamma)$.*

Proof It suffices to prove the case when Γ does not admit a nontrivial join decomposition and Γ is not a point.

By the construction of Θ_S, we know the q–image of any S–flat which intersects K is Hausdorff close to an S'–flat which contains the identity. Moreover, if the S–flat is maximal, then the corresponding S'–flat is unique. Since $G(\Gamma') \cdot v(K) = G(\Gamma)$, the equivariance of q implies the q–image of every S–flat is Hausdorff close to an S'–flat. Since q is a quasi-isometry, images of parallel S–geodesics are Hausdorff closed to each other. This induces $q_*\colon \mathcal{P}(\Gamma, S) \to \mathcal{P}(\Gamma', S')$, which is injective since q is a quasi-isometry, and surjective by the $G(\Gamma')$–equivariance.

Pick $x \in G(\Gamma)$, and let $\{F_i\}_{i\in I}$ be the collection of maximal S–flats containing x. For each i, let F_i' be the unique maximal S'–flat such that $d_H(q(F_i), F_i') < \infty$. Note that $\bigcap_{i\in I} F_i = x$ by our assumption on Γ. So $\bigcap_{i\in I} F_i'$ is either empty or one point. Note that if $x \in K$, then $\bigcap_{i\in I} F_i' = \mathrm{id}$. The equivariance of q_* implies that for every x, $\bigcap_{i\in I} F_i'$ is a point, which is defined to be $r(x)$. It is clear that $v(K) \subset r^{-1}(\mathrm{id})$, but $|G(\Gamma)\colon G(\Gamma')| \le |v(K)|$, so $v(K) = r^{-1}(\mathrm{id})$. It follows that $v(K)$ is the strict fundamental domain for the left action of $G(\Gamma')$, and r is a $G(\Gamma')$–equivariant map which maps $v(K)$ to id.

Note that $r(\mathrm{id}) = \mathrm{id}$. Then the $G(\Gamma')$–equivariance of r implies $r(g) = g$ for any $g \in G(\Gamma') \subset G(\Gamma)$. Thus r is a retraction. Similarly, by using the $G(\Gamma')$–equivariance of r, we deduce that r sends every S–flat that intersects K to an S'–flat passing through the identity element of $G(\Gamma')$. Thus r sends every S–flat to an S'–flat by the equivariance of r. It is easy to see r extends to a cubical map $r\colon X(\Gamma, S) \to X(\Gamma', S')$ such that $r^{-1}(\mathrm{id}) = K$. □

Remark 6.12 We can generalize some of the results in Lemma 6.11 to infinite convex subcomplexes of $X(\Gamma, S)$. A convex subcomplex $K \subset X(\Gamma, S)$ is *admissible* if for any standard geodesic l, the CAT(0) projection $\pi_l(K)$ is either a finite interval or the whole of l (a ray is not allowed). Let $\{l_\lambda\}_{\lambda\in\Lambda}$ be a maximal collection of standard geodesics such that (1) $l_\lambda \cap K \neq \varnothing$; (2) l_λ and $l_{\lambda'}$ are not parallel for $\lambda \neq \lambda'$; (3) $\pi_{l_\lambda}(K)$ is a finite interval. For each l_λ, let $\alpha_\lambda \in G(\Gamma)$ be an element which translates along l_λ with translation length $= 1 + \mathrm{length}(\pi_{l_\lambda}(K))$. Let G_K be the subgroup generated by $S' = \{\alpha_\lambda\}_{\lambda\in\Lambda}$. If K is admissible, we can prove $G_K \cdot v(K) = G(\Gamma)$ as before. Moreover, for any finite subset $S_1' \subset S'$, the subgroup G_1 generated by S_1' is a right-angled Artin group, and $G_1 \hookrightarrow G_K$ is an isometric embedding with respect to the word metric. We can define an S'–flat as before and view each vertex of G_K as a 0–dimensional S'–flat.

Now we show $v(K)$ is a strict fundamental domain for the action $G_K \curvearrowright G(\Gamma)$. It suffices to show $\alpha(K) \cap K = \varnothing$ for each nontrivial $\alpha \in G_K$. We can assume there is a right-angled Artin group G_1 such that $\alpha \in G_1 \subset G_K$. Let $\alpha = w_1 w_2 \cdots w_n$ be a canonical form of α; see [16, Section 2.3]. Then:

(1) Each w_i belongs to an abelian standard subgroup of G_1.

(2) For each i, let $w_i = r_{i,1}^{k_{i,1}} r_{i,2}^{k_{i,2}} \cdots r_{i,n_i}^{k_{i,n_i}}$ ($r_{i,j} \in S'$). Then for each $r_{i+1,j}$ ($1 \le j \le n_{i+1}$), there exists $r_{i,j'}$ which does not commute with $r_{i+1,j}$.

We associate each generator $r_{i,j}$ with a subset $X_{i,j} \subset X(\Gamma, S)$ as in the proof of Lemma 6.4, and claim there exists j with $1 \le j \le n_1$ such that $\alpha(K) \subset X_{1,j}$; then $\alpha(K) \cap K = \varnothing$ follows. We prove by induction on n and assume $w_2 w_3 \cdots w_n(K) \subset X_{2,j'}$. By (2), there is $r_{1,j}$ such that $r_{1,j}$ and $r_{2,j'}$ does not commute, so we have $r_{1,j}^{k_{1,j}}(X_{2,j'}) \subset X_{1,j}$. Moreover, by (1), $r_{1,h}^{k_{1,h}}(X_{1,j}) = X_{1,j}$ for $h \neq j$, so $\alpha(K) \subset w_1(X_{2,j'}) \subset X_{1,j}$.

Now we can define a G_K–equivariant map $r\colon G(\Gamma) \to G_K$ by sending $v(K)$ to the identity of G_K. We prove as before that r maps S–flats to (possibly lower-dimensional or 0–dimensional) S'–flats; thus r is 1–Lipschitz with respect to the word metric. Let $i\colon G_K \hookrightarrow G(\Gamma)$ be the inclusion. Then by the equivariance of r, the composition $r \circ i$ is a left translation of G_K. In particular, if K contains the identity, then r is a retraction. It follows that if S' is finite, then i is a quasi-isometric embedding.

Note that a related construction in the case of right-angled Coxeter groups has been discussed in [34]. By taking larger and larger convex compact subcomplexes of $X(\Gamma, S)$, we know $G(\Gamma)$ is residually finite. Moreover, pick $\beta \in \mathrm{Stab}(K) \subset G(\Gamma)$. By definition of S', we have $S' = \beta S' \beta^{-1}$, so $\mathrm{Stab}(K)$ normalizes G_K. Now we have obtained a direct proof of the fact that every word-quasiconvex subgroup of a finitely generated right-angled Artin group is separable (Theorem F of [34]) by using the above discussion together with the outline in Section 1.5 of [34].

The following result follows readily from the above discussion.

Theorem 6.13 *Let $G(\Gamma)$ be a RAAG with $\mathrm{Out}(G(\Gamma))$ finite. We pick a standard generating set S for $G(\Gamma)$. Then there is a one-to-one correspondence between nonnegative convex compact subcomplexes of $X(\Gamma, S)$ that contain the identity and finite-index RAAG subgroups of $G(\Gamma)$. In particular, these subgroups are generated by conjugates of powers of elements in S.*

In particular, Theorem 1.3 in the introduction follows from Theorem 6.13.

Remark 6.14 If we drop the finite automorphism group assumption in the above theorem, then there exist a RAAG $G(\Gamma_1)$ and its finite index RAAG subgroup $G(\Gamma_2)$ such that $G(\Gamma_2)$ is not isomorphic to any special subgroup of $G(\Gamma_1)$. To see this, let $G(\Gamma_1)$ be a right-angled Artin group such that $\mathrm{Out}(G(\Gamma_1))$ is transvection-free. Then Lemma 6.11 and Theorem 3.20 imply each special subgroup of $G(\Gamma_1)$ does not admit a nontrivial transvection in its outer automorphism group. Let Γ_1 and Γ_2 be the graphs in Example 3.22. Then $G(\Gamma_2)$ is a right-angled Artin subgroup of $G(\Gamma_1)$, and there are nontrivial transvections in $\mathrm{Out}(G(\Gamma_2))$. Thus $G(\Gamma_2)$ is not isometric to any special subgroup of $G(\Gamma_1)$.

Remark 6.15 Pick $G(\Gamma)$ such that $\mathrm{Out}(G(\Gamma))$ is finite; then Theorem 6.13 can be used to show a certain subgroup of $G(\Gamma)$ is not a RAAG. For example, let $\{v_i\}_{i=1}^k$ be a subset of some standard generating set for $G(\Gamma)$. We define a homomorphism $h\colon G(\Gamma)\to\mathbb{Z}/2$ by sending each v_i to the nontrivial element in $\mathbb{Z}/2$ and killing all other generators. Then $\ker(h)$ is a RAAG if and only if $k=1$. One can compare this example to Example 3.22.

Remark 6.16 It is shown in [42, Theorem 2] that if $F(\Gamma')$ embeds into $\mathcal{P}(\Gamma)$ as a full subcomplex, then there exists a monomorphism $G(\Gamma')\hookrightarrow G(\Gamma)$. This result can be recovered by our previous discussion as follows. Let Γ be an arbitrary finite simplicial graph. Let S be a standard generating set for $G(\Gamma)$. For any vertex $w\in\mathcal{P}(\Gamma)$, let $\alpha_w\in G(\Gamma)$ be a conjugate of some element in S such that $\alpha_w(l)=l$ for every standard geodesic $l\subset X(\Gamma,S)$ with $\Delta(l)=w$.

Suppose $M\subset\mathcal{P}(\Gamma,S)$ is a compact full subcomplex and Γ' is the 1–skeleton of M. Denote the vertex set of M by $\{w_i\}_{i=1}^n$, and let l_i be a standard geodesic with $\Delta(l_i)=w_i$. We identify each l_i in an orientation-preserving way with $\mathbb{R}$ such that $0\in\mathbb{R}$ is identified with $\pi_{l_i}(\mathrm{id})\subset l_i$, where π_{l_i} is the CAT(0) projection to l_i and id is the identity element of $G(\Gamma)$.

For $1\le i\le n$, define $\Lambda_i=\{1\le j\le n\mid d(w_i,w_j)\ge 2\}$. For each i, we define a pair of integers a_i and k_i as follows. If $\Lambda_i\neq\varnothing$, then let $[a_i,a_i+k_i]\subset\mathbb{R}$ be the minimal interval such that $\bigcup_{j\in\Lambda_i}\pi_{l_i}(l_j)\subset[a_i,a_i+k_i]$ (recall that l_i is identified with $\mathbb{R}$). If $\Lambda_i=\varnothing$, then we pick an arbitrary a_i and set $k_i=0$. Define $X_i=\pi_{c_i}^{-1}\bigl((-\infty,a_i-\frac{1}{2}]\bigr)\cup\pi_{c_i}^{-1}\bigl([a_i+k_i+\frac{1}{2},\infty)\bigr)$. Then by construction, $X_i\cap X_j=\varnothing$ for i,j satisfying $d(w_i,w_j)\ge 2$. Using the argument in Section 6.1, we can show the subgroup generated by $S'=\{\alpha_{w_i}^{k_i+1}\}_{i=1}^n$ is a RAAG with defining graph Γ'.

At this point it is natural to ask the following question.

Question 6.17 *Let S be a standard generating set of $G(\Gamma)$, and let S' be a finite collection of elements of the form $\alpha r^k \alpha^{-1}$, where $r \in S$, $k \in \mathbb{Z}$ and $\alpha \in G(\Gamma)$. Suppose G is the subgroup generated by S'. Is G a right-angled Artin group?*

6.3 Generalized star extension

Our goal in this subsection is to find an algorithm to determine whether $G(\Gamma)$ and $G(\Gamma')$ are quasi-isometric or not, given that $\mathrm{Out}(G(\Gamma))$ is finite.

For a convex subcomplex $E \subset X(\Gamma)$, we denote the full subcomplex in $\mathcal{P}(\Gamma, S)$ spanned by $\{\Delta(l_\lambda)\}_{\lambda\in\Lambda}$ by $\widehat{E}$, where $\{l_\lambda\}_{\lambda\in\Lambda}$ is the collection of standard geodesics in $X(\Gamma)$ with $l_\lambda \cap E \neq \varnothing$.

Now we describe a process to construct a graph Γ' from Γ such that $G(\Gamma')$ is isomorphic to a special subgroup of $G(\Gamma)$. Let $\Gamma_1 = \Gamma$, and let K_1 be one point. We will construct a pair (Γ_i, K_i) inductively such that:

(1) K_i is a compact CAT(0) cube complex, and there is a cubical embedding $f\colon K_i \to X(\Gamma)$ such that $f(K_i)$ is convex in $X(\Gamma)$.

(2) Γ_i is a finite simplicial graph, and there is a simplicial isomorphism $g\colon F(\Gamma_i) \to \widehat{f(K_i)}$.

Note that these assumptions are true for $i = 1$.

We associate each edge $e \subset K_i$ with a vertex in Γ_i, denoted by v_e, as follows. Let l_e be the standard geodesic in $X(\Gamma)$ that contains $f(e)$. We define $v_e := g^{-1}(\Delta(l_e))$. Each vertex $x \in K_i$ can be associated with a full subcomplex $\Phi(x) \subset F(\Gamma_i)$ defined by $\Phi(x) = g^{-1}(\widehat{x})$.

To define (Γ_{i+1}, K_{i+1}), pick a vertex $v \in \Gamma_i$, and let $\{x_j\}_{j=1}^m$ be the collection of vertices in K_i such that $v \in \Phi(x_j)$. Then $\{f(x_j)\}_{j=1}^k$ are exactly the vertices in $P_l \cap f(K_i)$, where l is a standard geodesic such that $\Delta(l) = g(v)$. Let L be the convex hull of $\{x_j\}_{j=1}^m$ in K_i. Then $e \subset L$ for any edge $e \subset K_i$ with $v_e = v$.

Since $f(L) = P_l \cap f(K_i)$, the natural product decomposition $P_l \cong l \times l^{\perp}$ induces a product decomposition of $L = h \times [0, a]$. Note that it is possible that $a = 0$, and $a > 0$ if and only if there exists an edge $e \subset K_i$ with $v_e = v$. If $a > 0$, then h is isomorphic to the hyperplane dual to e, and for any edge $e' \in K_i$ with $v_{e'} = v$, the projection of e' to the interval factor $[0, a]$ is an edge.

Let $L_i = h \times \{a\} \subset L$, and let $M_i = \bigcup_{x\in L_i} \Phi(x)$ (where x is a vertex). We define $F(\Gamma_{i+1})$ to be the simplicial complex obtained by gluing $F(\Gamma_i)$ and M_i along $\mathrm{St}(v, M_i)$ (see Section 2.1 for the notation), and define K_{i+1} to be the CAT(0) cube complex obtained by gluing K_i and $L_i \times [0, 1]$ along L_i. One readily verifies that one can extend f to a cubical embedding $f'\colon K_{i+1}\to X(\Gamma)$ such that $f'(K_{i+1})$ is convex. This also induces an isomorphism $g'\colon F(\Gamma_{i+1})\to \widehat{K}_{i+1}$ which is an extension of g.

By construction, each $G(\Gamma_i)$ is isomorphic to a special subgroup of $G(\Gamma)$; moreover, the associated convex subcomplex of this special subgroup is K_i. Also note that the above induction process actually does not depend on knowing what $X(\Gamma)$ is. Thus it also provides a way to construct convex subcomplexes of $X(\Gamma)$ by hand.

The above process of obtaining (Γ_{i+1}, K_{i+1}) from (Γ_i, K_i) is called a *generalized star extension* (GSE) at v. Note that the following are equivalent:

(1) $\Gamma_i \subsetneq \Gamma_{i+1}$.

(2) $P_l \subsetneq X(\Gamma)$, where l is the standard geodesic in $X(\Gamma)$ such that $\Delta(l)=g(v)$.

(3) $\mathrm{St}(\pi(g(v)))\subsetneq F(\Gamma)$, where $\pi\colon \mathcal{P}(\Gamma)\to F(\Gamma)$ is the natural label-preserving projection defined in (4-1).

A GSE is *nontrivial* if $\Gamma_i \subsetneq \Gamma_{i+1}$. If Γ is not a clique, then at each stage, there exists a vertex $v\in\Gamma_i$ such that the GSE at v is nontrivial.

Lemma 6.18 *Suppose $G(\Gamma')$ is isomorphic to a special subgroup of $G(\Gamma)$. Then we can construct Γ' from Γ by using finitely many* GSE*s.*

Proof Let Θ_S and $\mathrm{CN}(\Gamma,S)$ be the objects defined in Section 6.1. Suppose $G(\Gamma')$ is isomorphic to $\Theta_S(K)$ for $K\in \mathrm{CN}(\Gamma,S)$. We define a sequence of convex subcomplexes in K by induction. Let K_1 be the identity element in $G(\Gamma)$. Suppose K_i is already defined. If $K_i=K$, then the induction terminates. If $K_i\subsetneq K$, pick an edge $e_i\subset K$ such that $e_i\cap K_i$ is a vertex and let K_{i+1} be the convex hull of $K_i\cup e_i$. Let $\{K_i\}_{i=1}^s$ be the resulting collection of convex subcomplexes. An alternative way of describing K_{i+1} is the following. If h_i is the hyperplane in K dual to e_i, and N_i is the carrier of h_i in K, then $h_i\cap K_i=\varnothing$ by the convexity of K_i. Thus $K_i\cap N_i$ is disjoint from h_i. Hence there is a copy of $(K_i\cap N_i)\times[0,1]$ inside N_i, which is denoted by M_i. Then $K_{i+1}=K_i\cup M_i$. Now one readily verifies that one can obtain $(\widehat{K}_{i+1},K_{i+1})$ from $(\widehat{K}_i,K_i)$ by a GSE. □

The above construction gives rise to an algorithm to detect whether $G(\Gamma')$ is isomorphic to a special subgroup of $G(\Gamma)$. If there are n vertices in Γ', then Γ' can be obtained from Γ by at most n nontrivial GSEs. So we can start with Γ, enumerate all possible n–step nontrivial GSEs from Γ, and compare each resulting graph with Γ'. By Theorem 5.7 and Theorem 6.13, we have the following result.

Theorem 6.19 *If $\mathrm{Out}(G(\Gamma))$ is finite, then $G(\Gamma')$ is quasi-isometric to $G(\Gamma)$ if and only if Γ' can be obtained from Γ by finitely many* GSE*s. In particular, there is an algorithm to determine whether $G(\Gamma')$ and $G(\Gamma)$ are quasi-isometric.*

Note that a GSE gives rise to a pair (Γ_i, K_i). If one does not care about the associated convex subcomplex K_i, then there is a simpler description of GSE when $\mathrm{Out}(G(\Gamma))$ is finite. Suppose we have already obtained $F(\Gamma_i)$ together with a finite collection of full subcomplexes $\{G_\lambda\}_{\lambda\in\Lambda_i}$ such that:

(1) $\{G_\lambda\}_{\lambda\in\Lambda_i}$ is a covering of $F(\Gamma_i)$.

(2) Each G_λ is isomorphic to $F(\Gamma)$.

When $i=1$, we pick the trivial cover of $F(\Gamma)$ by itself. To construct Γ_{i+1}, pick a vertex $v\in F(\Gamma_i)$, let $\Lambda_v=\{\lambda\in\Lambda_i \mid v\in G_\lambda\}$ and let $\Gamma_v=\bigcup_{\lambda\in\Lambda_v} G_\lambda$. Suppose $\{C_j\}_{j=1}^m$ is the collection of connected components of $\Gamma_v\setminus \mathrm{St}(v,\Gamma_v)$, and suppose $C_j'=C_j\cup \mathrm{St}(v,\Gamma_v)$. Then $F(\Gamma_{i+1})$ is defined by gluing C_1' and $F(\Gamma_i)$ along $\mathrm{St}(v,\Gamma_v)$, and Γ_{i+1} is the 1–skeleton of $F(\Gamma_{i+1})$.

Lemma 6.20 *Suppose* $\mathrm{Out}(G(\Gamma))$ *is finite. Then the above simplified process is consistent with* GSE*.*

Proof We assume inductively that there is a CAT(0) cube complex K_i such that the two induction assumptions for GSE are satisfied; moreover, $\{G_\lambda\}_{\lambda\in\Lambda_i}$ coincides with $\{\Phi(x)\}_{x\in K_i}$ (where x is a vertex). Let $L=h\times[0,a]$ be as before and let $L_j=h\times\{j\}\subset L$ for each integer $j\in[0,a]$. It suffices to show there is a one-to-one correspondence between $\{L_j\}_{j=0}^a$ and $\{C_j'\}_{j=1}^m$ such that for each j, there exists a unique j' with $\widehat{f(L_j)}=g(C_{j'}')$. Pick adjacent vertices $x_1,x_2\in f(L_j)$ and let $\bar w\in\Gamma$ be the label of edge $\overline{x_1x_2}$. Suppose $\bar v=\pi(g(v))$. Then $d(\bar w,\bar v)=1$. Since $\mathrm{Out}(G(\Gamma))$ is finite, the orthogonal complement of $\bar w$ satisfies $\bar w^\perp\not\subseteq \mathrm{St}(\bar v)$. Then there is a vertex $\bar u\in\bar w^\perp$ such that $d(\bar u,\bar v)=2$. The lifts of $\bar u$ in $\hat x_1$ and $\hat x_2$ are the same point, so $(\hat x_1\cap\hat x_2)\setminus\mathrm{St}(g(v))$ contains a vertex. Since $F(\Gamma)$ does not have separating closed stars, $\hat x_i\setminus\mathrm{St}(g(v))$ is connected for $i=1,2$. Thus $(\hat x_1\cap\hat x_2)\setminus\mathrm{St}(g(v))$ is connected. It follows that $\widehat{f(L_j)}\setminus\mathrm{St}(g(v))$ is connected. Moreover, from Lemma 4.9, $\widehat{f(L_{j_1})}\setminus\mathrm{St}(g(v))$ and $\widehat{f(L_{j_2})}\setminus\mathrm{St}(g(v))$ are in different components of $\mathcal{P}(\Gamma)\setminus\mathrm{St}(g(v))$ when $j_1\neq j_2$, so there exists a unique j' such that $\widehat{f(L_j)}=g(C_{j'}')$. □

References

[1] **A Abrams**, **N Brady**, **P Dani**, **M Duchin**, **R Young**, *Pushing fillings in right-angled Artin groups*, J. Lond. Math. Soc. 87 (2013) 663–688 MR

[2] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR

[3] **H Baik**, **S-h Kim**, **T Koberda**, *Right-angled Artin subgroups of the C^∞ diffeomorphism group of the real line*, preprint (2014) arXiv

[4] **J Behrstock**, **R Charney**, *Divergence and quasimorphisms of right-angled Artin groups*, Math. Ann. 352 (2012) 339–356 MR

[5] **J A Behrstock**, **T Januszkiewicz**, **W D Neumann**, *Quasi-isometric classification of some high dimensional right-angled Artin groups*, Groups Geom. Dyn. 4 (2010) 681–692 MR

[6] **J Behrstock**, **B Kleiner**, **Y Minsky**, **L Mosher**, *Geometry and rigidity of mapping class groups*, Geom. Topol. 16 (2012) 781–888 MR

[7] **J A Behrstock**, **W D Neumann**, *Quasi-isometric classification of graph manifold groups*, Duke Math. J. 141 (2008) 217–240 MR

[8] **M Bestvina**, **N Brady**, *Morse theory and finiteness properties of groups*, Invent. Math. 129 (1997) 445–470 MR

[9] **M Bestvina**, **B Kleiner**, **M Sageev**, *The asymptotic geometry of right-angled Artin groups, I*, Geom. Topol. 12 (2008) 1653–1699 MR

[10] **M Bestvina**, **B Kleiner**, **M Sageev**, *Quasiflats in* CAT(0) *2–complexes*, Algebr. Geom. Topol. 16 (2016) 2663–2676 MR

[11] **N Brady**, **J Meier**, *Connectivity at infinity for right angled Artin groups*, Trans. Amer. Math. Soc. 353 (2001) 117–132 MR

[12] **N Brady**, **T Riley**, **H Short**, *The geometry of the word problem for finitely generated groups*, Birkhäuser, Basel (2007) MR

[13] **M R Bridson**, **A Haefliger**, *Metric spaces of non-positive curvature*, Grundl. Math. Wissen. 319, Springer (1999) MR

[14] **P-E Caprace**, **N Monod**, *Isometry groups of non-positively curved spaces: structure theory*, J. Topol. 2 (2009) 661–700 MR

[15] **P-E Caprace**, **M Sageev**, *Rank rigidity for* CAT(0) *cube complexes*, Geom. Funct. Anal. 21 (2011) 851–891 MR

[16] **R Charney**, *An introduction to right-angled Artin groups*, Geom. Dedicata 125 (2007) 141–158 MR

[17] **R Charney**, **M W Davis**, *Finite* $K(\pi,1)$*s for Artin groups*, from "Prospects in topology" (F Quinn, editor), Ann. of Math. Stud. 138, Princeton Univ. Press (1995) 110–124 MR

[18] **R Charney**, **M Farber**, *Random groups arising as graph products*, Algebr. Geom. Topol. 12 (2012) 979–995 MR

[19] **M T Clay**, **C J Leininger**, **J Mangahas**, *The geometry of right-angled Artin subgroups of mapping class groups*, Groups Geom. Dyn. 6 (2012) 249–278 MR

[20] **C B Croke**, **B Kleiner**, *Spaces with nonpositive curvature and their ideal boundaries*, Topology 39 (2000) 549–556 MR

[21] **M W Davis**, *Groups generated by reflections and aspherical manifolds not covered by Euclidean space*, Ann. of Math. 117 (1983) 293–324 MR

[22] **M B Day**, *Finiteness of outer automorphism groups of random right-angled Artin groups*, Algebr. Geom. Topol. 12 (2012) 1553–1583 MR

[23] **C Droms**, *Isomorphisms of graph groups*, Proc. Amer. Math. Soc. 100 (1987) 407–408 MR

[24] **M J Dunwoody**, *The accessibility of finitely presented groups*, Invent. Math. 81 (1985) 449–457 MR

[25] **A Eskin**, *Quasi-isometric rigidity of nonuniform lattices in higher rank symmetric spaces*, J. Amer. Math. Soc. 11 (1998) 321–361 MR

[26] **A Eskin**, **B Farb**, *Quasi-flats and rigidity in higher rank symmetric spaces*, J. Amer. Math. Soc. 10 (1997) 653–692 MR

[27] **B Farb**, **M Handel**, *Commensurations of* $\mathrm{Out}(\mathrm{F}_n)$, Publ. Math. Inst. Hautes Études Sci. 105 (2007) 1–48 MR

[28] **V N Gerasimov**, *Semi-splittings of groups and actions on cubings*, from "Algebra, geometry, analysis and mathematical physics" (Y G Reshetnyak, L A Bokut', S K Vodop'yanov, I A Taĭmanov, editors), Izdat. Ross. Akad. Nauk Sib. Otd. Inst. Mat., Novosibirsk (1997) 91–109 MR In Russian; translated as *Fixed-point-free actions on cubings*, Siberian Adv. Math. 8 (1998) 36–58

[29] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR

[30] **M F Hagen**, *Weak hyperbolicity of cube complexes and quasi-arboreal groups*, J. Topol. 7 (2014) 385–418 MR

[31] **M F Hagen**, **D T Wise**, *Cubulating hyperbolic free-by-cyclic groups: the general case*, Geom. Funct. Anal. 25 (2015) 134–179 MR

[32] **M F Hagen**, **D T Wise**, *Cubulating hyperbolic free-by-cyclic groups: the irreducible case*, Duke Math. J. 165 (2016) 1753–1813 MR

[33] **F Haglund**, *Isometries of* $\mathrm{CAT}(0)$ *cube complexes are semi-simple*, preprint (2007) arXiv

[34] **F Haglund**, *Finite index subgroups of graph products*, Geom. Dedicata 135 (2008) 167–209 MR

[35] **F Haglund**, **D T Wise**, *Special cube complexes*, Geom. Funct. Anal. 17 (2008) 1551–1620 MR

[36] **F Haglund**, **D T Wise**, *Coxeter groups are virtually special*, Adv. Math. 224 (2010) 1890–1903 MR

[37] **U Hamenstädt**, *Geometry of the mapping class groups, III: Quasi-isometric rigidity*, preprint (2005) arXiv

[38] **J Huang**, *Quasi-isometry classification of right-angled Artin groups, II: Several infinite out cases*, preprint (2016) arXiv

[39] **J Huang**, *Top-dimensional quasiflats in* CAT(0) *cube complexes*, Geom. Topol. 21 (2017) 2281–2352 MR

[40] **J Huang**, **B Kleiner**, *Groups quasi-isometric to RAAG's*, preprint (2016) arXiv To appear in Duke Math. J.

[41] **M Kapovich**, **B Kleiner**, **B Leeb**, *Quasi-isometries and the de Rham decomposition*, Topology 37 (1998) 1193–1211 MR

[42] **S-h Kim**, **T Koberda**, *Embedability between right-angled Artin groups*, Geom. Topol. 17 (2013) 493–530 MR

[43] **S-H Kim**, **T Koberda**, *The geometry of the curve graph of a right-angled Artin group*, Internat. J. Algebra Comput. 24 (2014) 121–169 MR

[44] **S-h Kim**, **T Koberda**, *Anti-trees and right-angled Artin subgroups of braid groups*, Geom. Topol. 19 (2015) 3289–3306 MR

[45] **B Kleiner**, **B Leeb**, *Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings*, C. R. Acad. Sci. Paris Sér. I Math. 324 (1997) 639–643 MR

[46] **B Kleiner**, **B Leeb**, *Groups quasi-isometric to symmetric spaces*, Comm. Anal. Geom. 9 (2001) 239–260 MR

[47] **T Koberda**, *Ping-pong lemmas with applications to geometry and topology*, from "Geometry, topology and dynamics of character varieties" (W Goldman, C Series, S P Tan, editors), Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap. 23, World Scientific, Hackensack, NJ (2012) 139–158 MR

[48] **T Koberda**, *Right-angled Artin groups and a generalized isomorphism problem for finitely generated subgroups of mapping class groups*, Geom. Funct. Anal. 22 (2012) 1541–1590 MR

[49] **M R Laurence**, *A generating set for the automorphism group of a graph group*, J. London Math. Soc. 52 (1995) 318–334 MR

[50] **L Mosher**, **M Sageev**, **K Whyte**, *Quasi-actions on trees, II: Finite depth Bass–Serre trees*, Mem. Amer. Math. Soc. 1008, Amer. Math. Soc., Providence, RI (2011) MR

[51] **G D Mostow**, *Strong rigidity of locally symmetric spaces*, Annals of Mathematics Studies 78, Princeton Univ. Press (1973) MR

[52] **Y Ollivier**, **D T Wise**, *Cubulating random groups at density less than* $1/6$, Trans. Amer. Math. Soc. 363 (2011) 4701–4733 MR

[53] **P Papasoglu**, **K Whyte**, *Quasi-isometries between groups with infinitely many ends*, Comment. Math. Helv. 77 (2002) 133–144 MR

[54] **P Przytycki**, **D T Wise**, *Mixed 3-manifolds are virtually special*, preprint (2012) arXiv

[55] **P Przytycki**, **D T Wise**, *Graph manifolds with boundary are virtually special*, J. Topol. 7 (2014) 419–435 MR

[56] **M Sageev**, *Ends of group pairs and non-positively curved cube complexes*, Proc. London Math. Soc. 71 (1995) 585–617 MR

[57] **H Servatius**, *Automorphisms of graph groups*, J. Algebra 126 (1989) 34–60 MR

[58] **H Servatius**, **C Droms**, **B Servatius**, *Surface subgroups of graph groups*, Proc. Amer. Math. Soc. 106 (1989) 573–578 MR

[59] **S J Taylor**, *Right-angled Artin groups and* $\mathrm{Out}(\mathbb{F}_n)$*, I: Quasi-isometric embeddings*, Groups Geom. Dyn. 9 (2015) 275–316 MR

[60] **D T Wise**, *Cubulating small cancellation groups*, Geom. Funct. Anal. 14 (2004) 150–214 MR

[61] **D T Wise**, *Research announcement: the structure of groups with a quasiconvex hierarchy*, Electron. Res. Announc. Math. Sci. 16 (2009) 44–55 MR

[62] **D T Wise**, *The structure of groups with a quasiconvex hierarchy*, preprint (2011) Available at `http://www.math.mcgill.ca/wise/papers.html`

The Department of Mathematics and Statistics, McGill University
Montreal, QC, Canada

`jingyin.huang@mcgill.ca`

Proposed: Walter Neumann | Received: 4 November 2015
Seconded: Dmitri Burago, Bruce Kleiner | Revised: 16 September 2016

Geometry & Topology 21 (2017) 3539–3599

Maximal representations, non-Archimedean Siegel spaces, and buildings

MARC BURGER
MARIA BEATRICE POZZETTI

Let $\mathbb{F}$ be a real closed field. We define the notion of a maximal framing for a representation of the fundamental group of a surface with values in $\mathrm{Sp}(2n,\mathbb{F})$. We show that ultralimits of maximal representations in $\mathrm{Sp}(2n,\mathbb{R})$ admit such a framing, and that all maximal framed representations satisfy a suitable generalization of the classical collar lemma. In particular, this establishes a collar lemma for all maximal representations into $\mathrm{Sp}(2n,\mathbb{R})$. We then describe a procedure to get from representations in $\mathrm{Sp}(2n,\mathbb{F})$ interesting actions on affine buildings, and in the case of representations admitting a maximal framing, we describe the structure of the elements of the group acting with zero translation length.

20-XX, 22E40

1 Introduction

Let Σ be a connected, orientable surface of genus g with $p \geq 0$ punctures and negative Euler characteristic, and let V be a symplectic vector space over $\mathbb{R}$. A current theme in higher Teichmüller theory is to which extent classical hyperbolic geometry and some fundamental structures on the Teichmüller space of Σ carry over to the geometry and the moduli space of maximal representations of $\Gamma = \pi_1(\Sigma)$ into $\mathrm{Sp}(V)$ or Hitchin representations into $\mathrm{SL}(V)$. For instance, compactifications of spaces of representations of Γ have been introduced and studied by Alessandrini [1], Le [14] and Parreau [22]. In the context of Hitchin representations, asymptotic properties of diverging sequences were studied by Collier and Li [7], Katzarkov, Noll, Pandit and Simpson [10], Loftin [18], Mazzeo, Swoboda, Weiss and Witt [19], Parreau [23] and Zhang [29; 30].

The purpose of this paper is to study the action on an asymptotic cone of the symmetric space $\mathcal{X}$ associated to $\mathrm{Sp}(V)$ defined by a sequence $(\rho_k)_{k\in\mathbb{N}}$ of maximal representations $\rho_k\colon \Gamma \to \mathrm{Sp}(V)$. More precisely, we fix a nonprincipal ultrafilter ω on $\mathbb{N}$ and

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3539

let $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ be a sequence of basepoints. We say that a sequence of scales $\lambda = (\lambda_k)_{k\in\mathbb{N}}$ is adapted to $(\rho_k, x_k)_{k\in\mathbb{N}}$ if

$$\lim_{\omega} \frac{D_S(\rho_k)(x_k)}{\lambda_k} < \infty.$$

Here for a representation ρ and a finite generating set S for Γ, we define $D_S(\rho)(x) = \max_{\gamma\in S} d(\rho(\gamma)x, x)$, where d denotes the Riemannian distance on $\mathcal{X}$. Observe that the above property is independent of the choice of the finite generating set S.

In this situation, we obtain an action ${}^{\omega}\rho_{\lambda}\colon \Gamma \to \mathrm{Iso}({}^{\omega}\mathcal{X}_{\lambda})$ by isometries on the asymptotic cone ${}^{\omega}\mathcal{X}_{\lambda}$ of the sequence $(\mathcal{X}, x_k, d/\lambda_k)$. The space ${}^{\omega}\mathcal{X}_{\lambda}$ is not only CAT(0)-complete, but when the limit $\lim_{\omega}\lambda_k$ is infinite, it is an affine building associated to the algebraic group $\mathrm{Sp}(V)$ over a specific field (more on this below); see Kleiner and Leeb [11], Kramer and Tent [13], Parreau [20] and Thornton [28]. Depending on the choice of scales, the representation ${}^{\omega}\rho_{\lambda}$ might have a global fixed point, but as it turns out, if the representations ρ_k are maximal, the limiting action is always faithful. Our main result gives then the underlying geometric structure of the set of elements γ in Γ whose translation length $L({}^{\omega}\rho_{\lambda}(\gamma))$ in ${}^{\omega}\mathcal{X}_{\lambda}$ is zero; notice that for an isometry of an affine building, having zero translation length is equivalent to having a fixed point.

For convenience, we fix once and for all a complete hyperbolic metric on Σ of finite area, and identify Γ with a subgroup of $\mathrm{PSL}(2,\mathbb{R})$. In order to state the main result, we recall that a decomposition $\Sigma = \bigcup_{v\in\mathcal{V}} \Sigma_v$ into subsurfaces with geodesic boundary gives rise to a presentation of Γ as fundamental group of a graph of groups with vertex set $\mathcal{V}$ and vertex groups $\pi_1(\Sigma_v)$. The group Γ acts on the associated Bass–Serre tree $\mathcal{T}$ and, in particular, on its vertex set $\widetilde{\mathcal{V}}$; observe that for $v \in \mathcal{V}$ and $w \in \widetilde{\mathcal{V}}$ lying above v, the stabilizer Γ_w of w in Γ is isomorphic to $\pi_1(\Sigma_v)$.

Theorem 1.1 *Let $\rho_k\colon \Gamma \to \mathrm{Sp}(V)$ be a sequence of maximal representations, $(\lambda_k)_{k\geq 1}$ an adapted sequence of scales and ${}^{\omega}\rho_{\lambda}$ the action of Γ on the asymptotic cone ${}^{\omega}\mathcal{X}_{\lambda}$. Then ${}^{\omega}\rho_{\lambda}$ is faithful. Moreover, there is a decomposition $\Sigma = \bigcup_{v\in\mathcal{V}} \Sigma_v$ of Σ into subsurfaces with geodesic boundary such that:*

(1) *for every $\gamma \in \Gamma$ whose corresponding closed geodesic is not contained in any subsurface, $L({}^{\omega}\rho_{\lambda}(\gamma)) > 0$;*

(2) *for every $v \in \mathcal{V}$, there is the following dichotomy:*

  (PT) *for every $w \in \widetilde{\mathcal{V}}$ lying above v, and any $\gamma \in \Gamma_w$ which is not boundary parallel, ${}^{\omega}\rho_{\lambda}(\gamma)$ has positive translation length;*

  (FP) *for every $w \in \widetilde{\mathcal{V}}$ lying above v, the stabilizer Γ_w has a common fixed point $b_w \in {}^{\omega}\mathcal{X}_{\lambda}$.*

A natural question is, given a sequence of maximal representations, how the choice of basepoints and scales influences the action of Γ on the asymptotic cone and, in particular, the decomposition given in Theorem 1.1. Turning to this issue, recall that for a maximal representation $\rho\colon \Gamma \to \mathrm{Sp}(V)$, the displacement function $x \mapsto D_S(\rho)(x)$ with respect to a generating set $S \subset \Gamma$ achieves its minimum $\mu_S(\rho)$ in a compact region of the symmetric space $\mathcal{X}$. Given a sequence $(\rho_k)_{k\in\mathbb{N}}$ of maximal representations, we have $\lim_\omega \mu_S(\rho_k) < \infty$ if and only if, up to modifying the sequence on a set of ω–measure zero, $(\rho_k)_{k\in\mathbb{N}}$ is contained in a compact subset of the character variety of maximal representations.

Assume thus that $\lim_\omega \mu_S(\rho_k) = \infty$. Choosing a sequence $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ of basepoints such that $D_S(\rho_k)(x_k) = \mu_S(\rho_k)$, the sequence of scales $(\mu_k := \mu_S(\rho_k))_{k\in\mathbb{N}}$ is obviously adapted to the sequence $(\rho_k, x_k)_{k\in\mathbb{N}}$, and the resulting Γ–action ${}^\omega\rho_\mu$ on ${}^\omega\mathcal{X}_\mu$ has no global fixed point. We show then (see Proposition 10.6) that if $(y_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ is a sequence of basepoints and $(\lambda_k)_{k\in\mathbb{N}}$ is an adapted sequence of scales such that ${}^\omega\rho_\lambda$ has no global fixed point, then ${}^\omega\mathcal{X}_\lambda$ equals ${}^\omega\mathcal{X}_\mu$ with homothetic distance function, and the actions ${}^\omega\rho_\lambda$ and ${}^\omega\rho_\mu$ coincide. In particular, the decomposition of Σ into subsurfaces given by Theorem 1.1 is uniquely determined by the sequence $(\rho_k)_{k\in\mathbb{N}}$.

We say that a subsurface is of type (PT) (resp. (FP)) if the first (resp. the second) possibility in Theorem 1.1(2) holds. One can show that any decomposition of the surface Σ and any assignment of type (PT) or (FP) to the subsurfaces can be realized by the limiting action for an appropriate sequence $(\rho_k)_{k\in\mathbb{N}}$. On the other hand, Theorem 1.1 suggests that in a generic limiting action without a global fixed point, no element of Γ should have zero translation length. We plan on analyzing the properties of such representations in future work.

In case there is a subsurface of type (FP), the restriction $\rho_k|_{\Gamma_w}\colon \Gamma_w \to \mathrm{Sp}(V)$ is a sequence of maximal representations to which the preceding discussion applies; that is, either up to ω–measure zero the sequence is relatively compact in the character variety of Γ_w, or there is an essentially unique choice of basepoints and scales such that the limiting action does not have a global fixed point. Since, at each step, the topological complexity of the surface decreases, this procedure stops after finitely many iterations and can be seen as an asymptotic expansion of the initial sequence $(\rho_k)_{k\in\mathbb{N}}$.

When each subsurface in the decomposition of Theorem 1.1 is of type (FP), we can use the fixed points b_w to construct a map from the Bass–Serre tree $\mathcal{T}$ to the asymptotic cone:

Theorem 1.2 *Assume that for any subsurface of the decomposition, possibility (FP) holds. Then there is a ${}^\omega\rho_\lambda$–equivariant quasi-isometric embedding $\mathcal{T} \to {}^\omega\mathcal{X}_\lambda$.*

In the case of a vector space of dimension 2, maximal representations correspond to holonomies of hyperbolizations; in this case, the second possibility in Theorem 1.1(2) occurs, for example, for sequences of hyperbolizations obtained by pinching a multicurve. In this case, the image of the quasi-isometric embedding of Theorem 1.2 is a simplicial subtree of the asymptotic cone ${}^{\omega}\mathcal{X}_{\lambda}$. In higher rank, it is possible to construct examples in which the image of the Bass–Serre tree is not totally geodesic in the affine building ${}^{\omega}\mathcal{X}_{\lambda}$.

We finish our discussion about ultralimits of maximal representations mentioning two interesting geometric properties of maximal representations that can be deduced from our work. Let S be a connected generating set, namely a generating set for Γ such that the union of the closed geodesics representing the elements of S is a connected subset of Σ, and let $L_S(\rho)$ denote the maximal displacement of an element in the generating set S:

$$L_S(\rho) = \max_{\gamma \in S} L(\rho(\gamma)).$$

Corollary 1.3 *Let $S \subset \Gamma$ be a connected generating set for Γ. Then there is a constant C depending only on S and $2n = \dim V$ such that for any maximal representation $\rho\colon \Gamma \to \mathrm{Sp}(V)$, we have*

$$(\ln 2)\sqrt{n} \le L_S(\rho) \le \mu_S(\rho) \le C L_S(\rho).$$

We say that two diverging sequences of real numbers $(\lambda_k)_{k\in\mathbb{N}}$ and $(\mu_k)_{k\in\mathbb{N}}$ have the same growth rate according to the ultrafilter ω if $\lim_{\omega} \lambda_k/\mu_k$ is finite and nonzero.

Corollary 1.4 *Let $(\rho_k)_{k\in\mathbb{N}}$ be a sequence of maximal representations of the fundamental group Γ of a surface of genus g with p punctures. Then, varying $\gamma \in \Gamma$, there are at most $8g-8+4p$ distinct growth rate classes among the sequences $L(\rho_k(\gamma))_{k\in\mathbb{N}}$.*

1.1 Real closed fields

The building structure on ${}^{\omega}\mathcal{X}_{\lambda}$ alluded to previously comes about as follows. Assume that the sequence of scales $(\lambda_k)_{k\in\mathbb{N}}$ is unbounded. Then $\sigma = (e^{-\lambda_k})_{k\in\mathbb{N}}$ is an infinitesimal in the field $\mathbb{R}_{\omega}$ of the hyperreals, and the building ${}^{\omega}\mathcal{X}_{\lambda}$ is associated to $\mathrm{Sp}(V \otimes \mathbb{R}_{\omega,\sigma})$ [20; 28]. Here $\mathbb{R}_{\omega,\sigma}$ is the valuation field introduced by Robinson [24]. The characterizing properties of the representations arising as ultralimits of maximal representations make sense in the more general context of symplectic groups over arbitrary real closed fields.[1] When $V_{\mathbb{F}}$ is a symplectic vector space over a real closed field $\mathbb{F}$, the Kashiwara cocycle classifies the orbits of $\mathrm{Sp}(V_{\mathbb{F}})$ on triples $\mathcal{L}(V_{\mathbb{F}})^{(3)}$ of pairwise transverse Lagrangians and can be used to select maximal triples (see Section 2.3 for a precise definition of maximal triples). The general objects of our study are representations which admit a maximal framing:

[1] We refer to Kaplansky [9] for general facts about linear algebra over real closed fields.

Definition 1.5 A representation $\rho\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{F}})$ *admits a maximal framing* if there exist a Γ–invariant subset $S \subset \partial\mathbb{H}^2$ including the fixed points of hyperbolic elements of Γ, and an equivariant map $\phi\colon S \to \mathcal{L}(V_{\mathbb{F}})$, such that for every positively oriented triple (x, y, z) in S^3, the image $(\phi(x), \phi(y), \phi(z))$ is maximal. If we want to emphasize the domain of definition, we will refer to a *maximal S–framing*.

Remark 1.6 For the conclusion of Theorem 1.8 (see below) to hold, the existence of a maximal S–framing for S the set of fixed points of hyperbolic elements is sufficient. However, the fact that the reduction (see Theorem 1.7) of a maximal S–framed representation admits a maximal S–framing will be used in subsequent papers where we study the structure of the real spectrum compactification of maximal representation varieties.

If $\mathbb{F} = \mathbb{R}$, any maximal representation admits a maximal framing (see Burger, Iozzi and Wienhard [6, Theorem 8]), and we show in Corollary 10.4 that this is also true for all ultralimits of maximal representations. Even more, the class of representations admitting a maximal framing is closed under the natural reduction process we are now going to describe. Let $\mathcal{O} \subset \mathbb{F}$ be an order convex local subring.[2] Its quotient by the maximal ideal, denoted by $\mathbb{F}_{\mathcal{O}}$, is real closed as well. Assume now that there exists a symplectic basis of $V_{\mathbb{F}}$ such that $\rho(\Gamma) \subset \mathrm{Sp}(2n, \mathcal{O})$. We can then consider the composition $\rho_{\mathcal{O}}$ of ρ with the quotient homomorphism $\mathrm{Sp}(2n, \mathcal{O}) \to \mathrm{Sp}(2n, \mathbb{F}_{\mathcal{O}})$:

Theorem 1.7 *Assume that $\rho\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{F}})$ admits a maximal S–framing. Then the reduction $\rho_{\mathcal{O}}\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{F}_{\mathcal{O}}})$ admits a maximal S–framing as well.*

Theorem 1.7 allows us in general to obtain well controlled actions on affine buildings. Indeed, for each infinitesimal $\sigma > 0$, the set of elements of $\mathbb{F}$ comparable with σ,

$$\mathcal{O}_\sigma = \{x \in \mathbb{F} : |x| \le \sigma^{-k} \text{ for some } k \in \mathbb{Z}\},$$

forms an order convex subring of $\mathbb{F}$. We denote by $\mathbb{F}_\sigma$ its residue field, which inherits from $\mathcal{O}_\sigma$ an order compatible valuation. As a consequence, to any reductive algebraic group over $\mathbb{F}_\sigma$ is associated an affine Bruhat–Tits building [2]. Since Γ is finitely generated, for each representation $\rho\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{F}})$ and every choice of a basis, it is possible to choose an infinitesimal σ such that $\rho(\Gamma) \subset \mathrm{Sp}(2n, \mathcal{O}_\sigma)$. By passing to the quotient $\rho_\sigma\colon \Gamma \to \mathrm{Sp}(2n, \mathbb{F}_\sigma)$, we get an action on the affine building associated to $\mathrm{Sp}(2n, \mathbb{F}_\sigma)$. The main result for maximal framed representations over real closed fields with valuation is:

[2] The definition of an order convex subring is recalled in Section 5.

Theorem 1.8 *Let $\rho\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{L}})$ be a maximal framed representation, where $\mathbb{L}$ is real closed with order compatible valuation, and let $\mathcal{B}$ be the Bruhat–Tits affine building associated to $\mathrm{Sp}(V_{\mathbb{L}})$. Then the action of Γ on $\mathcal{B}$ satisfies the conclusions of Theorem 1.1.*

When $\mathbb{L}$ is a real closed field with order compatible valuation, we denote by $\mathcal{U}$ the order convex valuation ring with residue field $\mathbb{L}_{\mathcal{U}}$. We already mentioned that the action on the affine building associated to a representation $\rho\colon \Gamma \to \mathrm{Sp}(V_{\mathbb{L}})$ might have a global fixed point. However, when this is the case, it is possible to find a symplectic basis of $V_{\mathbb{L}}$ such that $\rho(\Gamma) \subset \mathrm{Sp}(2n,\mathcal{U})$, and if ρ admits a maximal framing, then it follows from Theorem 1.7 that the reduction $\rho_{\mathcal{U}}\colon \Gamma \to \mathrm{Sp}(2n,\mathbb{L}_{\mathcal{U}})$ has the same property. In particular, this can be used to study the restriction of the representation ρ to the subsurfaces defined in Theorem 1.8.

As a consequence of Theorem 1.8, we get a concrete way of checking if a representation ρ admitting a maximal framing has a global fixed point: if S is a connected generating set for Γ, then ρ has a global fixed point if and only if each element of S has a fixed point (see Corollary 7.6 for a precise formulation of this result and some further comments).

1.2 Tools

We now turn to a short description of the key tools we develop in this paper. In the context of his approach to the compactification of the Teichmüller space [3], Brumfiel studied non-Archimedean hyperbolic planes [4]: for any ordered field $\mathbb{F}$, he associates to $\mathrm{PSL}(2,\mathbb{F})$ a nonstandard hyperbolic plane $\mathbb{HF}^2$, and for fields with valuation, he introduces a pseudodistance on $\mathbb{HF}^2$ whose Hausdorff quotient is the $\mathbb{R}$–tree associated to $\mathrm{PSL}(2,\mathbb{F})$. Inspired by Brumfiel's work (see also [13]), we associate to a symplectic group $\mathrm{Sp}(2n,\mathbb{F})$ over a real closed field $\mathbb{F}$ the space

$$\mathcal{X}_{\mathbb{F}} = \{X + iY \mid X, Y \in \mathrm{Sym}(n,\mathbb{F}),\ Y \text{ positive definite}\},$$

where $\mathrm{Sym}(n,\mathbb{F})$ denotes the vector space of symmetric $n\times n$ matrices with coefficients in $\mathbb{F}$. The group $\mathrm{Sp}(2n,\mathbb{F})$ acts on $\mathcal{X}_{\mathbb{F}}$ by fractional linear transformations, and the $\mathrm{Sp}(2n,\mathbb{F})$–space $\mathcal{X}_{\mathbb{F}}$ can be thought of as a nonstandard version of the Siegel upper half-space. Using a matrix-valued cross-ratio, we define, for any two transverse Lagrangians $a, b \in \mathcal{L}(\mathbb{F}^{2n})$, the $\mathbb{F}$*–tube* $\mathcal{Y}_{a,b}$ which is the nonstandard symmetric space associated to the stabilizer in $\mathrm{Sp}(2n,\mathbb{F})$ of the pair (a,b), a group isomorphic to $\mathrm{GL}(n,\mathbb{F})$. In the case of the hyperbolic plane, the $\mathbb{F}$–tubes are just the Euclidean half-circles joining the ideal points a, b. Given a representation $\rho\colon \Gamma \to \mathrm{Sp}(2n,\mathbb{F})$ admitting a maximal framing $\phi\colon S \to \mathcal{L}(\mathbb{F}^{2n})$, we can associate to every hyperbolic element $\gamma \in \Gamma$

the $\mathbb{F}$–tube $\mathcal{Y}_\gamma = \mathcal{Y}_{\phi(\gamma^-),\phi(\gamma^+)}$, where γ^-, γ^+ are the fixed points of γ in $\partial\mathbb{H}^2$. One key property that we exploit is that the intersection pattern of the axes of hyperbolic elements in Γ is reflected in the intersection pattern of the corresponding $\mathbb{F}$–tubes. When the field $\mathbb{F}$ has an order compatible valuation, there is a natural $\mathbb{R}_{\geq 0}$–valued pseudodistance on $\mathcal{X}_\mathbb{F}$, and the relation between cross-ratios and this pseudodistance allows us to quantify the intersection pattern of the $\mathbb{F}$–tubes. Finally, we exploit that the Hausdorff quotient of $\mathcal{X}_\mathbb{F}$ can be identified with the set of vertices of the affine Bruhat–Tits building associated to $\mathrm{Sp}(2n,\mathbb{F})$.

1.3 Collar lemma

We finish this introduction discussing another geometric property of representations admitting a maximal framing, which is at the basis of most of the results we discussed so far. Recall that, since any element $g \in \mathrm{Sp}(V)$ is conjugate to ${}^t g^{-1}$, the set of eigenvalues of a symplectic element is closed with respect to inverse: if λ is an eigenvalue of g, the same is true for λ^{-1}. With a slight abuse of terminology, we say that two hyperbolic elements $\gamma, \eta \in \Gamma < \mathrm{PSL}(2,\mathbb{R})$ intersect if their axes do.

Theorem 1.9 (collar lemma) *Let $\mathbb{F}$ be a real closed field, and let $\rho\colon \Gamma \to \mathrm{Sp}(V_\mathbb{F})$ be a representation admitting a maximal framing. Then if $\gamma \in \Gamma$ is hyperbolic, $\rho(\gamma)$ has no eigenvalue of absolute value 1. Let $|\lambda_1(\gamma)| \geq \cdots \geq |\lambda_n(\gamma)| > 1$ be the eigenvalues of absolute value larger than 1. If the hyperbolic elements γ, η in Γ intersect, then*

$$|\lambda_1(\gamma)|^{2n} \geq \frac{1}{|\lambda_n(\eta)|^2 - 1}, \tag{1}$$

$$\left(\prod_{i=1}^{n} |\lambda_i(\gamma)|^{2/n} - 1\right)\left(\prod_{i=1}^{n} |\lambda_i(\eta)|^{2/n} - 1\right) \geq 1. \tag{2}$$

Here $|\cdot|$ denotes the $\mathbb{F}$–valued absolute value on $\mathbb{F}[i]$, and we count the eigenvalues with their multiplicity as roots of the characteristic polynomial. We immediately get from Theorem 1.9(2):

Corollary 1.10 *Under the same assumptions of Theorem 1.9, we have:*

(1) *If γ is self-intersecting, then $|\lambda_1(\gamma)| \geq \sqrt{2}$.*

(2) *If γ satisfies $|\lambda_1(\gamma)| < \sqrt{2}$, then γ is simple and any η intersecting γ satisfies $|\lambda_1(\eta)| > \sqrt{2}$. In particular, there are at most $(3g-3+p)$ conjugacy classes of hyperbolic elements γ with $|\lambda_1(\gamma)| < \sqrt{2}$.*

(3) *There exists $\epsilon > 0$ in $\mathbb{F}$ with $|\lambda_1(\gamma)| > 1+\epsilon$ for any hyperbolic $\gamma \in \Gamma$.*

As an application of the collar lemma, we establish a uniform discontinuity property of the $\Gamma = \pi_1(\Sigma)$ action on the non-Archimedean Siegel space $\mathcal{X}_{\mathbb{F}}$ by a maximal S–framed representation in the case where Σ has no boundary. Recall here that $\mathbb{F}$ has a natural topology given by the order, and so does $\mathcal{X}_{\mathbb{F}}$ as an open subset of $M(n, \mathbb{F}[i])$. Given an open subset $\mathcal{U} \subset \mathcal{X}_{\mathbb{F}} \times \mathcal{X}_{\mathbb{F}}$ containing the diagonal and $x \in \mathcal{X}_{\mathbb{F}}$, we let $\mathcal{U}_x$ denote the open neighborhood consisting of all $y \in \mathcal{X}_{\mathbb{F}}$ with $(x, y) \in \mathcal{U}$.

Corollary 1.11 *Let $\Gamma = \pi_1(\Sigma)$ where Σ has no boundary, and let $\rho\colon \Gamma \to \mathrm{Sp}(2n, \mathbb{F})$ be a representation admitting a maximal framing. Then there is an open neighborhood of the diagonal $\mathcal{U} \subset \mathcal{X}_{\mathbb{F}} \times \mathcal{X}_{\mathbb{F}}$ which is invariant for the diagonal $\mathrm{Sp}(2n, \mathbb{F})$–action and such that for every $x \in \mathcal{X}_{\mathbb{F}}$,*

$$\rho(\gamma)\mathcal{U}_x \cap \mathcal{U}_x = \varnothing \quad \textit{for all } \gamma \in \Gamma \setminus \{e\}.$$

We finish the introduction drawing some consequences of the collar lemma in the case of classical maximal representations. It was established by Siegel in [26, Theorem 3] that, under suitable normalizations, the translation length of an isometry $g \in \mathrm{Sp}(2n, \mathbb{R})$ on the symmetric space $\mathcal{X}_{\mathbb{R}}$ is

$$L(g) = 2\sqrt{\sum_{i=1}^{n} \ln^2 |\lambda_i(g)|}\,.$$

Using this formula, we get, from Theorem 1.9(2) and the Cauchy–Schwarz inequality, the following:

Corollary 1.12 *Let $\rho\colon \Gamma \to \mathrm{Sp}(2n, \mathbb{R})$ be a maximal representation. If γ and η intersect, then*

$$(e^{L(\rho(\gamma))/\sqrt{n}} - 1)(e^{L(\rho(\eta))/\sqrt{n}} - 1) \geq 1.$$

In particular, if γ is not simple then $L(\rho(\gamma)) \geq \log(2)\sqrt{n}$.

Using that $e^x - 1 \leq 2x$ for $0 \leq x \leq 1$, we get that, if $L(\rho(\eta)) \leq \sqrt{n}$, then

$$\frac{L(\rho(\gamma))}{\sqrt{n}} \geq \ln\left(\frac{\sqrt{n}}{2L(\rho(\eta))}\right),$$

which exhibits the same asymptotic growth relation as in the Teichmüller setting. However, it is worth remarking that, as opposed to the classical collar lemma, Corollary 1.12 is not just a consequence of the Margulis Lemma: in our setting, the sets of minimal displacement of the isometries $\rho(\gamma)$ and $\rho(\eta)$ do not necessarily intersect. A similar version of the collar lemma in the framework of Hitchin representations has been recently established by Lee and Zhang [15]; see Remark 3.5 for a comparison with our results.

Outline of the paper In Section 2, we define three different models for the nonstandard symmetric space, and we study the action of $\mathrm{Sp}(V)$ on n–tuples of transverse Lagrangians. Section 3 is devoted to the proof of the collar lemma, Theorem 3.3, for representations admitting a maximal framing. The matrix-valued cross-ratio and the $\mathbb{F}$–tubes are introduced and studied in Section 4. In Section 5, we focus on order convex subrings and describe how to obtain representations over the residue field. The main result of the section is Theorem 5.9 (Theorem 1.7 in the introduction), whose proof also exploits the geometric input coming from the collar lemma. In Section 6, we restrict to fields with valuations and use the cross-ratio to describe the projection from the nonstandard symmetric space to the affine Bruhat–Tits building. In Section 7, we initiate our study of elements with zero translation length: to each such element, we associate a pair of canonical fixed points (Proposition 7.1) and give sufficient conditions for these points to coincide (Proposition 7.3). The proof of the decomposition Theorem 1.8 (Theorem 8.1) occupies Section 8, while Theorem 1.2 is proven in Section 9. In the last section of the paper, we discuss the relation between ultralimits of maximal representations and representations in symplectic groups over the Robinson field $\mathbb{R}_{\omega,\sigma}$. This allows us to deduce Theorem 1.1 from the more general Theorem 1.8 and, in the case of closed surfaces, to completely characterize representations in $\mathrm{Sp}(2n,\mathbb{R}_{\omega,\sigma})$ which admit a maximal framing (Theorem 10.5).

Acknowledgements The major part of this work was done during the MSRI program "Dynamics on Moduli Spaces of Geometric structures" 2015; we thank the organizers and the participants of this program who have created an intellectually stimulating atmosphere. We thank in particular Anne Parreau for important discussions and, more specifically, for encouraging us to prove Theorem 1.1, Domingo Toledo for his helpful comments, Thomas Huber for providing a proof of Proposition A.1 valid in the general context of real closed fields and the referee for carefully reading our paper and for several insightful comments. Burger thanks the Clay foundation and the Simons foundation for support. This material is based upon work supported by the National Science Foundation under Grant No. 0932078 000 while the authors were in residence at MSRI.

2 Symplectic geometry over real closed fields

2.1 Basic objects

Let V be a $2n$–dimensional vector space over a field $\mathbb{F}$, endowed with a symplectic form $\langle\cdot,\cdot\rangle$. The symplectic group $\mathrm{Sp}(V)$ is the subgroup of elements of $\mathrm{GL}(V)$ preserving the form $\langle\cdot,\cdot\rangle$. Recall that a Lagrangian subspace is a maximal isotropic subspace of V; they form a subset of the Grassmannian $\mathrm{Gr}_n(V)$ of n–dimensional subspaces of V, denoted by $\mathcal{L}(V)$. Whenever a Lagrangian l is fixed, we denote by $\mathcal{L}(V)^l$ the set of Lagrangians transverse to l, and by $\mathcal{Q}(l)$ the vector space of quadratic forms on l.

Given a, b in $\mathcal{L}(V)$ transverse, we recall the construction of an affine chart

$$j_{a,b}\colon \mathcal{Q}(a) \to \mathcal{L}(V)^b.$$

For each element f in $\mathcal{Q}(a)$, we denote by $b_f\colon a \times a \to \mathbb{F}$ the associated symmetric bilinear form. Since a and b are transverse, the symplectic pairing induces an isomorphism of b with the dual of a. We denote by $T_f\colon a \to b$ the unique linear map satisfying

$$\langle v, T_f(w)\rangle = b_f(v, w), \quad \text{for } v, w \in a.$$

The subspace of V defined by

$$j_{a,b}(f) := \{v + T_f(v) \mid v \in a\}$$

is a Lagrangian subspace transverse to b.

Conversely, if l is transverse to b, any vector v in a can be written uniquely as a combination of a vector in b and a vector in l. This allows us to define a linear map $T^l_{a,b}\colon a \to b$ by requiring that $v + T^l_{a,b}(v) \in l$. In turn, we can use $T^l_{a,b}$ to define the quadratic form $Q_{a,l,b}$ on a:

$$Q_{a,l,b}(v) = \langle v, T^l_{a,b}(v)\rangle, \quad v \in a,$$

which satisfies $j_{a,b}(Q_{a,l,b}) = l$.

In the theory of maximal representations, positive-definite quadratic forms play a prominent role. If q_1, q_2 are quadratic forms we will write $q_1 \gg 0$ to indicate that q_1 is positive definite, and $q_1 \gg q_2$ to indicate that the difference $q_1 - q_2$ is positive definite.

2.2 Three models of the Siegel space

The symmetric space associated to the symplectic group $\mathrm{Sp}(2n, \mathbb{R})$ was extensively studied by Siegel [26] and is often referred to as the Siegel space. We now show that the three most studied models for the Siegel space can be defined over arbitrary ordered fields, are always equivariantly isomorphic, and give rise to interesting geometries.

We fix an ordered field $\mathbb{F}$. Clearly the polynomial $f(x) = x^2 + 1$ is irreducible in $\mathbb{F}[x]$. We denote by $i \in \overline{\mathbb{F}}$ a root of the polynomial f and by $\mathbb{K}$ the splitting field of f, the degree two extension $\mathbb{K} = \mathbb{F}[i]$. If V is a $2n$–dimensional vector space over $\mathbb{F}$ endowed with a symplectic form $\langle \cdot, \cdot \rangle$, we denote by $V_{\mathbb{K}}$ the "complexification" $V_{\mathbb{K}} = V \otimes \mathbb{K}$ and by $\langle \cdot, \cdot \rangle_{\mathbb{K}}\colon V_{\mathbb{K}}^2 \to \mathbb{K}$ the $\mathbb{K}$–linear extension to $V_{\mathbb{K}}$ of $\langle \cdot, \cdot \rangle$.

The first model of the Siegel space consists of *compatible complex structures* on V:

$$\mathbb{X}_V = \{J \in \mathrm{GL}(V) \mid J^2 = -\mathrm{Id},\ \langle J\cdot,\cdot\rangle \text{ is a scalar product}\}.$$

The set $\mathbb{X}_V$ is a semialgebraic subset of $\mathrm{End}(V)$ on which the symplectic group $\mathrm{Sp}(V)$ acts by conjugation. For $J \in \mathbb{X}_V$, we will denote by $(\cdot,\cdot)_J := \langle J\cdot,\cdot\rangle$ the corresponding scalar product.

The second model of the Siegel space corresponds to the image of the *Borel embedding*; see [5, Section 2.1.1; 25]. As in the real case, we realize $\mathbb{X}_V$ as a semialgebraic subset $\mathcal{T}_V$ of $\mathcal{L}(V_{\mathbb{K}})$. Indeed, if $J \in \mathrm{GL}(V)$ is an element of $\mathbb{X}_V$, the complexification $J \otimes \mathbb{I}_{\mathbb{K}}$ is diagonalizable over $\mathbb{K}$. It is easy to verify that the eigenspaces $L_J^{\pm}$ of $J \otimes \mathbb{I}_{\mathbb{K}}$ with respect to the eigenvalues $\pm i$ are elements of $\mathcal{L}(V_{\mathbb{K}})$. If we denote by $\sigma\colon V_{\mathbb{K}} \to V_{\mathbb{K}}$ the complex conjugation with respect to the real form V, we get that $\sigma(L_J^{\pm}) = L_J^{\mp}$. The image $\mathcal{T}_V$ of the Borel embedding can be characterized as the set

$$\mathcal{T}_V = \{L \in \mathcal{L}(V_{\mathbb{K}}) \mid i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}|_{L\times L} \text{ is positive definite}\}.$$

The group $\mathrm{Sp}(V)$ acts by extension of scalars on $V_{\mathbb{K}}$, preserves the symplectic form $\langle\cdot,\cdot\rangle_{\mathbb{K}}$ and commutes with the complex conjugation σ; thus it acts on $\mathcal{T}_V$.

Lemma 2.1 *The algebraic map*

$$\mathbb{X}_V \to \mathcal{T}_V, \quad J \mapsto L_J^{+},$$

induces an $\mathrm{Sp}(V)$*–equivariant bijection.*

Proof If $v = x + iy$ is an eigenvector for the endomorphism $J \otimes \mathbb{I}_{\mathbb{K}}$ of eigenvalue i, it follows that $y = -Jx$. In particular, the restriction of $i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}$ satisfies

$$\begin{aligned} i\langle v,\sigma(v)\rangle_{\mathbb{K}} &= i\langle x, iJx\rangle_{\mathbb{K}} + i\langle -iJx, x\rangle_{\mathbb{K}} \\ &= 2\langle Jx, x\rangle, \end{aligned}$$

and this implies that the image of $\mathbb{X}_V$ is contained in $\mathcal{T}_V$.

Conversely, if $L \in \mathcal{L}(V_{\mathbb{K}})$ is such that $i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}|_{L\times L}$ is positive definite, L is transverse to $\sigma(L)$ since the restriction of the aforementioned Hermitian form to $\sigma(L)$ is negative definite. We denote by J_L the endomorphism of $V_{\mathbb{K}}$ defined by imposing that $J_L(v) = iv$ for each v in L and $J_L(v') = -iv'$ for each $v' \in \sigma(L)$.

Since any element w of V can be written uniquely as $w = v + \sigma(v)$ for some $v \in L$, and in particular, $J_L w = iv - i\sigma(v) = iv + \sigma(iv) \in V$, the endomorphism J_L preserves the real structure V. Let $J := J_L|_V$. Since the Hermitian form $i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}|_{L\times L}$ is by assumption positive definite, the quadratic form $\langle J\cdot,\cdot\rangle$ is positive definite.

If J is a point in $\mathbb{X}_V$ and g belongs to $\mathrm{Sp}(V)$, then since g commutes with σ, we get that $gL_J^{\pm}$ is the $\pm i$–eigenspace of $gJg^{-1}\otimes\mathbb{I}_{\mathbb{K}}$. It follows that the map $J\mapsto L_J^{+}$ is $\mathrm{Sp}(V)$–equivariant. □

The third and most concrete model for the Siegel space is the *upper half-space* $\mathcal{X}_{\mathbb{F}}$, a specific set of $\mathbb{K}$–valued symmetric matrices:

$$\mathcal{X}_{\mathbb{F}}=\{X+iY \mid X\in\mathrm{Sym}(n,\mathbb{F}),\ Y\in\mathrm{Sym}^+(n,\mathbb{F})\}.$$

Here $\mathrm{Sym}(n,\mathbb{F})$ denotes the vector space of symmetric $n\times n$ matrices with coefficients in $\mathbb{F}$ and $\mathrm{Sym}^+(n,\mathbb{F})$ denotes the properly convex cone in $\mathrm{Sym}(n,\mathbb{F})$ consisting of positive-definite symmetric matrices.

In order to establish a bijection between $\mathcal{T}_V$ and $\mathcal{X}_{\mathbb{F}}$, we fix a Lagrangian l_∞ in $\mathcal{L}(V)$, a complex structure $J\in\mathbb{X}_V$ and a basis $e_1,\dots,e_n$ of l_∞ which is orthonormal for $(\cdot,\cdot)_J$. The matrix representing the symplectic form with respect to the basis

$$\mathcal{B}=\{e_1,\dots,e_n,-Je_1,\dots,-Je_n\}$$

of V is $\left(\begin{smallmatrix}0 & \mathrm{Id}\\ -\mathrm{Id} & 0\end{smallmatrix}\right)$. Moreover, using the basis $\mathcal{B}$, we can associate to any $2n\times n$ matrix M of maximal rank the n–dimensional subspace of V spanned by the columns of M. We use this to give an explicit identification of $\mathrm{Sym}(n,\mathbb{K})$ with the affine chart of $\mathcal{L}(V_{\mathbb{K}})$ which consists of subspaces transverse to l_∞:

$$\iota\colon \mathrm{Sym}(n,\mathbb{K})\to\mathcal{L}(V_{\mathbb{K}}),\quad Z\mapsto\begin{pmatrix}Z\\ \mathrm{Id}\end{pmatrix}.$$

It is easy to verify that if we use the basis $\{-Je_1,\dots,-Je_n\}$ to identify the space $\mathrm{Sym}(n,\mathbb{K})$ with $\mathcal{Q}(Jl_\infty)$, we get that the map ι corresponds to the map j_{Jl_∞,l_∞} described in Section 2.1.

Since the restriction of $i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}$ to l_∞ is identically zero, every element l of $\mathcal{T}_V$ belongs to the image of ι, and it is easy to verify that the restriction of $i\langle\cdot,\sigma(\cdot)\rangle_{\mathbb{K}}$ to $\iota(X+iY)$ can be represented by the matrix $2Y$. In particular, ι restricts to a bijection between $\mathcal{T}_V$ and $\mathcal{X}_{\mathbb{F}}$. Notice that the restriction of ι to the subset of $\mathbb{F}$–valued symmetric matrices has image in $\mathcal{L}(V)$ and gives a parametrization of the affine chart of $\mathcal{L}(V)$ consisting of Lagrangians transverse to l_∞.

It follows from the identification between $\mathbb{X}_V$ and $\mathcal{X}_{\mathbb{F}}$ that the symplectic group $\mathrm{Sp}(2n,\mathbb{F})$ acts on $\mathcal{X}_{\mathbb{F}}$ by fractional linear transformations:

$$\begin{pmatrix}A & B\\ C & D\end{pmatrix}\cdot Z=(AZ+B)(CZ+D)^{-1}.$$

It will be useful in the following to record that, with our choice for a basis of the symplectic form, an element $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ belongs to $\mathrm{Sp}(2n,\mathbb{F})$ if and only if

$$ {}^tAD - {}^tCB = \mathrm{Id}, \quad {}^tAC = {}^tCA, \quad \text{and} \quad {}^tBD = {}^tDB. $$

In order to achieve transitivity of the symplectic group on the Siegel upper half-space, we need to restrict to real closed fields:

Definition 2.2 A *real closed field* is an ordered field $\mathbb{F}$ in which every positive element is a square and such that every polynomial in one variable over $\mathbb{F}$ factors into linear and quadratic factors.

Lemma 2.3 *If $\mathbb{F}$ is a real closed field, the symplectic group $\mathrm{Sp}(2n,\mathbb{F})$ acts transitively on $\mathcal{X}_{\mathbb{F}}$.*

Proof Since $\mathbb{F}$ is, by assumption, real closed, every symmetric matrix is diagonalizable by an orthogonal matrix, and as soon as it is positive definite, it admits a unique positive square root [9, Sections 2–4]. Let now $X+iY$ be a point in $\mathcal{X}_{\mathbb{F}}$ and let S be the square root of Y. We have

$$ X+iY = \begin{pmatrix} S & XS^{-1} \\ 0 & S^{-1} \end{pmatrix} \cdot i\,\mathrm{Id}. \qquad \square $$

2.3 Action on $\mathbb{F}$–Lagrangians

We now want to understand the action of $\mathrm{Sp}(V)$ on n–tuples of pairwise transverse Lagrangians. We denote this set by $\mathcal{L}(V)^{(n)}$:

$$ \mathcal{L}(V)^{(n)} = \{(l_1,\dots,l_n) \in \mathcal{L}(V)^n \mid l_i \pitchfork l_j\}. $$

It is a general fact that, for any field $\mathbb{F}$, the symplectic group acts transitively on pairs of transverse Lagrangians.

Lemma 2.4 *The symplectic group $\mathrm{Sp}(V)$ acts transitively on $\mathcal{L}(V)^{(2)}$.*

Recall from Section 2.1 that, whenever two transverse Lagrangians a,b are fixed, we have an identification $j_{a,b}\colon \mathcal{Q}(a) \cong \mathcal{L}(V)^b$, and we denote by $Q_{a,l,b}$ the inverse image $j_{a,b}^{-1}(l)$. Clearly for any element g in $\mathrm{Sp}(V)$, the quadratic forms Q_{l_1,l_2,l_3} and Q_{gl_1,gl_2,gl_3} are equivalent. As it turns out, the equivalence class of the quadratic form Q_{l_1,l_2,l_3} is a complete invariant of the triple (l_1,l_2,l_3) up to the symplectic group action:

Proposition 2.5 *The triples $(l_1, l_2, l_3), (m_1, m_2, m_3)$ in $\mathcal{L}(V)^{(3)}$ are equivalent modulo the symplectic group action if and only if the quadratic forms Q_{l_1,l_2,l_3} and Q_{m_1,m_2,m_3} are equivalent.*

Proof Since $\mathrm{Sp}(V)$ is transitive on pairs of transverse subspaces, we can assume that $l_1 = m_1 = a$ and $l_3 = m_3 = b$. The result now follows from the fact that the stabilizer in $\mathrm{Sp}(V)$ of the pair a, b is $\mathrm{GL}(n, \mathbb{F})$ acting on $\mathcal{Q}(a)$ by congruence. □

In particular, Sylvester's theorem allows us to count the number of $\mathrm{Sp}(V)$–orbits when the field $\mathbb{F}$ is real closed: since in this case, the signature $\mathrm{sign}(Q)$ is a complete invariant of a quadratic form Q up to equivalence (see [9, Theorem 9]), we have

Corollary 2.6 *Let $\mathbb{F}$ be a real closed field, and let V be a symplectic $\mathbb{F}$–vector space of dimension $2n$. Then there are $n+1$ orbits of $\mathrm{Sp}(V)$ in $\mathcal{L}(V)^{(3)}$.*

A fundamental tool in the study of Lagrangian subspaces is the Kashiwara cocycle, which, at least when $\mathbb{F} = \mathbb{R}$, is also known as the Maslov cocycle:

Definition 2.7 The Kashiwara cocycle is the function

$$\tau\colon \mathcal{L}^3(V) \to \mathbb{Z}, \quad (l_1, l_2, l_3) \mapsto \mathrm{sign}(Q),$$

where Q is the quadratic form on the direct sum $l_1 \oplus l_2 \oplus l_3$ defined by

$$Q(x_1 + x_2 + x_3) = \langle x_1, x_2 \rangle + \langle x_2, x_3 \rangle + \langle x_3, x_1 \rangle.$$

The following properties of the Kashiwara cocycle are well known:

Proposition 2.8 (see [17, Section 1.5]) *Let $(V, \langle\cdot,\cdot\rangle)$ be a $2n$–dimensional symplectic vector space over a real closed field.*

(1) *τ is alternating and invariant for the diagonal action of $\mathrm{Sp}(V)$ on $\mathcal{L}(V)^3$.*

(2) *τ has values in $\{-n, -n+1, \dots, n\}$. On triples consisting of pairwise transverse Lagrangians, it only achieves the values $\{-n, -n+2, \dots, n\}$. If $|\tau(l_1, l_2, l_3)| = n$, then the l_i are pairwise transverse.*

(3) *If (l_1, l_2, l_3) are pairwise transverse, then*

$$\tau(l_1, l_2, l_3) = \mathrm{sign}(Q_{l_1,l_2,l_3}).$$

(4) *τ is a cocycle: for each 4–tuple (l_1, l_2, l_3, l_4), we have*

$$\tau(l_2, l_3, l_4) - \tau(l_1, l_3, l_4) + \tau(l_1, l_2, l_4) - \tau(l_1, l_2, l_3) = 0.$$

The second and the third statement in Proposition 2.8 justify the following definition:

Definition 2.9 A triple $(l_1, l_2, l_3) \in \mathcal{L}(V)^{(3)}$ is *maximal* if Q_{l_1,l_2,l_3} is positive definite. More generally, an n–tuple $(l_1, \dots, l_n)$ is *maximal* if Q_{l_i,l_j,l_k} is positive definite for any ordered triple of indices $i < j < k$.

Let S^1 denote the unit circle in $\mathbb{C}$ endowed with its canonical orientation as boundary of the unit disc. Given a pair a, b of distinct points, $((a,b))$ denotes the connected component of $S^1 \setminus \{a, b\}$ consisting of the points crossed by a positively oriented C^1–path joining a to b. More generally, if $a, b \in \mathcal{L}(V)$ are transverse, we denote by $((a,b))$ the subset of $\mathcal{L}(V)$ consisting of points c such that the triple (a, c, b) is maximal:

$$((a,b)) = \{c \in \mathcal{L}(V) \mid (a, c, b) \in \mathcal{L}(V)^{(3)} \text{ is maximal}\}.$$

The key property of maximal triples that will be exploited throughout the paper is that they correspond to positive-definite quadratic forms:

Lemma 2.10 (1) *A triple (a, l, b) is maximal if and only if $l = j_{a,b}(q)$ for a positive-definite quadratic form $q \in \mathcal{Q}(a)$.*

(2) *A 4–tuple (l_1, l_2, l_3, l_4) is maximal if and only if it holds that $Q_{l_1,l_2,l_4} \gg 0$ and $Q_{l_1,l_3,l_4} - Q_{l_1,l_2,l_4} \gg 0$.*

Proof This follows from Proposition 2.8 together with the observation that the unipotent radical of the stabilizer in $\mathrm{Sp}(V)$ of b is isomorphic to $\mathrm{Sym}(n, \mathbb{F})$ and acts on $\mathcal{Q}(a)$ by translation. □

We finish this subsection by analyzing the $\mathrm{Sp}(V)$–orbits in $\mathcal{L}(V)^{(4)}$. Using the objects and notation introduced in Section 2.2, we fix a Lagrangian subspace l_∞, a complex structure J and a symplectic basis $\mathcal{B}$ of V of the form

$$\mathcal{B} = \{e_1, \dots, e_n, -Je_1, \dots, -Je_n\}.$$

Moreover, when this does not cause confusion, we suppress $\iota\colon \mathrm{Sym}(n, \mathbb{F}) \to \mathcal{L}(V)$ and simply represent an element in $\mathcal{L}(V)^{l_\infty}$ by an $\mathbb{F}$–valued symmetric matrix.

Proposition 2.11 *Let $\mathbb{F}$ be a real closed field, and let $(l_1, l_2, l_3, l_4) \in \mathcal{L}(V)^{(4)}$ be a maximal 4–tuple. Then there exist a diagonal matrix $D = \mathrm{diag}(d_1, \dots, d_n)$ satisfying $d_1 \geq \dots \geq d_n > 0$, and an element $g_1 \in \mathrm{Sp}(V)$ such that*

$$g_1(l_1, l_2, l_3, l_4) = (-\mathrm{Id}, 0, D, l_\infty).$$

Moreover, there exists $g_2 \in \mathrm{Sp}(V)$ such that $g_2(l_1, l_2, l_3, l_4) = (-\mathrm{Id}, \Lambda, 0, l_\infty)$, where Λ is diagonal with eigenvalues $-1 < \lambda_i = -d_i/(1 + d_i) < 0$.

Proof Since $\mathrm{Sp}(V)$ is transitive on maximal triples of Lagrangians and the triple $(-\mathrm{Id}, 0, l_\infty)$ is maximal, we have an element $g_1' \in \mathrm{Sp}(V)$ such that $g_1'(l_1, l_2, l_3, l_4) = (-\mathrm{Id}, 0, Z, l_\infty)$ for some positive-definite matrix Z.

It is easy to verify that the stabilizer of the triple $(-\mathrm{Id}, 0, l_\infty)$ in $\mathrm{Sp}(2n, \mathbb{F})$ consists of matrices that have the form

$$\mathrm{Stab}_{\mathrm{Sp}(2n,\mathbb{F})}(-\mathrm{Id}, 0, l_\infty) = \left\{ \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \mid A \in O(n) \right\}$$

with respect to the basis $\mathcal{B}$ and acts by congruence. This allows us to conclude: since $\mathbb{F}$ is real closed, every positive-definite matrix Z is orthogonally congruent to a diagonal matrix, namely there exists $A \in O(n)$ with $AZA^{-1} = D$, where $D = \mathrm{diag}(d_1, \dots, d_n)$ with $d_1 \geq \cdots \geq d_n$; see [9, Theorem 48]. Then

$$g_1 = \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} g_1'$$

satisfies the first assertion. For the second assertion, it is enough to take

$$g_2 = \begin{pmatrix} (\mathrm{Id}+D)^{-1/2} & -D(\mathrm{Id}+D)^{-1/2} \\ 0 & (\mathrm{Id}+D)^{1/2} \end{pmatrix} g_1. \qquad \square$$

An important role in the rest of the paper will be played by *Shilov hyperbolic* elements of $\mathrm{Sp}(V)$. We denote by $|\cdot| \colon \mathbb{K} \to \mathbb{F}^{\geq 0}$ the absolute value $|a + ib| = \sqrt{a^2 + b^2}$.

Definition 2.12 An element $g \in \mathrm{Sp}(V)$ is *Shilov hyperbolic* if there exists a g–invariant decomposition $V = L_g^+ \oplus L_g^-$, with $L_g^\pm \in \mathcal{L}(V)$, such that all eigenvalues of the restriction of g to L_g^- have absolute value strictly smaller than one and all eigenvalues of the restriction of g to L_g^+ have absolute value strictly bigger than one. In this case, we denote by M_g the restriction of g to L_g^+.

Remark 2.13 When V is a real vector space, the set of Lagrangians $\mathcal{L}(V)$ is the Shilov boundary of the symmetric space $\mathcal{T}_V$. Moreover, if $g \in \mathrm{Sp}(V)$ is Shilov hyperbolic, then there exists a Zariski open subset of $\mathcal{L}(V)$, the set of points transverse to L_g^-, which is contracted by g to L_g^+.

3 Representations admitting a maximal framing: the collar lemma

Let Σ be an oriented surface of negative Euler characteristic, genus g and p punctures. As mentioned in the introduction, we endow Σ with a complete hyperbolic metric of finite area and identify it with $\Gamma \backslash \mathbb{H}^2$ where $\mathbb{H}^2$ is the Poincaré upper half-plane.

We now turn to the study of representations $\rho\colon \Gamma \to \mathrm{Sp}(V)$ where V is a symplectic space over a real closed field $\mathbb{F}$. Recall from the introduction that we denote by $S \subseteq \partial\mathbb{H}^2$ any Γ–invariant subset containing all the fixed points of hyperbolic elements in Γ.

Definition 3.1 We say that the representation $\rho\colon \Gamma \to \mathrm{Sp}(V)$ admits a maximal S–framing if there exists an equivariant map $\phi\colon S \to \mathcal{L}(V)$ such that, whenever the triple (x, y, z) in S^3 is positively oriented, the triple of Lagrangians $(\phi(x), \phi(y), \phi(z))$ is maximal.

Remark 3.2 It is a fundamental result [6, Theorem 8] that if $\mathbb{F} = \mathbb{R}$, then any maximal representation admits a maximal framing. In addition, one can take $S = \partial\mathbb{H}^2$ and ϕ either left or right continuous.

In this section, we prove a generalization of the classical collar lemma of hyperbolic geometry to the context of representations which admit a maximal framing. In the case where $\mathbb{F}$ is the field of ordinary reals $\mathbb{R}$, this establishes a collar lemma for all maximal representations and gives a quantitative form of the fact due to Strubel [27] that for every hyperbolic element γ in Γ, the image $\rho(\gamma)$ is Shilov hyperbolic.

Theorem 3.3 (collar lemma) *If $\rho\colon \Gamma \to \mathrm{Sp}(V)$ is a representation admitting a maximal framing, then for every hyperbolic element γ, the image $\rho(\gamma)$ is Shilov hyperbolic. Let a, b be elements of Γ which intersect, and denote by $|\alpha_1| \geq \cdots \geq |\alpha_n| > 1$ the eigenvalues of the restriction of $\rho(a)$ to the attractive invariant Lagrangian $L^+_{\rho(a)}$, and analogously for $|\beta_1| \geq \cdots \geq |\beta_n| > 1$ and $\rho(b)$. Then:*

(1) $(\det M^{2/n}_{\rho(a)} - 1)(\det M^{2/n}_{\rho(b)} - 1) \geq 1$;

(2) $|\beta_1|^{2n} \geq \dfrac{1}{|\alpha_n|^2 - 1}$.

We isolate a useful lemma which is used many times in the proof:

Lemma 3.4 *Let $M \in \mathrm{GL}(n, \mathbb{F})$. Denote by $0 < \tau_n \leq \cdots \leq \tau_1$ the eigenvalues of M^tM and by $|\mu_n| \leq \cdots \leq |\mu_1|$ the absolute values of the eigenvalues of M. Then $\tau_n \leq |\mu_n|^2$ and $\tau_1 \geq |\mu_1|^2$.*

Proof If $S = M^tM$, then $S \gg 0$, and if $(\cdot, \cdot)$ denotes the standard scalar product, we have

$$\tau_n = \min_{v \neq 0} \frac{(Sv, v)}{(v, v)}.$$

Since $(Sv, v) = ({}^tMv, {}^tMv)$, we get $\tau_n \le ({}^tMv, {}^tMv)/(v, v)$ for every nonzero v. If now μ_n belongs to $\mathbb{F}$, we get the statement applying this inequality to a corresponding eigenvector of ${}^tM$. If instead $\mu_n \in \mathbb{K} \setminus \mathbb{F}$, then there is a two-dimensional subspace $E \cong \mathbb{F}^2$ in $\mathbb{F}^n$ which is invariant under ${}^tM$ and where this latter matrix acts like $\left(\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right)$ for some $a, b \in \mathbb{F}$ with $a^2 + b^2 = |\mu_n|^2$. Then for $\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right) \in E$, we have

$$\left({}^tM\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right), {}^tM\left(\begin{smallmatrix} x \\ y \end{smallmatrix}\right)\right) = (ax + by)^2 + (-bx + ay)^2 = (a^2 + b^2)(x^2 + y^2),$$

which implies the first assertion in the lemma. The second inequality follows from applying the first inequality to ${}^tM^{-1}$ and observing that

$$\max_{v \neq 0} \frac{(Sv, v)}{(v, v)} = \left(\min_{v \neq 0} \frac{(v, v)}{({}^tMv, {}^tMv)}\right)^{-1} = \left(\min_{v \neq 0} \frac{({}^tM^{-1}v, {}^tM^{-1}v)}{(v, v)}\right)^{-1}. \qquad \square$$

Proof of Theorem 3.3 Given two hyperbolic elements $a, b \in \Gamma$, we denote by $\mathrm{ax}(a)$ and $\mathrm{ax}(b)$ the axes of a and b, and by a^+ and b^+ (resp. a^- and b^-) the attractive (resp. repulsive) fixed points of a and b in $\partial\mathbb{H}^2$.

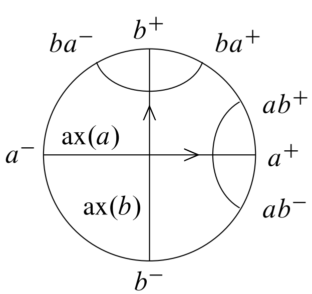
We can assume, without loss of generality, that a and b translate as represented by Figure 1 (left) and that the points $(a^-, b^-, ab^-, a^+, ab^+, ba^+, b^+, ba^-)$ are cyclically positively ordered; see [15, Lemma 2.2].

Let $\phi\colon S \to \mathcal{L}(V)$ be the maximal framing for ρ. Then the six points

$$\big(\phi(b^-),\ \phi(a^+),\ \rho(a)\phi(b^+),\ \rho(b)\phi(a^+),\ \phi(b^+),\ \phi(a^-)\big)$$

in $\mathcal{L}(V)^6$ form a maximal 6–tuple. This implies that they are pairwise transverse and every ordered subtriple forms a maximal triple.

We are going to perform our computations in the upper half-space model. As in Section 2.2, fix a symplectic basis $\{e_1, \dots, e_n, -Je_1, \dots, -Je_n\}$ of V, set $l_\infty = \langle e_1, \dots, e_n\rangle$ and parametrize the set of Lagrangians transverse to l_∞ by symmetric



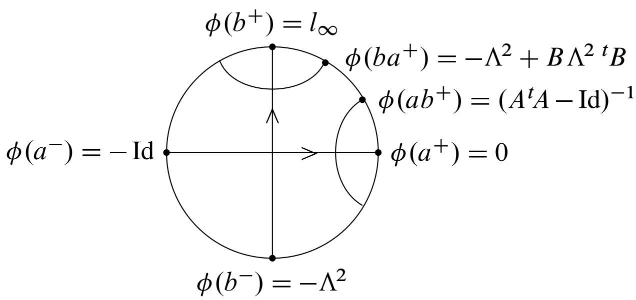


Figure 1: The points involved in the proof of Theorem 3.3

matrices. In view of Proposition 2.11, we may, modulo conjugating ρ, assume that the 4–tuple $(\phi(a^-),\phi(b^-),\phi(a^+),\phi(b^+))$ is equal to $(-\mathrm{Id},-\Lambda^2,0,l_\infty)$, where Λ is diagonal with eigenvalues $0<\lambda_i<1$. Since $\rho(a)$ fixes 0 and $-\mathrm{Id}$, and $\rho(b)$ fixes $-\Lambda^2$ and l_∞, we have

$$\rho(a)=\begin{pmatrix} {}^tA^{-1} & 0 \\ -{}^tA^{-1}+A & A\end{pmatrix} \quad \text{and} \quad \rho(b)=\begin{pmatrix} B & B\Lambda^2-\Lambda^2\,{}^tB^{-1} \\ 0 & {}^tB^{-1}\end{pmatrix}$$

for some matrices A, B. Let $\alpha_1,\dots,\alpha_n$ (resp. $\beta_1,\dots,\beta_n$) denote the eigenvalues of A (resp. B) counted with multiplicity and ordered so that $|\alpha_i|\ge|\alpha_{i+1}|$ (resp. $|\beta_i|\ge|\beta_{i+1}|$).

An easy computation gives

$$\begin{aligned}\rho(b)\phi(a^+)&=\rho(b)\cdot 0 &&=-\Lambda^2+B\Lambda^2\,{}^tB,\\ \rho(a)\phi(b^+)&=\rho(a)\cdot l_\infty &&=(A{}^tA-\mathrm{Id})^{-1}.\end{aligned}$$

We summarize this information in Figure 1 (right) for the reader's convenience.

The maximality of the triple

$$(\phi(a^+),\phi(ab^+),\phi(b^+))=(0,(A{}^tA-\mathrm{Id})^{-1},l_\infty)$$

implies that the quadratic form represented by $(A{}^tA-\mathrm{Id})^{-1}$ is positive definite, and in particular, all the eigenvalues of $A{}^tA$ are bigger than one. Thus if we denote by $\tau_1\ge\dots\ge\tau_n>1$ the eigenvalues of $A{}^tA$, it follows from Lemma 3.4 that $1<\tau_n\le|\alpha_n|^2$, and hence we get that the eigenvalues of A satisfy $1<|\alpha_n|\le\dots\le|\alpha_1|$; in particular, $\rho(a)$ is Shilov hyperbolic.

We now exploit the maximality of the triple

$$(\phi(a^+),\phi(ba^+),\phi(b^+))=(0,B\Lambda^2\,{}^tB-\Lambda^2,l_\infty),$$

which is equivalent to the fact that the quadratic form

$$\Lambda((\Lambda^{-1}B\Lambda)^t(\Lambda^{-1}B\Lambda)-\mathrm{Id})\Lambda=B\Lambda^2\,{}^tB-\Lambda^2$$

is positive definite. Denoting by C the matrix $\Lambda^{-1}B\Lambda$, we get that all the eigenvalues of $C\,{}^tC$ are bigger than 1. Let $1<\sigma_n\le\dots\le\sigma_1$ denote the eigenvalues of $C\,{}^tC$. From Lemma 3.4, we get that the eigenvalues of B satisfy $1<|\beta_n|\le\dots\le|\beta_1|$. This implies that $\rho(b)$ is Shilov hyperbolic as well. Moreover, we have

$$\sigma_1\le\det(C\,{}^tC)=\det(C)^2\le|\beta_1|^{2n}.$$

Last we exploit the maximality of the quadruple

$$(\phi(a^+),\phi(ab^+),\phi(ba^+),\phi(b^+))=(0,(A{}^tA-\mathrm{Id})^{-1},B\Lambda^2\,{}^tB-\Lambda^2,l_\infty),$$

which is equivalent to the property that

$$\Lambda\big((\Lambda^{-1}B\Lambda)^t(\Lambda^{-1}B\Lambda)-\mathrm{Id}\big)\Lambda-(A^tA-\mathrm{Id})^{-1}\gg 0; \tag{3}$$

see Lemma 2.10(2).

Taking into account that $1<\sigma_n\le\cdots\le\sigma_1$, we obtain that if $x_n\le\cdots\le x_1$ are the eigenvalues of

$$X=\Lambda\big((\Lambda^{-1}B\Lambda)^t(\Lambda^{-1}B\Lambda)-\mathrm{Id}\big)\Lambda=\Lambda(C^tC-\mathrm{Id})\Lambda,$$

then (3) implies

$$x_i\ge\frac{1}{\tau_{n+1-i}-1},\quad\text{for all } 1\le i\le n. \tag{4}$$

Next we claim that $x_i<(\sigma_i-1)$. Indeed, by the minmax theorem, we have

$$\begin{aligned}x_k&=\min_{\dim W=n+1-k}\ \max_{v\in W\setminus\{0\}}\frac{(\Lambda(C^tC-\mathrm{Id})\Lambda v,v)}{\|v\|^2}\\&=\min_{\dim W=n+1-k}\ \max_{v\in W\setminus\{0\}}\left(\frac{((C^tC-\mathrm{Id})\Lambda v,\Lambda v)}{\|\Lambda v\|^2}\frac{\|\Lambda v\|^2}{\|v\|^2}\right)\\&\le(\sigma_k-1)\max_{v\in\mathbb{F}^n\setminus\{0\}}\frac{\|\Lambda v\|^2}{\|v\|^2}=(\sigma_k-1)\lambda_1^2<\sigma_k-1,\end{aligned}$$

where the last inequality takes into account that $\lambda_1<1$.

Setting $i=1$ in the above inequalities, we obtain $\sigma_1-1\ge 1/(\tau_n-1)$ which, together with the inequalities previously obtained, namely that $|\beta_1|^{2n}\ge\sigma_1$ and $\tau_n\le|\alpha_n|^2$, shows assertion (2).

We establish now the inequality (1). Since $x_i<\sigma_i-1$, we get

$$(\det B)^2=\prod_{i=1}^n\sigma_i>\prod_{i=1}^n(1+x_i),$$

and we deduce from (4) that

$$\prod_{i=1}^n(1+x_i)\ge\prod_{i=1}^n\frac{\tau_i}{(\tau_i-1)}.$$

Since over any real closed field $\mathbb{F}$, one has $\prod_{i=1}^n(a_i^n-1)\le(a_1a_2\cdots a_n-1)^n$ for any $a_1,\dots,a_n>1$ (see Proposition A.1), we deduce, choosing $a_i=\tau_i^{1/n}$, that

$$\left(\prod_{i=1}^n\frac{\tau_i}{(\tau_i-1)}\right)^{1/n}\ge\frac{(\tau_1\cdots\tau_n)^{1/n}}{(\tau_1\cdots\tau_n)^{1/n}-1}.$$

Using $\tau_1\cdots\tau_n=(\det A)^2$, this establishes the first inequality. □

Remark 3.5 In the specific case of a maximal representation with values in $\mathrm{Sp}(2n,\mathbb{R})$ and which in addition belongs to the Hitchin component, assertion (2) is a weaker version of the collar lemma for Hitchin representations proven by Lee and Zhang [15]: their result implies, under these hypotheses, that

$$\beta_1^2 \geq \frac{\alpha_n^2}{(\alpha_n^2 - 1)}.$$

This is Proposition 2.12(1) in their paper.

3.1 Proper discontinuity on the non-Archimedean Siegel space

We now turn to the topological property of maximal S–framed representations stated in Corollary 1.11; this will follow from the following fact of independent interest.

Proposition 3.6 *There exists a continuous,* $\mathrm{Sp}(V)$*–invariant, multiplicative distance function*

$$D\colon \mathbb{X}_V \times \mathbb{X}_V \to \mathbb{F}_{\geq 1}$$

with the following property: for any $g \in \mathrm{Sp}(V)$ *and* $J \in \mathbb{X}_V$*, we have*

$$D(gJ, J) \geq |\lambda_1(g)|^2,$$

where $|\lambda_1(g)|$ *is the maximum modulus of an eigenvalue of* g*.*

More precisely, the properties of D alluded to in Proposition 3.6 are

- (MD1) $D(J_1, J_2) \geq 1$, with equality if and only if $J_1 = J_2$;
- (MD2) $D(J_1, J_2) = D(J_2, J_1)$ for all J_1, J_2;
- (MD3) $D(J_1, J_2) \leq D(J_1, J_3)D(J_3, J_2)$ for all J_1, J_2, J_3.

We begin with three observations concerning positive-definite forms Q_1, Q_2 on an $\mathbb{F}$–vector space W. We have that

$$\left|\frac{Q_2}{Q_1}\right| := \max_{x \neq 0} \frac{Q_2(x)}{Q_1(x)} \tag{5}$$

exists and coincides with the largest eigenvalue of the symmetric endomorphism S representing Q_2 with respect to Q_1. If moreover Q_1, Q_2 have the same determinant, that is $\det(S) = 1$, then

$$\left|\frac{Q_2}{Q_1}\right| \geq 1, \quad \text{with equality if and only if } Q_2 = Q_1. \tag{6}$$

The third observation is that if Q is positive definite and $g \in \mathrm{Sp}(V)$, then

$$\max_{x \neq 0} \frac{Q(gx)}{Q(x)} \geq |\lambda_1(g)|^2, \tag{7}$$

as follows immediately from Lemma 3.4 upon taking an orthonormal basis for Q.

Let $\mathbb{X}_V$ be the model of the Siegel space given by the set of compatible complex structures on V (see Section 2.2); given $J \in \mathbb{X}_V$, we let $Q_J(x) := \langle Jx, x \rangle$. Define, for $J_1, J_2 \in \mathbb{X}_V$,

$$D(J_1, J_2) := \max\left(\left| \frac{Q_{J_2}}{Q_{J_1}} \right|, \left| \frac{Q_{J_1}}{Q_{J_2}} \right| \right) \in \mathbb{F}_{\geq 1}.$$

Then D is well defined and continuous by (5); it verifies (MD1) as follows from (6). The $\mathrm{Sp}(V)$–invariance, as well as properties (MD2) and (MD3), are formal verifications. The inequality in Proposition 3.6 is then a direct consequence of (7).

Proof of Corollary 1.11 It follows from Corollary 1.10 and the assumption that Σ has no boundary that there exists $\epsilon > 0$ in $\mathbb{F}$ such that $|\lambda_1(\rho(\gamma))| \geq 1 + \epsilon$ for all $\gamma \in \Gamma \setminus \{e\}$. As a result, we have (Proposition 3.6)

$$D(\rho(\gamma)J, J) \geq (1+\epsilon)^2$$

for all $J \in \mathbb{X}_V$ and $\gamma \in \Gamma \setminus \{e\}$. It follows then from the fact that D is a continuous, $\mathrm{Sp}(V)$–invariant, multiplicative distance that

$$\mathcal{U} = \{(J_1, J_2) \in \mathbb{X}_V \times \mathbb{X}_V \mid D(J_1, J_2) < (1+\epsilon)\}$$

fulfills all the properties of Corollary 1.11. □

4 Cross-ratios and the geometry of $\mathbb{F}$–tubes

4.1 Cross-ratios

We now introduce a useful tool to study the geometry of the Siegel space. Let V be a $2n$–dimensional vector space over a field $\mathbb{L}$. Observe that if a, b are n–dimensional subspaces which are transverse ($a \pitchfork b$), then we have a direct sum decomposition $V = a \oplus b$, and thus we can define the projection $p_a^{/\!/b} \colon V \to a$ onto a parallel to b. Let now (l_1, l_2, l_3, l_4) be a quadruple in $\mathrm{Gr}_n(V)$ with the property that $l_1 \pitchfork l_2$ and $l_3 \pitchfork l_4$.

Definition 4.1 The cross-ratio of (l_1, l_2, l_3, l_4) is the endomorphism of l_1 defined by

$$R(l_1, l_2, l_3, l_4) = p_{l_1}^{/\!/l_2} \circ p_{l_4}^{/\!/l_3}|_{l_1}.$$

The cross-ratio has the following equivariance property: for all $g \in \mathrm{GL}(V)$, we have

$$R(gl_1, gl_2, gl_3, gl_4) = gR(l_1, l_2, l_3, l_4)g^{-1}.$$

It will be useful, in the following, to have an explicit expression for R once a basis $\mathcal{B} = \{e_1, \dots, e_{2n}\}$ of V is fixed. Recall that, as in Section 2.2, the choice of the basis $\mathcal{B}$ allows us to represent an element m of $\mathrm{Gr}_n(V)$ with a $2n \times n$ matrix M of maximal rank: the columns of the matrix M are understood to be the coordinates, with respect to $\mathcal{B}$, of a basis of m. With this notation, we have the following:

Lemma 4.2 *Let us assume that the columns of the matrix $\binom{X_i}{\mathrm{Id}_n}$ form a basis $\mathcal{B}_i$ of the n–dimensional vector space l_i. Then the expression for $R(l_1, l_2, l_3, l_4)$ with respect to the basis $\mathcal{B}_1$ of l_1 is given by*

$$R(l_1, l_2, l_3, l_4) = (X_1 - X_2)^{-1}(X_4 - X_2)(X_4 - X_3)^{-1}(X_1 - X_3).$$

Proof The matrix representing the linear map $p_{l_4}^{/\!/ l_3}|_{l_1}$ with respect to the bases $\mathcal{B}_1$ of l_1 and $\mathcal{B}_4$ of l_4 is the unique $A \in \mathrm{GL}(n, \mathbb{L})$ such that

$$\begin{pmatrix} X_1 \\ \mathrm{Id}_n \end{pmatrix} = \begin{pmatrix} X_4 \\ \mathrm{Id}_n \end{pmatrix} A + \begin{pmatrix} X_3 \\ \mathrm{Id}_n \end{pmatrix} (\mathrm{Id} - A).$$

Solving for A, we obtain

$$A = (X_4 - X_3)^{-1}(X_1 - X_3).$$

Notice that $X_4 - X_3$ is invertible since, by assumption, l_3 and l_4 are transverse.

Similarly, we get that the matrix representing the restriction of the linear map $p_{l_1}^{/\!/ l_2}$ to l_4 with respect to the bases $\mathcal{B}_4$ of l_4 and $\mathcal{B}_1$ of l_1 is given by

$$B = (X_1 - X_2)^{-1}(X_4 - X_2).$$

Since, by definition, the endomorphism $R(l_1, l_2, l_3, l_4)$ is the composition of $p_{l_1}^{/\!/ l_2}$ and $p_{l_4}^{/\!/ l_3}|_{l_1}$, and $p_{l_4}^{/\!/ l_3}|_{l_1}$ has image contained in l_4, we get that

$$R(l_1, l_2, l_3, l_4) = BA,$$

which gives the desired result. □

Let us now fix a basis $\mathcal{B}$ of V, set, as usual, $l_\infty = \langle e_1, \dots, e_n \rangle$ and represent with a matrix $M \in M(n, \mathbb{L})$ the subspace spanned by the columns of $\binom{M}{\mathrm{Id}}$. Here $M(n, \mathbb{L})$ is the set of $n \times n$ matrices. By a similar computation, we have

Lemma 4.3 *Assume 0, Z, X, l_∞ are pairwise transverse. Then*

$$R(0, Z, X, l_\infty) = Z^{-1}X.$$

It will be useful to understand how the cross-ratio varies with respect to permutations of the factors. In particular, we need to be able to compare endomorphisms of different vector spaces. Given two vector spaces l_1, l_2 of the same dimension, we say that two endomorphism $R_1 \in \mathrm{End}(l_1)$ and $R_2 \in \mathrm{End}(l_2)$ are *conjugate* if there exists an isomorphism $g\colon l_1 \to l_2$ such that $gR_1g^{-1} = R_2$. In this case, we write $R_1 \cong R_2$.

Lemma 4.4 *Assume that the subspaces l_i are pairwise transverse. Then*

(1) $R(l_1, l_2, l_4, l_3) = \mathrm{Id} - R(l_1, l_2, l_3, l_4)$;

(2) $R(l_4, l_1, l_2, l_3) \cong (\mathrm{Id} - R(l_1, l_2, l_3, l_4)^{-1})^{-1}$;

(3) $R(l_2, l_3, l_1, l_4) \cong R(l_1, l_4, l_2, l_3) = (\mathrm{Id} - R(l_1, l_2, l_3, l_4))^{-1}$.

Proof (1) By definition, we have

$$p_{l_1}^{/\!/l_2} \circ p_{l_4}^{/\!/l_3}|_{l_1} + p_{l_1}^{/\!/l_2} \circ p_{l_3}^{/\!/l_4}|_{l_1} = p_{l_1}^{/\!/l_2} \circ (p_{l_4}^{/\!/l_3} + p_{l_3}^{/\!/l_4})|_{l_1} = p_{l_1}^{/\!/l_2} \circ \mathrm{Id}\,|_{l_1} = \mathrm{Id}_{l_1}.$$

(2) Up to the $\mathrm{GL}(V)$ action, we can assume that $l_1 = 0$, $l_2 = Z$, $l_3 = X$ and $l_4 = l_\infty$. In particular, $R(0, Z, X, l_\infty) = Z^{-1}X$. In order to compute $R(l_\infty, 0, Z, X)$, we compute $p_{l_\infty}^{/\!/0}|_X = X$ and $p_X^{/\!/Z}|_{l_\infty} = (X - Z)^{-1}$.

(3) Similarly, one gets that $p_Z^{/\!/X}|_{l_\infty} = (Z - X)^{-1}$. The second equality follows from the fact that $p_0^{/\!/l_\infty}|_X = \mathrm{Id}$ and $p_X^{/\!/Z}|_0 = (\mathrm{Id} - Z^{-1}X)^{-1}$. □

4.2 $\mathbb{F}$–tubes

Let $(V, \langle \cdot, \cdot \rangle)$ be a symplectic vector space over a real closed field $\mathbb{F}$. Recall from Section 2.2 that $\mathbb{K}$ denotes the quadratic extension $\mathbb{F}[i]$, that $\sigma\colon \mathcal{L}(V_{\mathbb{K}}) \to \mathcal{L}(V_{\mathbb{K}})$ is induced by the complex conjugation with respect to the real structure V of $V_{\mathbb{K}}$ and that $\mathcal{T}_V$ is the model of the Siegel space contained in $\mathcal{L}(V_{\mathbb{K}})$. For any pair of transverse Lagrangians (a, b) in $\mathcal{L}(V)^{(2)}$, we introduce here an algebraic subset $\mathcal{Y}_{a,b}$ of the Siegel space $\mathcal{T}_V$ that is determined by the pair (a, b) and whose dimension is half the dimension of $\mathcal{T}_V$. We call such subsets $\mathbb{F}$–tubes. In the case when $\mathbb{F} = \mathbb{R}$, the subsets $\mathcal{Y}_{a,b}$ are Lagrangian submanifolds of the same rank as $\mathcal{X}_{\mathbb{R}}$; the $\mathbb{F}$–tube $\mathcal{Y}_{a,b}$ can be seen as the higher-rank generalization of a geodesic of the Poincaré model which is more suited to our purposes.

With the notation of Section 2.1, we define

$$\mathcal{Y}_{a,b} = \{l \in \mathcal{T}_V \mid R(a, l, \sigma(l), b) = -\mathrm{Id}\}.$$

Notice that requiring that an endomorphism of a vector space is equal to $-\mathrm{Id}$ does not depend on the choice of a basis. From the equivariance property of the cross-ratio and the fact that the symplectic group commutes with the complex conjugation σ, we deduce that

$$g\mathcal{Y}_{a,b} = \mathcal{Y}_{ga,gb} \quad \text{for any } g \in \mathrm{Sp}(V). \tag{8}$$

Our first goal is to give equations for $\mathcal{Y}_{a,b}$ in the Siegel upper half-space for some specific choice of the pair (a,b). In the sequel, if Z denotes a matrix with coefficients in $\mathbb{K}$, denote by $\bar{Z}$ the matrix obtained applying complex conjugation in $\mathbb{K}$ to all coefficients of Z. If Z is symmetric, this is the same as applying the complex conjugation σ to the corresponding Lagrangian.

Lemma 4.5 *The $\mathbb{F}$–tube with endpoints $0, l_\infty$ is*

$$\mathcal{Y}_{0,l_\infty} = \{iY \mid Y \in \mathrm{Sym}^+(n,\mathbb{F})\}.$$

Proof It follows from Lemma 4.3 that $R(0,Z,\sigma(Z),l_\infty) = Z^{-1}\bar{Z}$. Clearly we have $Z^{-1}\bar{Z} = -\mathrm{Id}$ if and only if $\bar{Z} = -Z$, and this concludes the proof. □

An immediate consequence of Lemma 4.5 and the equivariance property (8) is that if $\mathbb{F}$ is a real closed field, the stabilizer of $\mathcal{Y}_{a,b}$ is isomorphic to $\mathrm{GL}(n,\mathbb{F})$, and it acts transitively on $\mathcal{Y}_{a,b}$.

It will also be useful to have explicit expression for the set $\mathcal{Y}_{a,b}$ when a and b are transverse to l_∞. This has a particularly nice expression when $a = \langle e_1 - e_{n+1}, \dots, e_n - e_{2n}\rangle$ and $b = \langle e_1 + e_{n+1}, \dots, e_n + e_{2n}\rangle$:

Lemma 4.6 *If $a, b \in \mathcal{L}(V)$ correspond to the matrices $-\mathrm{Id}$ and Id, then*

$$\begin{aligned}\mathcal{Y}_{-\mathrm{Id},\mathrm{Id}} &= U(n) \cap \mathcal{X}_{\mathbb{F}} \\ &= \{X + iY \in \mathcal{X}_{\mathbb{F}} \mid YX = XY,\ X^2 + Y^2 = \mathrm{Id}\}.\end{aligned}$$

Proof Lemma 4.2 implies

$$R(-\mathrm{Id}, Z, \sigma(Z), \mathrm{Id}) = (-\mathrm{Id} - Z)^{-1}(\mathrm{Id} - Z)(\mathrm{Id} - \bar{Z})^{-1}(-\mathrm{Id} - \bar{Z}).$$

Since $\mathrm{Id} + \bar{Z}$ and $(\mathrm{Id} - \bar{Z})^{-1}$ commute, the equality $R(-\mathrm{Id}, Z, \sigma(Z), \mathrm{Id}) = -\mathrm{Id}$ reads

$$(\mathrm{Id} - Z)(\mathrm{Id} + \bar{Z}) = -(\mathrm{Id} + Z)(\mathrm{Id} - \bar{Z}),$$

which implies

$$\mathrm{Id} - Z + \bar{Z} - Z\bar{Z} = -\mathrm{Id} + \bar{Z} - Z + Z\bar{Z},$$

and hence, $Z\bar{Z} = Z^*Z = \mathrm{Id}$. □

As a consequence of the explicit parametrization of the sets $\mathcal{Y}_{0,l_\infty}$ and $\mathcal{Y}_{-\mathrm{Id},\mathrm{Id}}$, we obtain:

Proposition 4.7 *Assume that $\mathbb{F}$ is a real closed field. Let $(a,b,c,d) \in \mathcal{L}(V)^{(4)}$ be a maximal 4–tuple. The $\mathbb{F}$–tubes $\mathcal{Y}_{a,c}$ and $\mathcal{Y}_{b,d}$ meet exactly in one point.*

Proof Up to the symplectic group action, we can assume $(a,b,c,d)=(-\mathrm{Id},0,D,l_\infty)$ for some diagonal matrix $D=\mathrm{diag}(d_1,\dots,d_n)$ with $d_i>0$; see Proposition 2.11. Let y be a point in $\mathcal{Y}_{0,l_\infty}\cap\mathcal{Y}_{-\mathrm{Id},D}$. Since y belongs to $\mathcal{Y}_{0,l_\infty}$, we know that y has expression $y=iY$ for some positive-definite matrix Y. From the definition of $\mathcal{Y}_{-\mathrm{Id},D}$, we get

$$(-\mathrm{Id}-iY)^{-1}(D-iY)(D+iY)^{-1}(-\mathrm{Id}+iY)=-\mathrm{Id}.$$

This is equivalent to

$$(D-iY)(D+iY)^{-1}=(\mathrm{Id}+iY)(-\mathrm{Id}+iY)^{-1},$$

which in turn, using that $(\mathrm{Id}+iY)$ and $(-\mathrm{Id}+iY)^{-1}$ commute, is equivalent to

$$(-\mathrm{Id}+iY)(D-iY)=(\mathrm{Id}+iY)(D+iY).$$

This last equation reads $Y^2=D$, which has a unique positive solution. □

Remark 4.8 If the ordered field $\mathbb{F}$ is not real closed, one can similarly get that, if (a,b,c,d) is maximal, the $\mathbb{F}$–tubes $\mathcal{Y}_{a,c}$ and $\mathcal{Y}_{b,d}$ meet in at most one point.

4.3 Reflection with respect to $\mathcal{Y}_{a,b}$

In this subsection, we introduce a notion of orthogonality for $\mathbb{F}$–tubes and establish that the set of $\mathbb{F}$–tubes orthogonal to a fixed one foliate the space $\mathcal{T}_V$. Our main tool will be the characterization of $\mathcal{Y}_{a,b}$ as the fixed point set of an involution $\sigma_{a,b}$ which we now define. Let a,b be transverse Lagrangians in $\mathcal{L}(V)$. We consider the real form $V_{a,b}$ of $V_{\mathbb{K}}$ given by

$$V_{a,b}=\langle v+iw \mid v\in a,\ w\in b\rangle,$$

and denote by $\sigma_{a,b}$ the complex conjugation of $V_{\mathbb{K}}$ fixing $V_{a,b}$. The following properties of $\sigma_{a,b}$ can be checked easily:

Lemma 4.9 (1) *$\sigma_{a,b}$ is $\mathbb{K}$–antilinear;*

(2) *$\sigma_{a,b}\sigma=\sigma\sigma_{a,b}$, and in particular, $\sigma_{a,b}$ preserves V;*

(3) *$\langle\sigma_{a,b}(\cdot),\sigma_{a,b}(\cdot)\rangle_{\mathbb{K}}=-\overline{\langle\cdot,\cdot\rangle_{\mathbb{K}}}$;*

(4) *$g\sigma_{a,b}=\sigma_{ga,gb}\,g$, for every g in $\mathrm{Sp}(V)$.*

As a consequence of the first two facts of Lemma 4.9, we get that $\sigma_{a,b}$ induces a map on $\mathrm{Gr}_n(V)$ that, with a slight abuse of notation, will be also denoted by $\sigma_{a,b}$. The third fact of Lemma 4.9 implies that $\sigma_{a,b}$ restricts to a map

$$\sigma_{a,b}\colon \mathcal{L}(V_{\mathbb{K}}) \to \mathcal{L}(V_{\mathbb{K}}),$$

which preserves the subspaces we are interested in:

Lemma 4.10 *The involution $\sigma_{a,b}$ preserves the subspaces $\mathcal{T}_V$ and $\mathcal{L}(V)$ of $\mathcal{L}(V_{\mathbb{K}})$. It commutes with the cross-ratio.*

Proof Since the $\mathbb{F}$–linear map $\sigma_{a,b}$ preserves V, the induced map on $\mathcal{L}(V_{\mathbb{K}})$ preserves the subspace $\mathcal{L}(V)$. The fact that $\sigma_{a,b}$ induces a map of $\mathcal{T}_V$ can be seen from the following computation which uses Lemma 4.9(3): for every $v, w \in V_{\mathbb{K}}$,

$$\begin{aligned} i\,\langle \sigma_{a,b}(v), \sigma_{a,b}(w)\rangle_{\mathbb{K}} &= -i\,\overline{\langle v, w\rangle}_{\mathbb{K}} \\ &= \overline{i\,\langle v, w\rangle}_{\mathbb{K}}. \end{aligned}$$

In particular, the restriction of $i\,\langle\,\cdot\,, \sigma(\,\cdot\,)\rangle_{\mathbb{K}}$ to a Lagrangian $l \in \mathcal{L}(V_{\mathbb{K}})$ is positive definite if and only if its restriction to $\sigma_{a,b}(l)$ is.

For any pair $a, b \in \mathcal{L}(V)^{(2)}$ and for any 4–tuple (l_1, l_2, l_3, l_4) in the domain of the definition of R, we have

$$\sigma_{a,b} R(l_1, l_2, l_3, l_4)\sigma_{a,b} = R(\sigma_{a,b}(l_1), \sigma_{a,b}(l_2), \sigma_{a,b}(l_3), \sigma_{a,b}(l_4));$$

this follows from the equivariance property of the cross-ratio and that $\sigma_{a,b}^2 = \mathrm{Id}$. □

It is easy to check from the very definition of σ_{0,l_∞} that for any $Z \in \mathcal{X}_{\mathbb{F}}$, we have $\sigma_{0,l_\infty}(Z) = -\overline{Z}$. In particular, $\mathcal{Y}_{0,l_\infty} = \mathcal{T}_V \cap \mathrm{Fix}(\sigma_{0,l_\infty})$. An immediate corollary of the transitivity of the symplectic group action on $\mathcal{L}(V)^{(2)}$ is the following:

Corollary 4.11 *For any pair (a, b), we have $\mathcal{Y}_{a,b} = \mathcal{T}_V \cap \mathrm{Fix}(\sigma_{a,b})$.*

Another useful characterization of the $\mathbb{F}$–tubes is the following:

Lemma 4.12 *In the model $\mathbb{X}_V$,*

$$\mathcal{Y}_{a,b} = \{J \in \mathbb{X}_V \mid a \text{ and } b \text{ are orthogonal for } \langle J\cdot, \cdot\rangle\}.$$

Proof In the notation of Section 2, let $J \in \mathbb{X}_V$. Then $\sigma_{a,b}(L_J^+) = L_J^+$ if and only if $\sigma_{a,b}(L_J^-) = L_J^-$. Hence, since $\sigma_{a,b}$ is $\mathbb{K}$–antilinear, we deduce $\sigma_{a,b}(J \otimes \mathbb{I}_{\mathbb{K}}) = -(J \otimes \mathbb{I}_{\mathbb{K}})\sigma_{a,b}$, which, by restriction to $V = a \oplus b$, is equivalent to $\sigma_{a,b} J = -J\sigma_{a,b}$. The latter is equivalent to $J(a) = b$; that is, a and b are orthogonal with respect to $\langle J\cdot, \cdot\rangle$. □

The restriction of $\sigma_{a,b}$ to the subset of $\mathcal{L}(V)$ consisting of points that are transverse to a and b can also be characterized in term of the cross-ratio:

Proposition 4.13 *For each $c \in \mathcal{L}(V)$ transverse to a and b, we have that $\sigma_{a,b}(c)$ is the unique point satisfying*

$$R(a, c, \sigma_{a,b}(c), b) = -\operatorname{Id}.$$

Proof Up to the symplectic group action, we can assume that $a = 0$ and $b = l_\infty$. Since c is transverse to l_∞, it can be represented by a symmetric matrix S with coefficients in $\mathbb{F}$. The formula of Lemma 4.3 implies that $R(0, S, \sigma_{0,l_\infty}(S), l_\infty) = S^{-1}\sigma_{0,l_\infty}(S)$, and hence the unique point satisfying $R(0, S, \sigma_{0,l_\infty}(S), l_\infty) = -\operatorname{Id}$ is $-S$. □

When $\mathbb{F} = \mathbb{R}$ and the 4–tuple (a, b, c, d) is maximal, the two $\mathbb{R}$–tubes $\mathcal{Y}_{a,c}$ and $\mathcal{Y}_{b,d}$ are orthogonal as totally geodesic submanifolds of the Riemannian manifold $\mathcal{X}_\mathbb{R}$ precisely when $R(a, b, c, d) = 2\operatorname{Id}$. For arbitrary real closed fields, we take this property as a definition of orthogonality.

Definition 4.14 Let (a, b, c, d) be maximal. Two $\mathbb{F}$–tubes $\mathcal{Y}_{a,c}$ and $\mathcal{Y}_{b,d}$ are *orthogonal* if $R(a, b, c, d) = 2\operatorname{Id}$. In this case, we write $\mathcal{Y}_{a,c} \perp \mathcal{Y}_{b,d}$.

Notice that the orthogonality relation is symmetric since $R(d, a, b, c)$ is conjugate to $(\operatorname{Id} - R(a, b, c, d)^{-1})^{-1}$; see Lemma 4.4(2). The following lemma is a consequence of the property of the cross-ratio established in Lemma 4.4(1) and the characterization of the involution $\sigma_{a,b}$ in terms of the cross-ratio given in Proposition 4.13:

Lemma 4.15 *Let (a, b, c, d) be a maximal quadruple in $\mathcal{L}(V)^{(4)}$. The following are equivalent:*

(1) $\mathcal{Y}_{a,c} \perp \mathcal{Y}_{b,d}$;

(2) $d = \sigma_{a,c}(b)$;

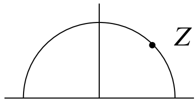
(3) $c = \sigma_{b,d}(a)$.

We now turn to an important geometric feature of the Siegel upper half-space, namely that the $\mathbb{F}$–tubes orthogonal to any fixed $\mathbb{F}$–tube foliate the whole space. We first verify this in a special case:

Proposition 4.16 *Assume that $\mathbb{F}$ is real closed. For any $Z = X + iY \in \mathcal{T}_V$, there exists a unique S in $\mathcal{L}(V)^{l_\infty}$ such that $(0, S, l_\infty)$ is maximal and $Z \in \mathcal{Y}_{-S,S}$. Moreover,*

$$S = Y^{1/2}\sqrt{\operatorname{Id} + (Y^{-1/2}XY^{-1/2})^2}\, Y^{1/2}.$$

Proof Given $Z = X + iY$, we look for a positive-definite matrix S with $Z \in \mathcal{Y}_{-S,S}$; see the following picture:



Denoting by $a(S^{-1/2})$ the element of $\mathrm{Sp}(2n, \mathbb{F})$ represented by the matrix $\left(\begin{smallmatrix} S^{-1/2} & 0 \\ 0 & S^{1/2} \end{smallmatrix}\right)$, we have $a(S^{-1/2})\mathcal{Y}_{-S,S} = \mathcal{Y}_{-\mathrm{Id},\mathrm{Id}}$. The condition $a(S^{-1/2})Z \in \mathcal{Y}_{-\mathrm{Id},\mathrm{Id}}$ leads, in view of the equations of Lemma 4.6, to

$$\begin{cases} (S^{-1/2}XS^{-1/2})(S^{-1/2}YS^{-1/2}) = (S^{-1/2}YS^{-1/2})(S^{-1/2}XS^{-1/2}), \\ (S^{-1/2}XS^{-1/2})^2 + (S^{-1/2}YS^{-1/2})^2 = \mathrm{Id}. \end{cases}$$

From the first equation, observing that Y is invertible, we get

$$XS^{-1} = YS^{-1}XY^{-1}.$$

Substituting this last equality in the second equation, and defining the matrix $V := Y^{-1/2}SY^{-1/2}$, we get

$$V^{-1}((Y^{-1/2}XY^{-1/2})^2 + \mathrm{Id}) = V,$$

which implies

$$V = \sqrt{\mathrm{Id} + (Y^{-1/2}XY^{-1/2})^2} \quad \text{and} \quad S = Y^{1/2}\sqrt{\mathrm{Id} + (Y^{-1/2}XY^{-1/2})^2}\, Y^{1/2}.$$

This shows the formula and implies uniqueness. □

Since all $\mathbb{F}$–tubes are $\mathrm{Sp}(V)$–conjugate, we obtain:

Corollary 4.17 *For any transverse pair $(a, b) \in \mathcal{L}(V)^{(2)}$ and any $z \in \mathcal{T}_V$, there exists a unique $c \in \mathcal{L}(V)$ such that (a, c, b) is maximal and z belongs to $\mathcal{Y}_{c,\sigma_{a,b}(c)}$.*

Corollary 4.17 allows us to define the orthogonal projection

$$\mathrm{pr}_{\mathcal{Y}_{a,b}} \colon \mathcal{T}_V \cup ((a, b)) \cup ((b, a)) \to \mathcal{Y}_{a,b}$$

as follows:

(1) if $c \in ((a, b)) \cup ((b, a))$, then we set $\mathrm{pr}_{\mathcal{Y}_{a,b}}(c) = \mathcal{Y}_{c,\sigma_{a,b}(c)} \cap \mathcal{Y}_{a,b}$;

(2) if $Z \in \mathcal{T}_V$, then we set $\mathrm{pr}_{\mathcal{Y}_{a,b}}(Z) = \mathcal{Y}_{c,\sigma_{a,b}(c)} \cap \mathcal{Y}_{a,b}$, where c is the unique Lagrangian in $\mathcal{L}(V)$ such that (a, c, b) is maximal and $Z \in \mathcal{Y}_{c,\sigma_{a,b}(c)}$.

It is easy to check that, when restricted to its set of definition in $\mathcal{L}(V)$, the orthogonal projection respects cross-ratios:

Lemma 4.18 *Let (a, b) be a pair of transverse Lagrangians, and let x, y be points in $((a, b))$. Then we have*

$$R(a, x, y, b) = R(a, \mathrm{pr}_{\mathcal{Y}_{a,b}}(x), \mathrm{pr}_{\mathcal{Y}_{a,b}}(y), b).$$

Proof Up to the symplectic group action, we can assume that $a=0$ and $b=l_\infty$. In that case, the result follows from the explicit formula for the cross-ratio and for the orthogonal projection. □

5 Reduction modulo an order convex subring

5.1 Order convex subrings

Let $\mathbb{F}$ be a real closed, non-Archimedean field. We denote by $\mathcal{O}<\mathbb{F}$ an *order convex* subring. This means that $\mathcal{O}$ is a subring with the additional property that for every positive element x in $\mathbb{F}$, if there exists y in $\mathcal{O}$ with $0<x<y$, then x belongs to $\mathcal{O}$ as well. It is easy to verify that, in this case, $\mathcal{O}$ is a local ring whose maximal ideal $\mathcal{I}$ is given by

$$\mathcal{I}=\{x\in\mathcal{O}\mid x^{-1}\notin\mathcal{O}\}.$$

We will denote by $\mathbb{F}_\mathcal{O}$ the quotient field $\mathbb{F}_\mathcal{O}:=\mathcal{O}/\mathcal{I}$. The field $\mathbb{F}_\mathcal{O}$ is real closed as well. The following examples of order convex subrings will play an important role in the sequel:

Example 5.1 Let $\sigma\in\mathbb{F}$ be an infinitesimal: this means that σ is a positive element satisfying $\sigma<1/n$ for any integer n. An example of an order convex subring of $\mathbb{F}$ is given by the set of elements comparable to σ:

$$\mathcal{O}_\sigma=\{x\in\mathbb{F}:|x|<\sigma^{-k}\text{ for some }k\in\mathbb{N}\};$$

in this case, the maximal ideal can also be characterized as

$$\mathcal{I}_\sigma=\{x\in\mathbb{F}:|x|<\sigma^{k}\text{ for all }k\in\mathbb{N}\}.$$

Example 5.2 Let us assume that $\mathbb{F}$ admits an order compatible valuation v. An example of order convex subring is given by the elements with nonnegative valuation

$$\mathcal{U}=\{x\in\mathbb{F}\mid v(x)\geq 0\},$$

and the maximal ideal can be characterized as

$$\mathcal{M}=\{x\in\mathbb{F}\mid v(x)>0\}.$$

5.2 $\mathcal{O}$–points

Let $\mathcal{O}$ be an order convex subring of $\mathbb{F}$, and let W be a finite-dimensional $\mathbb{F}$–vector space equipped with an $\mathbb{F}$–valued scalar product $(\cdot,\cdot)$. Then we set

$$W(\mathcal{O})=\{v\in W\mid(v,v)\in\mathcal{O}\}\quad\text{and}\quad W(\mathcal{I})=\{v\in W\mid(v,v)\in\mathcal{I}\}.$$

They are $\mathcal{O}$–submodules; if $e_1, \dots, e_m$ is any orthonormal basis of W, then one verifies

$$W(\mathcal{O}) = \sum_{i=1}^{m} \mathcal{O} e_i \quad \text{and} \quad W(\mathcal{I}) = \sum_{i=1}^{m} \mathcal{I} e_i .$$

This implies that the quotient $W_\mathcal{O} = W(\mathcal{O})/W(\mathcal{I})$ is an $\mathbb{F}_\mathcal{O}$–vector space of dimension $m = \dim(W)$, that the scalar product $(\cdot\,,\cdot)$ descends to a well-defined scalar product $(\cdot\,,\cdot)_\mathcal{O}$ on $W_\mathcal{O}$ and that, if $p_\mathcal{O}\colon W(\mathcal{O}) \to W_\mathcal{O}$ denotes the quotient map, $\{p_\mathcal{O}(e_1), \dots, p_\mathcal{O}(e_m)\}$ is again an orthonormal basis of $W_\mathcal{O}$. Notice, however, that the map $p_\mathcal{O}$ depends on the choice of the scalar product on W.

The subgroup

$$\mathrm{GL}(W)(\mathcal{O}) := \{g \in \mathrm{GL}(W) \mid g(W(\mathcal{O})) = W(\mathcal{O})\}$$

preserves $W(\mathcal{I})$, and we obtain this way a natural homomorphism $\pi_\mathcal{O}\colon \mathrm{GL}(W)(\mathcal{O}) \to \mathrm{GL}(W_\mathcal{O})$. The choice of an orthonormal basis of W induces an identification of the group $\mathrm{GL}(W)(\mathcal{O})$ with $\mathrm{GL}(m, \mathcal{O})$.

Let $\mathcal{Q}(W)$ be the vector space of $\mathbb{F}$–valued quadratic forms on W. As in Section 2, we associate to $f \in \mathcal{Q}(W)$ the symmetric bilinear form $b_f(\cdot\,,\cdot)$. We fix a basis $e_1, \dots, e_m$ of W which is orthonormal for $(\cdot\,,\cdot)$ and let $(A_f)_{ij} = b_f(e_i, e_j)$ be the associated symmetric matrix. We endow $\mathcal{Q}(W)$ with the scalar product $(f, g) = \mathrm{tr}(A_f A_g)$. Our next task is to understand the relationship between $\mathcal{Q}(W_\mathcal{O})$ and $\mathcal{Q}(W)_\mathcal{O}$.

Lemma 5.3 *For a quadratic form $f \in \mathcal{Q}(W)$, the following are equivalent:*

(1) $f \in \mathcal{Q}(W)(\mathcal{O})$;

(2) $f(W(\mathcal{O})) \subseteq \mathcal{O}$;

(3) $b_f(W(\mathcal{O}), W(\mathcal{O})) \subseteq \mathcal{O}$ *and* $b_f(W(\mathcal{O}), W(\mathcal{I})) \subseteq \mathcal{I}$.

Proof Clearly $\|f\|^2 = \mathrm{tr}(A_f^2) = \sum (A_f)_{ij}^2$ belongs to $\mathcal{O}$ if and only if $(A_f)_{ij}$ belongs to $\mathcal{O}$ for all i, j, which easily implies the desired equivalences. □

Thus, if f belongs to $\mathcal{Q}(W)(\mathcal{O})$, then b_f induces an $\mathbb{F}_\mathcal{O}$–valued bilinear symmetric form $\overline{b}_f$ on $W_\mathcal{O}$. In turn, $\overline{b}_f$ defines a quadratic form $\overline{f} \in \mathcal{Q}(W_\mathcal{O})$. If A_f is the matrix of f with respect to the orthonormal basis $\{e_1, \dots, e_m\}$, then the matrix $A_{\overline{f}}$ representing $\overline{f}$ with respect to the basis $\{p_\mathcal{O}(e_1), \dots, p_\mathcal{O}(e_m)\}$ is just the reduction modulo $\mathcal{I}$ of the matrix A_f. With this at hand, one verifies easily that the map

$$\overline{p}_\mathcal{O}\colon \mathcal{Q}(W)(\mathcal{O}) \to \mathcal{Q}(W_\mathcal{O}), \quad f \mapsto \overline{f},$$

induces an isomorphism of $\mathbb{F}_\mathcal{O}$–vector spaces

$$\mathcal{Q}(W)_\mathcal{O} \to \mathcal{Q}(W_\mathcal{O}).$$

We end the discussion concerning quadratic forms with the following remark:

Remark 5.4 Let $f \in \mathcal{Q}(W)$, and let $\{e_1, \dots, e_n\}$ be an orthonormal basis in which f is diagonal, that is, $b_f(e_i, e_j) = \lambda_i \delta_{ij}$. Let

$$\begin{aligned} m_f &= \operatorname{card}\{i \mid \lambda_i > 0\}, \\ n_f &= \operatorname{card}\{i \mid \lambda_i < 0\}, \\ z_f &= \operatorname{card}\{i \mid \lambda_i = 0\}. \end{aligned}$$

Then clearly $m_{\bar{f}} \leq m_f$, $n_{\bar{f}} \leq n_f$ and $z_{\bar{f}} \geq z_f$.

There is also a reduction process for Grassmannians, and it will play an important role for the construction of framings. Thus let $L \in \mathrm{Gr}_l(W)$ be an l–dimensional subspace of W. Then $L(\mathcal{O}) = L \cap W(\mathcal{O})$, and if $e_1, \dots, e_l$ is an orthonormal basis of L, we have $L(\mathcal{O}) = \mathcal{O}e_1 + \cdots + \mathcal{O}e_l$. This implies that the image $p_{\mathcal{O}}(L)$ of $L(\mathcal{O})$ in $W_{\mathcal{O}}$ is an $\mathbb{F}_{\mathcal{O}}$–vector subspace of dimension l. In this way, we obtain a map $q_{\mathcal{O}}\colon \mathrm{Gr}_l(W) \to \mathrm{Gr}_l(W_{\mathcal{O}})$ which is equivariant with respect to $\pi_{\mathcal{O}}\colon GL(W)(\mathcal{O}) \to \mathrm{GL}(W_{\mathcal{O}})$.

Remark 5.5 The map $q_{\mathcal{O}}$ does not preserve transversality: if $V = \mathbb{F}^2$ with the standard scalar product, and x is a nonzero element of $\mathcal{I}$, the two distinct lines $\mathbb{F} \cdot (1, 0)$ and $\mathbb{F} \cdot (1, x)$ of $\mathbb{P}V$ have the same image in $\mathbb{P}V_{\mathcal{O}}$.

We apply now the preceding remarks to the following situation. Let V be a $\mathbb{F}$–vector space with a symplectic form $\langle \cdot, \cdot \rangle$, and fix a compatible complex structure J. We will use the associated scalar product $(\cdot, \cdot) := \langle J \cdot, \cdot \rangle$ to define the $\mathcal{O}$ points. If L is a Lagrangian, then JL is orthogonal to L, and if $\{e_1, \dots, e_n\}$ is an orthonormal basis of L, the basis $\mathcal{B} = \{e_1, \dots, e_n, -Je_1, \dots, -Je_n\}$ is orthonormal and symplectic. With this at hand, one shows readily that $J \in \mathrm{Sp}(V)(\mathcal{O}) := \mathrm{Sp}(V) \cap \mathrm{GL}(V)(\mathcal{O})$, and that $\langle \cdot, \cdot \rangle$ induces a symplectic form $\langle \cdot, \cdot \rangle_{\mathcal{O}}$ of $V_{\mathcal{O}}$ compatible with $p_{\mathcal{O}}\colon V(\mathcal{O}) \to V_{\mathcal{O}}$. If in addition, one sets $J_{\mathcal{O}} = \pi_{\mathcal{O}}(J)$, then $J_{\mathcal{O}}$ is a complex structure on $V_{\mathcal{O}}$ compatible with $\langle \cdot, \cdot \rangle_{\mathcal{O}}$ and with associated scalar product $(\cdot, \cdot)_{\mathcal{O}}$. From the above, it follows that, if $L \in \mathrm{Gr}_n(V)$ is a Lagrangian, then $q_{\mathcal{O}}(L) \in \mathrm{Gr}_n(V_{\mathcal{O}})$ is a Lagrangian as well.

Lemma 5.6 *The map*

$$q_{\mathcal{O}}\colon \mathcal{L}(V) \to \mathcal{L}(V_{\mathcal{O}})$$

is surjective.

Proof Let L_0 be a k–dimensional totally isotropic subspace of V, and let $v_0 \in V$ be such that $\langle v, v_0 \rangle \in \mathcal{I}$ for all $v \in L_0$. Let $e_1, \dots, e_k$ be an orthonormal basis of L_0. By completing it to a symplectic basis of V, it is easy to verify that the map

$$V(\mathcal{I}) \to \mathcal{I}^k, \quad w \mapsto (\langle e_1, w \rangle, \dots, \langle e_k, w \rangle),$$

is surjective. Thus we can find $w_0 \in V(\mathcal{I})$ with $\langle e_i, v_0\rangle = \langle e_i, w_0\rangle$ for all $1 \le i \le k$. Then $v_1 = v_0 - w_0$ has the same projection in $V_{\mathcal{O}}$ as v_0 and is orthogonal to L_0 with respect to the symplectic form. The lemma follows then by recurrence on the dimension. □

5.3 Affine charts on Lagrangian Grassmannians and reduction modulo $\mathcal{I}$

Now we turn to a more detailed study of the map $q_{\mathcal{O}}$ and certain transversality properties. Recall from Section 2.1 that given transverse Lagrangians l_1, l_2 in V, we have a map

$$j_{l_1,l_2}\colon \mathcal{Q}(l_1) \to \mathcal{L}(V)^{l_2}$$

which to $f \in \mathcal{Q}(l_1)$ associates the Lagrangian

$$L_f = \{v + T_f v \mid v \in l_1\},$$

where $T_f\colon l_1 \to l_2$ is defined by the equation

$$b_f(v,w) = \langle v, T_f w\rangle = \langle w, T_f v\rangle, \quad v, w \in l_1.$$

If l_1, l_2 are orthogonal for $(\cdot,\cdot)$ and $\{e_1,\dots,e_n\}$ is an orthonormal basis of l_1, then $\{Je_1,\dots,Je_n\}$ is an orthonormal basis for l_2, and the symmetric matrix A_f of f in this basis is given by $(A_f)_{ij} = \langle e_i, T_f(e_j)\rangle = -(Je_i, T_f(e_j))$. Thus it follows from Lemma 5.3 that f belongs to $\mathcal{Q}(l_1)(\mathcal{O})$ if and only if the matrix coefficients of T_f with respect to the basis $\{e_1,\dots,e_n\}$ and $\{Je_1,\dots,Je_n\}$ are in $\mathcal{O}$, which in turn is equivalent to $T_f(l_1(\mathcal{O})) \subseteq l_2(\mathcal{O})$.

Lemma 5.7 *If l_1, l_2 are orthogonal Lagrangians in the symplectic vector space V, then $q_{\mathcal{O}}(l_1)$ and $q_{\mathcal{O}}(l_2)$ are orthogonal, and the diagram*

$$\begin{tikzcd}
\mathcal{Q}(l_1)(\mathcal{O}) \arrow[r] \arrow[d, "\bar{p}_{\mathcal{O}}"] & \mathcal{Q}(l_1) \arrow[r, "j_{l_1,l_2}", "\sim"'] & \mathcal{L}(V)^{l_2} \arrow[r] & \mathcal{L}(V) \arrow[d, "q_{\mathcal{O}}"] \\
\mathcal{Q}(q_{\mathcal{O}}(l_1)) \arrow[rr, "j_{q_{\mathcal{O}}(l_1),q_{\mathcal{O}}(l_2)}", "\sim"'] & & \mathcal{L}(V_{\mathcal{O}})^{q_{\mathcal{O}}(l_2)} \arrow[r] & \mathcal{L}(V_{\mathcal{O}})
\end{tikzcd}$$

commutes. The image under $q_{\mathcal{O}}$ of a Lagrangian that does not belong to $j_{l_1,l_2}(\mathcal{Q}(l_1)(\mathcal{O}))$ is not transverse to $q_{\mathcal{O}}(l_2)$.

Proof Since l_1 and l_2 are orthogonal, we have for $f \in \mathcal{Q}(l_1)$,

$$j_{l_1,l_2}(f)(\mathcal{O}) = \{v + T_f(v) \mid v \in l_1(\mathcal{O}),\ T_f(v) \in l_2(\mathcal{O})\}.$$

First notice that, if f belongs to $\mathcal{Q}(l_1)(\mathcal{O})$, then $T_f(l_1(\mathcal{O}))$ is contained in $l_2(\mathcal{O})$, and thus we get

$$j_{l_1,l_2}(f)(\mathcal{O}) = \{v + T_f(v) \mid v \in l_1(\mathcal{O})\}.$$

Now T_f induces a well-defined map $\overline{T}_f\colon q_{\mathcal{O}}(l_1) \to q_{\mathcal{O}}(l_2)$ with the property that

$$q_{\mathcal{O}}(j_{l_1,l_2}(f)) = \{v + \overline{T}_f(v) \mid v \in q_{\mathcal{O}}(l_1)\}.$$

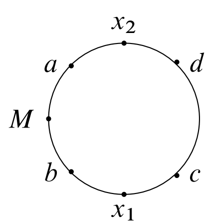
But $b_f(v,w)$ is, by definition, equal to $\langle v, T_f(w)\rangle$, and thus $b_{\bar{f}}(v,w)$ is equal to $\langle v, \overline{T}_f(w)\rangle_{\mathcal{O}}$ for $v, w \in q_{\mathcal{O}}(l_1)$. This implies that $j_{q_{\mathcal{O}}(l_1),q_{\mathcal{O}}(l_2)}(\overline{p}_{\mathcal{O}}(f))$ is equal to $q_{\mathcal{O}}(j_{l_1,l_2}(f))$ and proves the commutativity of the diagram.

If f does not belong to $\mathcal{Q}(l_1)(\mathcal{O})$, we can assume without loss of generality that $T_f(e_1)$ is not in $l_2(\mathcal{O})$. Writing $T_f(e_1) = \sum_{i=1}^n \mu_i Je_i$, let i_0 be such that $|\mu_{i_0}| = \max\{|\mu_i| : 1 \le i \le n\}$. Then μ_{i_0} does not belong to $\mathcal{O}$, and hence $\mu = \mu_{i_0}^{-1}$ belongs to $\mathcal{I}$. This implies that $T_f(\mu e_1)$ belongs to $l_2(\mathcal{O})$ and its e_{i_0} coordinate is equal to 1. Thus $\mu e_1 + T_f(\mu e_1)$ belongs to $j_{l_1,l_2}(f)(\mathcal{O})$, and

$$0 \neq p_{\mathcal{O}}(\mu e_1 + T_f(\mu e_1)) \in q_{\mathcal{O}}(j_{l_1,l_2}(f)) \cap q_{\mathcal{O}}(l_2). \qquad \square$$

Lemma 5.8 *Assume $(a,b,c,d) \in \mathcal{L}(V)^{(4)}$ is a maximal 4–tuple such that $q_{\mathcal{O}}(a)$ is transverse to $q_{\mathcal{O}}(b)$, and $q_{\mathcal{O}}(c)$ is transverse to $q_{\mathcal{O}}(d)$. Then for every $x_1 \in ((b,c))$ and $x_2 \in ((d,a))$, the subspace $q_{\mathcal{O}}(x_1)$ is transverse to $q_{\mathcal{O}}(x_2)$.*

Proof Pick $m \in ((q_{\mathcal{O}}(a), q_{\mathcal{O}}(b)))$ and $M \in \mathcal{L}(V)$ with $q_{\mathcal{O}}(M) = m$ (see Lemma 5.6). As a consequence of Remark 5.4 and the definition of the Kashiwara cocycle, we get that $M \in ((a,b))$. It follows then that (b, x_1, c, d, x_2, a) forms a maximal 6–tuple, and these six Lagrangians are all transverse to M, as illustrated in the following picture:



Thus these points are in the image of $j_{JM,M}\colon \mathcal{Q}(JM) \to \mathcal{L}(V)^M$. Denote by $f_l \in \mathcal{Q}(JM)$ the quadratic form with $j_{JM,M}(f_l) = l \in \mathcal{L}(V)^M$. We have from the maximality property of the 6–tuple that

$$f_b \ll f_{x_1} \ll f_c \ll f_d \ll f_{x_2} \ll f_a;$$

see Lemma 2.10(2). Applying now Lemma 5.7 to $l_1 = JM$ and $l_2 = M$, we deduce, from the fact that $q_{\mathcal{O}}(a)$ and $q_{\mathcal{O}}(b)$ are transverse to $m = q_{\mathcal{O}}(M)$, that f_a and f_b

are in $\mathcal{Q}(JM)(\mathcal{O})$. From the inequalities above, we deduce that f_{x_1} and f_{x_2} are in $\mathcal{Q}(JM)(\mathcal{O})$; it follows then from the commutativity of the diagram in Lemma 5.7 that $q_{\mathcal{O}}(x_1)$ and $q_{\mathcal{O}}(x_2)$ are transverse to $m = q_{\mathcal{O}}(M)$. Also,

$$f_{q_{\mathcal{O}}(x_2)} - f_{q_{\mathcal{O}}(x_1)} \gg f_{q_{\mathcal{O}}(d)} - f_{q_{\mathcal{O}}(c)} \gg 0,$$

where the last inequality follows from the hypothesis that $q_{\mathcal{O}}(d)$ is transverse to $q_{\mathcal{O}}(c)$. Thus $q_{\mathcal{O}}(x_2)$ is transverse to $q_{\mathcal{O}}(x_1)$. □

5.4 Choosing the scale and constructing the maximal framing

Let $\rho\colon \Gamma \to \mathrm{Sp}(V)$ be a representation admitting a maximal framing $\phi\colon S \to \mathcal{L}(V)$. We assume that there is a complex structure J in $\mathbb{X}_V$ and an order convex subring $\mathcal{O}$ of $\mathbb{F}$ such that $\rho(\Gamma) \subset \mathrm{Sp}(V)(\mathcal{O})$. We define then $\rho_{\mathcal{O}}\colon \Gamma \to \mathrm{Sp}(V_{\mathcal{O}})$ as the composition $\rho_{\mathcal{O}} := \pi_{\mathcal{O}} \circ \rho$ and $\phi_{\mathcal{O}}\colon S \to \mathcal{L}(V_{\mathcal{O}})$ as the composition $\phi_{\mathcal{O}} := q_{\mathcal{O}} \circ \phi$. Our goal is to show:

Theorem 5.9 *If ϕ is a maximal S–framing for $\rho\colon \Gamma \to \mathrm{Sp}(V)$, then $\phi_{\mathcal{O}}$ is a maximal S–framing for $\rho_{\mathcal{O}}\colon \Gamma \to \mathrm{Sp}(V_{\mathcal{O}})$.*

Remark 5.10 Since Γ is finitely generated, for any choice of a compatible complex structure J it is possible to find an infinitesimal σ such that $\rho(\Gamma) \subset \mathrm{Sp}(V)(\mathcal{O}_\sigma)$, where $\mathcal{O}_\sigma$ is the order convex subring described in Example 5.1. However, as we will discuss in Section 10, the choice of σ depends on the complex structure J; see Proposition 10.6.

In view of the definition of maximality of triples of Lagrangians and Remark 5.4, in order to prove Theorem 5.9, we have to show that if $x \neq y$ are distinct points in S, then $q_{\mathcal{O}}(\phi(x))$ and $q_{\mathcal{O}}(\phi(y))$ are transverse Lagrangians. As a first step, we show:

Lemma 5.11 *Assume that there exist two distinct points x, y in S such that $q_{\mathcal{O}}(\phi(x))$ and $q_{\mathcal{O}}(\phi(y))$ are not transverse. Then there exists a hyperbolic element $\gamma \in \Gamma$ such that $q_{\mathcal{O}}(\phi(\gamma^+))$ and $q_{\mathcal{O}}(\phi(\gamma^-))$ are not transverse.*

Proof From Lemma 5.8, it follows that we can choose I either $((x, y))$ or $((y, x))$ so that for every t_1, t_2 in I, we have that $q_{\mathcal{O}}(\phi(t_1))$ and $q_{\mathcal{O}}(\phi(t_2))$ are not transverse. Now pick a hyperbolic element $\gamma \in \Gamma$ with $\{\gamma^+, \gamma^-\} \subset I$. □

The strategy of the proof consists in showing that for every hyperbolic element $\gamma \in \Gamma$, the Lagrangians $q_{\mathcal{O}}(\phi(\gamma^-))$ and $q_{\mathcal{O}}(\phi(\gamma^+))$ are transverse. This will be a consequence of the properties of the eigenvalues of $\rho(\gamma)$ using the collar lemma.

We first observe that eigenvalues behave well with respect to reduction modulo $\mathcal{I}$:

Lemma 5.12 *Let $B \in \mathrm{GL}(m, \mathcal{O})$ be a matrix, and denote by $\beta_i \in \mathbb{K}$ the eigenvalues of B. Then:*

(1) $|\beta_i| \in \mathcal{O}$;

(2) *if $\overline{B}$ denotes the image of B in $\mathrm{GL}(m, \mathbb{F}_{\mathcal{O}})$, and $\overline{\beta}_i$ are the images of β_i in $\mathbb{K}_{\mathcal{O}}$, then the eigenvalues of $\overline{B}$ are precisely $\overline{\beta}_i$.*

Proof The first assertion follows from the fact that if β_i is an eigenvalue of B, then there exists a vector $v \in V(\mathcal{O}) \setminus V(\mathcal{I})$ such that $\|Bv\| = |\beta_i| \|v\|$; see Lemma 3.4. The second assertion follows from the fact that the characteristic polynomial of the reduction $\overline{B}$ is the reduction of the characteristic polynomial of B. □

Remark 5.13 Clearly, if g belongs to $\mathrm{GL}(V)(\mathcal{O})$, for each subspace W of V preserved by g, the restriction $g|_W$ belongs to $\mathrm{GL}(W)(\mathcal{O})$, and the restriction commutes with the reduction: $\pi_{\mathcal{O}}(g)|_{q_{\mathcal{O}}(W)} = \pi_{\mathcal{O}}(g|_W)$. However, it is worth pointing out that the Jordan decomposition of a matrix $B \in \mathrm{GL}(m, \mathcal{O})$ is not necessarily defined in $\mathrm{GL}(m, \mathcal{O})$, and in particular, the exponents of the minimal polynomial of a matrix B need not to be related with the exponents of the minimal polynomial of the reduction of B. For example, if ϵ belongs to $\mathcal{I}$, then the reduction of the not diagonalizable matrix $\begin{pmatrix} 2 & \epsilon \\ 0 & 2 \end{pmatrix}$ is diagonalizable, and the reduction of the diagonalizable matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1+\epsilon \end{pmatrix}$ is not diagonalizable.

This last example shows that generalized eigenspaces relative to distinct eigenvalues might not have transverse images in the quotient if the corresponding eigenvalues coincide modulo $\mathcal{I}$. We will now deduce from the collar lemma that, in case of maximal S–framed representations, the intermediate eigenvalues have distinct reductions:

Lemma 5.14 *Let $\rho\colon \Gamma \to \mathrm{Sp}(V)$ be a representation admitting a maximal framing. Assume that $\rho(\Gamma) \subset \mathrm{Sp}(V)(\mathcal{O})$. Then for every hyperbolic element $\gamma \in \Gamma$, we have*

$$|\lambda_n(\gamma)| - 1 \in \mathcal{O} \setminus \mathcal{I},$$

where $|\lambda_1(\gamma)| \geq \cdots \geq |\lambda_n(\gamma)| > 1$ are the eigenvalues of γ of absolute value greater than 1.

Proof Let $\delta \in \Gamma$ be a hyperbolic element with positive intersection number with γ, and let $\lambda_1(\delta)$ be the eigenvalue of $\rho(\delta)$ of largest modulus. If $|\lambda_n(\gamma)| < 2$, then the collar lemma (Theorem 3.3) implies

$$|\lambda_n(\gamma)| - 1 = \frac{|\lambda_n(\gamma)|^2 - 1}{|\lambda_n(\gamma)| + 1} \geq \frac{1}{3|\lambda_1(\delta)|^{2n}}.$$

Now observe that, since $\rho(\delta) \in \mathrm{Sp}(2n, \mathcal{O})$, we have that $|\lambda_1(\delta)|$ belongs to $\mathcal{O}$, from which the claim follows. □

We have now all the necessary ingredients to prove Theorem 5.9:

Proof of Theorem 5.9 Let us assume by contradiction that there exist x, y in S with $q_{\mathcal{O}}(\phi(x))$ nontransverse to $q_{\mathcal{O}}(\phi(y))$. As a consequence of Lemma 5.11, we can find a hyperbolic element γ in Γ such that $q_{\mathcal{O}}(\phi(\gamma^+))$ is nontransverse to $q_{\mathcal{O}}(\phi(\gamma^-))$.

If now $|\lambda_1(\gamma)| \geq \cdots \geq |\lambda_n(\gamma)| > 1$ are the absolute values of the eigenvalues of $\rho(\gamma)|_{\phi(\gamma^+)}$, counted with multiplicity, then it follows from Lemmas 5.14 and 5.12 that the absolute values $|\overline{\lambda_1(\gamma)}| \geq \cdots \geq |\overline{\lambda_n(\gamma)}| > 1$ of the eigenvalues of the restriction of $\rho_{\mathcal{O}}(\gamma)$ to $q_{\mathcal{O}}(\phi(\gamma^+))$ are all strictly larger than 1. Since $|\overline{\lambda_1(\gamma)}|^{-1} \leq \cdots \leq |\overline{\lambda_n(\gamma)}|^{-1} < 1$ are then the absolute values of the eigenvalues of the restriction of $\rho_{\mathcal{O}}(\gamma)$ to $q_{\mathcal{O}}(\phi(\gamma^-))$, this implies that the $\rho_{\mathcal{O}}$–invariant vector space $q_{\mathcal{O}}(\phi(\gamma^+)) \cap q_{\mathcal{O}}(\phi(\gamma^-))$ must be zero since otherwise $\rho_{\mathcal{O}}(\gamma)$ would have at least a nonzero eigenvalue which would be an element in $\mathbb{K}_{\mathcal{O}}$ both of absolute value strictly larger and smaller than 1. Thus $q_{\mathcal{O}}(\phi(\gamma^+)) \cap q_{\mathcal{O}}(\phi(\gamma^-)) = 0$, which is a contradiction. Hence, for every $x \neq y$ in S, we have that $q_{\mathcal{O}}(\phi(x))$ is transverse to $q_{\mathcal{O}}(\phi(y))$. □

6 Fields with valuation and the projection to the building

In this section, $\mathbb{F}$ will denote an ordered field with a compatible valuation $v\colon \mathbb{F} \to \mathbb{R} \cup \{\infty\}$, meaning that we require $v(y) \leq v(x)$ whenever $0 \leq x \leq y$.

Example 6.1 (compare Example 5.1) Let $\mathbb{E}$ be an ordered field, $\sigma \in \mathbb{E}$ an infinitesimal and $\mathcal{O}_\sigma$ the order convex local subring consisting of elements comparable with σ. On $\mathcal{O}_\sigma$, we define the valuation

$$v_\sigma(x) = \sup\{t \in \mathbb{R} : |x| \leq \sigma^t\}.$$

Then v_σ passes to the quotient $\mathbb{E}_\sigma := \mathcal{O}_\sigma / \mathcal{I}_\sigma$ by the maximal ideal $\mathcal{I}_\sigma$ and defines an order compatible valuation.

We introduce on $\mathbb{F}$ the norm $\|x\| := e^{-v(x)}$. This defines an ultrametric norm on $\mathbb{F}$ with valuation ring $\mathcal{U} := \{x \in \mathbb{F} : \|x\| \leq 1\}$ whose maximal ideal is $\mathcal{M} := \{x \in \mathbb{F} : \|x\| < 1\}$. Observe that since the valuation is order compatible, the norm is order compatible as well: if $0 < x < y$, then $\|x\| \leq \|y\|$.

Let $(V, \langle \cdot, \cdot \rangle)$ be a symplectic vector space over $\mathbb{F}$, $J_0 \in \mathbb{X}_V$ a compatible complex structure and $(\cdot, \cdot)_{J_0}$ the corresponding scalar product. We denote by $\mathcal{B}_V$ the affine building associated to $\mathrm{Sp}(V)$; see [20, Section 3.2; 13, Theorem 4.3]. It is well known that the set of vertices $\mathcal{B}_V^0$ of $\mathcal{B}_V$ can be identified with the homogeneous space $\mathrm{Sp}(V)/\mathrm{Sp}(V)(\mathcal{U})$, where, as in Section 5, we define

$$V(\mathcal{U}) = \{v \in V \mid (v,v) \in \mathcal{U}\}$$

and

$$\mathrm{Sp}(V)(\mathcal{U}) = \{g \in \mathrm{Sp}(V) \mid g(V(\mathcal{U})) = V(\mathcal{U})\}.$$

The stabilizer of the complex structure $J_0 \in \mathbb{X}_V$ is

$$\begin{aligned} U(J_0) &= \{g \in \mathrm{Sp}(V) \mid gJ_0g^{-1} = J_0\} \\ &= \{g \in \mathrm{Sp}(V) \mid g \text{ preserves the scalar product } (\cdot,\cdot)_{J_0}\}, \end{aligned}$$

and hence is contained in $\mathrm{Sp}(V)(\mathcal{U})$. As a result, we can define the projection

$$\pi_{\mathcal{B}}\colon \mathbb{X}_V = \mathrm{Sp}(V)/U(J_0) \to \mathcal{B}_V^0 = \mathrm{Sp}(V)/\mathrm{Sp}(V)(\mathcal{U}).$$

Remark 6.2 Parreau [20] gave an explicit description of the building associated to $\mathrm{SL}(2n,\mathbb{F})$ as the space of good norms on $\mathbb{F}^{2n}$ of determinant one. It is possible to verify that, considering the affine building associated to $\mathrm{Sp}(2n,\mathbb{F})$ as a subbuilding of the affine building associated to $\mathrm{SL}(2n,\mathbb{F})$, the map $\pi_{\mathcal{B}}$ corresponds to the map that associates to a point $J \in \mathbb{X}_V$ the corresponding good norm $\eta_J(v) = \|(v,v)_J\|$.

For $\mathbb{F} = \mathbb{R}$, Siegel [26] gave explicit formulas for the Riemannian distance on $\mathcal{X}_{\mathbb{R}}$. We use the cross-ratio R defined in Section 4.1 to define in our context a distance-like function as follows. Observe that, given $X, W \in \mathcal{T}_V$, the cross-ratio $R(X,\sigma(W),W,\sigma(X))$ is always well defined: indeed, the Hermitian form $i\langle\cdot,\sigma(\cdot)\rangle$ is positive definite on X and W and negative definite on $\sigma(W)$ and $\sigma(X)$; in particular, X and $\sigma(W)$ are transverse, and so are W and $\sigma(X)$. Moreover, all the eigenvalues of the cross-ratio $R(X,\sigma(W),W,\sigma(X))$ belong to $\mathbb{F}$ and are between 0 and 1: indeed, since $\mathbb{F}$ is real closed, for each pair $X, W \in \mathcal{X}_{\mathbb{F}}$, we can find $g \in \mathrm{Sp}(V)$ such that $g \cdot X = i\,\mathrm{Id}$ and $g \cdot W = iD$ for a diagonal matrix D with positive entries, and we have

$$gR(X,\sigma(W),W,\sigma(X))g^{-1} = R(i\,\mathrm{Id}, -iD, iD, -i\,\mathrm{Id}) = \frac{(\mathrm{Id}-D)^2}{(\mathrm{Id}+D)^2}.$$

We can thus define

$$d(Z,W) = \sqrt{\sum_{i=1}^{n}\left(\ln\left\|\frac{1+\sqrt{r_i}}{1-\sqrt{r_i}}\right\|\right)^2}, \tag{9}$$

where $r_1,\dots,r_n$ are the eigenvalues of $R(X,\sigma(W),W,\sigma(X))$.

In the case we considered above, where $X = i\,\mathrm{Id}$ and $W = iD$, (9) specializes to

$$d(i\,\mathrm{Id}, iD) = \sqrt{\sum_{i=1}^{n} (\ln \|d_i\|)^2},$$

where $d_1, \dots, d_n$ are the entries of D.

The function d is clearly $\mathrm{Sp}(V)$–invariant since the eigenvalues of the cross-ratio are. Denote by $d_{\mathcal{B}}$ the CAT(0) distance on $\mathcal{B}_V$. Using the transitivity of the symplectic group on apartments in $\mathcal{B}_V$ and the invariance of d, one verifies:

Proposition 6.3 *For any $X, Y \in \mathcal{T}_V$, we have*

$$d_{\mathcal{B}}(\pi_{\mathcal{B}}(X), \pi_{\mathcal{B}}(Y)) = d(X, Y).$$

As a result, we get that d is a pseudodistance on $\mathcal{T}_V$, and $\mathcal{B}_V$ is the Hausdorff quotient of $\mathcal{T}_V$ modulo this pseudodistance.

We will denote by $L_{\mathcal{B}}(g)$ the translation length of an element $g \in \mathrm{Sp}(V)$ considered as an isometry of the affine building $\mathcal{B}_V$.

7 On elements with fixed points

We place ourselves in the framework of Section 6 and consider a representation $\rho\colon \Gamma \to \mathrm{Sp}(V)$ admitting a maximal framing (S, ϕ). In this section, we want to analyze how elements of Γ which have zero translation length in the building $\mathcal{B}_V$ interact. As a crucial step in the analysis, we associate to any such $\gamma \in \Gamma$ a pair (b_γ^+, b_γ^-) of points in $\mathcal{B}_V$ which are fixed by $\rho(\gamma)$ and are canonically constructed from the maximal framing ϕ.

Recall from Section 6 that we denote by $\pi_{\mathcal{B}}\colon \mathcal{T}_V \to \mathcal{B}_V$ the $\mathrm{Sp}(V)$–equivariant projection from the Siegel upper half-space to the affine building associated to $\mathrm{Sp}(V)$, and given an element $g \in \mathrm{Sp}(V)$, we denote by $L_{\mathcal{B}}(g)$ the translation length of g on $\mathcal{B}_V$. Moreover, for ease of notation, we will denote by $\mathcal{Y}_\gamma$ the $\mathbb{F}$–tube $\mathcal{Y}_{\phi(\gamma^-),\phi(\gamma^+)}$ and by $\mathbb{Y}_\gamma$ its projection to $\mathcal{B}_V$:

$$\mathbb{Y}_\gamma = \pi_{\mathcal{B}}(\mathcal{Y}_\gamma).$$

It follows from the equivariance of $\pi_{\mathcal{B}}$ that $\mathbb{Y}_\gamma$ is a subbuilding of $\mathcal{B}_V$ associated to a subgroup of $\mathrm{Sp}(V)$ isomorphic to $\mathrm{GL}(n, \mathbb{F})$. Recall from Section 4.3 that given any pair of transverse Lagrangians $a, b \in \mathcal{L}(V)$, we defined an orthogonal projection

$$\mathrm{pr}_{\mathcal{Y}_{a,b}}\colon ((a,b)) \cup ((b,a)) \to \mathcal{Y}_{a,b}.$$

The first goal of the section is to prove:

Proposition 7.1 *Let $\gamma \in \Gamma$ be an element which is not boundary parallel. Assume that $L_{\mathcal{B}}(\rho(\gamma)) = 0$. Then both maps*

$$F_\gamma^+\colon ((\gamma^-, \gamma^+)) \to \mathbb{Y}_\gamma, \quad x \mapsto \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(x)))$$

and

$$F_\gamma^-\colon ((\gamma^+, \gamma^-)) \to \mathbb{Y}_\gamma, \quad x \mapsto \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(x)))$$

are constant.

Denoting by b_γ^+ (resp. b_γ^-) the constant images of the maps F_γ^+ (resp. F_γ^-) in Proposition 7.1 we have:

Corollary 7.2 *The points b_γ^+ and b_γ^- are fixed by $\rho(\gamma)$.*

If $\gamma \in \Gamma$ corresponds to a simple closed geodesic, it is possible to construct examples of representations $\rho\colon \Gamma \to \mathrm{Sp}(V)$ such that the points b_γ^+ and b_γ^- are different. The second main result of the section gives sufficient conditions for the two points to coincide:

Proposition 7.3 *Assume that γ and η in Γ are hyperbolic elements with intersecting axes, and that $L_{\mathcal{B}}(\rho(\gamma)) = L_{\mathcal{B}}(\rho(\eta)) = 0$. Then*

$$b_\gamma^+ = b_\gamma^- = b_\eta^+ = b_\eta^- = \pi_{\mathcal{B}}(\mathcal{Y}_\gamma \cap \mathcal{Y}_\eta).$$

Corollary 7.4 *Assume that $L_{\mathcal{B}}(\rho(\gamma)) = 0$. If the closed geodesic corresponding to γ is not simple, then $b_\gamma^+ = b_\gamma^-$.*

Before proceeding to the proofs of Propositions 7.1 and 7.3, we observe that in certain situations, one can get a uniform lower bound on the translation lengths $L_{\mathcal{B}}(\rho(\gamma))$ for all hyperbolic elements γ crossing a given hyperbolic element η. This is in fact an immediate corollary of the collar lemma:

Corollary 7.5 *Assume that $\eta \in \Gamma$ is a hyperbolic element, and let us denote by $|\lambda_1(\eta)| \geq \cdots \geq |\lambda_n(\eta)| > 1$ the eigenvalues of $\rho(\eta)$ of absolute value larger than 1. If $\delta = \| |\lambda_n(\eta)| - 1 \| < 1$, then for any element γ intersecting η, we have*

$$L_{\mathcal{B}}(\rho(\gamma)) \geq \frac{1}{2n\delta}.$$

In particular, if the closed geodesic represented by η is not simple, $\| |\lambda_n(\eta)| - 1 \| \geq 1$.

Proposition 7.3 also allows us to give sufficient conditions for a representation ρ to have a global fixed point. We say that a generating set X for Γ is connected if the graph (X, E), where E consists of the pairs (s_1, s_2) of elements of X whose axes intersect, is connected.

Corollary 7.6 *Let X be any connected generating set for Γ. If $\rho\colon \Gamma \to \mathrm{Sp}(V)$ is a representation admitting a maximal framing, the following are equivalent:*

(1) *ρ has a global fixed point in $\mathcal{B}_V$;*

(2) *$L_{\mathcal{B}}(\rho(s)) = 0$ for all $s \in X$.*

Remark 7.7 There exist connected generating sets consisting of $2g$ simple closed curves. In particular, Corollary 7.6 refines, in our setting, [22, Corollary 3].

Recall from Section 2.3 that we say that $g \in \mathrm{Sp}(V)$ is Shilov hyperbolic if there exists a g–invariant decomposition $V = L_g^+ \oplus L_g^-$ such that all the eigenvalues of the restriction M_g of g to L_g^+ are in absolute value strictly greater than one. It is however worth remarking that, in general, g does not necessarily have a hyperbolic dynamic on $\mathcal{L}(V)$. It follows from Lemma 5.14 that, as soon as ρ admits a maximal framing, for any hyperbolic element $\gamma \in \Gamma$, its image $\rho(\gamma)$ is Shilov hyperbolic.

Lemma 7.8 *Let $g \in \mathrm{Sp}(V)$ be Shilov hyperbolic, and let $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{F}[i]$ be the set of eigenvalues of M_g. Then*

$$L_{\mathcal{B}}(g) = 2\sqrt{\sum_{i=1}^{n} (\ln \|\lambda_i\|)^2}.$$

Proof Since g is Shilov hyperbolic, it stabilizes the $\mathbb{F}$–tube $\mathcal{Y}_{L_g^+, L_g^-}$, and similarly it stabilizes the projection

$$\mathbb{Y}_{L_g^+, L_g^-} = \pi_{\mathcal{B}}(\mathcal{Y}_{L_g^+, L_g^-}).$$

This latter is a subbuilding of $\mathcal{B}_V$ associated to $\mathrm{GL}(n, \mathbb{F})$. The desired statement then follows from [20]. □

Lemma 7.9 *Let $g \in \mathrm{Sp}(V)$ be Shilov hyperbolic. Then the following are equivalent:*

(1) *$L_{\mathcal{B}}(g) = 0$;*

(2) *$\|\det M_g\| = 1$;*

(3) *$\|\det R(L_g^+, S, gS, L_g^-)\| = 1$ for every S in $((L_g^+, L_g^-))$.*

Proof In view of Lemma 7.8, we have that

$$L_{\mathcal{B}}(g) = 2\sqrt{\sum_{i=1}^{n} (\ln \|\lambda_i\|)^2},$$

while

$$\|\det M_g\| = \prod_{i=1}^{n} \|\lambda_i\| \quad \text{and} \quad \det R(L_g^+, S, gS, L_g^-) = (\det M_g)^{-2}.$$

The equivalence follows easily from the assumption that $|\lambda_i| > 1$ for all i and the order compatibility of the norm. □

Lemma 7.10 *Let us assume that the 5–tuple of Lagrangians $(x_1, x_2, x_3, x_4, x_5)$ is maximal. Then*

$$\det R(x_1, x_2, x_3, x_5) \leq \det R(x_1, x_2, x_4, x_5).$$

Proof We may assume $x_1 = 0$ and $x_5 = l_\infty$; then we have $0 \ll x_2 \ll x_3 \ll x_4$. In this case, a computation gives that $R(x_1, x_2, x_3, x_5)$ is conjugate to $y_1 = x_2^{-1/2} x_3 x_2^{-1/2}$, and $R(x_1, x_2, x_4, x_5)$ is conjugate to $y_2 = x_2^{-1/2} x_4 x_2^{-1/2}$. Since each eigenvalue of y_1 is positive and smaller than the corresponding eigenvalue of y_2, one obtains the desired inequality. □

Lemma 7.11 *Assume that (a, x, y, b) in $\mathcal{L}(V)^4$ is maximal. Then*

(1) $\|\det R(a, x, y, b)\| \geq 1$*;*

(2) $d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x), \mathrm{pr}_{\mathcal{Y}_{a,b}}(y)) \leq \ln \|\det R(a, x, y, b)\| \leq \sqrt{n}\, d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x), \mathrm{pr}_{\mathcal{Y}_{a,b}}(y))$.

Proof Since $\mathrm{Sp}(V)$ is transitive on maximal triples, we can assume that $a = 0, b = l_\infty$ and x corresponds to the matrix $+\mathrm{Id}$. Since the triple (x, y, l_∞) is maximal, y corresponds to a positive-definite matrix Y with all eigenvalues strictly bigger than one. The first statement is immediate since $\det R(a, x, y, b) = \det(Y)$.

It follows from the definition of the orthogonal projection that $\mathrm{pr}_{\mathcal{Y}_{a,b}}(x) = i\,\mathrm{Id}$ and $\mathrm{pr}_{\mathcal{Y}_{a,b}}(y) = iY$. If $\lambda_1, \dots, \lambda_n$ are the eigenvalues of Y, the explicit formula for the distance d gives

$$d(i\,\mathrm{Id}, iY) = \sqrt{\sum_{i=1}^{n} (\ln \|\lambda_i\|)^2},$$

and we have

$$\ln \|\det R(a, x, y, b)\| = \sum_{i=1}^{n} \ln \|\lambda_i\|.$$

The second assertion in the lemma then follows from Cauchy–Schwartz and the fact that $\ln \|\lambda_i\| \geq 0$ for every i. □

Lemma 7.12 *If $L_{\mathcal{B}}(\rho(\gamma)) = 0$, then for any $x, y \in ((\gamma^-, \gamma^+))$ with $(\gamma^-, x, y, \gamma^+)$ positively oriented, we have*

$$\|\det R(\phi(\gamma^-), \phi(x), \phi(y), \phi(\gamma^+))\| = 1.$$

Proof Since $(\gamma^-, x, y, \gamma^+)$ is positively oriented, and γ^+ is the attractive fixed point of γ, we can pick $n \geq 1$ with $(x, y, \gamma^n x)$ positively oriented. Then by Lemma 7.10, we have

$$\begin{aligned} 1 &\leq \det R(\phi(\gamma^-), \phi(x), \phi(y), \phi(\gamma^+)) \\ &\leq \det R(\phi(\gamma^-), \phi(x), \rho(\gamma)^n \phi(x), \phi(\gamma^+)), \end{aligned}$$

and the latter has norm 1 by Lemma 7.9(3). □

Proof of Proposition 7.1 Let s and t be points in $((\gamma^-, \gamma^+))$ and assume without loss of generality that the quadruple $(\gamma^-, t, s, \gamma^+)$ is positively oriented. Then $(\phi(\gamma^-), \phi(t), \phi(s), \phi(\gamma^+))$ is a maximal quadruple; thus by Lemma 7.11, we have $d(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(t)), \mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(s))) \leq \ln \|\det R(\phi(\gamma^-), \phi(t), \phi(s), \phi(\gamma^+))\|$. The right hand side vanishes by Lemma 7.12, and hence we obtain, using Proposition 6.3, that $\pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(t))) = \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(s)))$. □

Let us now assume that there are two elements γ, η in $\pi_1(\Sigma)$ whose axes intersect. We want to show that if both $\rho(\gamma)$ and $\rho(\eta)$ fix a point in $\mathcal{B}_V$, then they share a fixed point. We begin with a preliminary computation:

Lemma 7.13 *Let γ and η be two hyperbolic elements of Γ with intersecting axes. Assume $L_{\mathcal{B}}(\rho(\gamma)) = L_{\mathcal{B}}(\rho(\eta)) = 0$ and that the quadruple $(\eta^-, \gamma^-, \eta^+, \gamma^+)$ is positively oriented. Then for every $x \in ((\gamma^-, \gamma^+))$, all eigenvalues of the cross-ratio $R(\phi(\eta^-), \phi(\gamma^-), \phi(x), \phi(\gamma^+))$ have the form $1+f$, where $f \in \mathbb{F}^{>0}$ satisfies $\|f\| = 1$.*

Proof Pick $g \in \mathrm{Sp}(V)$ such that $g(\phi(\eta^-), \phi(\gamma^-), \phi(\gamma^+)) = (-\mathrm{Id}, 0, l_\infty)$, and set $p = g(\phi(\eta^+))$; see Figure 2. Now pick $x \in ((\gamma^-, \eta^+))$ and set $q = g(\phi(x))$. Observe that $0 \ll q \ll p$.

By Lemma 7.12, since $L_{\mathcal{B}}(\rho(\gamma)) = 0$, we have

$$\|\det R(\phi(\gamma^-), \phi(x), \phi(\eta^+), \phi(\gamma^+))\| = 1,$$

which implies $\|\det p\| = \|\det q\|$.

Let $\mu_1 \geq \cdots \geq \mu_n > 0$ and $\lambda_1 \geq \cdots \geq \lambda_n > 0$ denote the eigenvalues of q and p, respectively. Since $0 \ll q \ll p$, we deduce that $0 < \mu_i < \lambda_i$ and hence $\|\mu_i\| \leq \|\lambda_i\|$. This implies that $\|\mu_i\| = \|\lambda_i\|$ since we know that their products are equal.

Exploiting that $L_{\mathcal{B}}(\rho(\eta)) = 0$ together with Lemma 7.12 we get

$$\|\det R(\phi(\eta^-), \phi(\gamma^-), \phi(x), \phi(\eta^+))\| = 1,$$

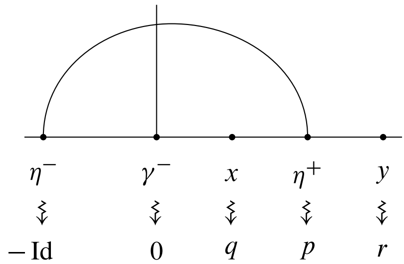


Figure 2: The points needed for the proof of Lemma 7.13

which implies that $\|(\det p)(\det(p-q))^{-1}\det(\mathrm{Id}+q)\| = 1$. From this, we deduce

$$\prod_{i=1}^{n}\|1+\mu_i\| = \|\det(\mathrm{Id}+q)\| = \left\|\frac{\det(p-q)}{\det p}\right\| \leq 1,$$

where the last inequality follows from $0 \ll p-q \ll p$. Together with the observation that $1+\mu_i \geq 1$ and the ultrametric inequality, this implies $\|\mu_i\| \leq 1$ for all i, and thus $\|\lambda_i\| = \|\mu_i\| \leq 1$.

Now let $y \in ((\eta^+,\gamma^+))$ and set $r = g(\phi(y))$. Then $0 \ll p \ll r$. Again by Lemma 7.12 we deduce that

$$\|\det R(\phi(\gamma^-),\phi(\eta^+),\phi(y),\phi(\gamma^+))\| = 1,$$

which implies $\|\det p\| = \|\det r\|$.

Let $\nu_1 \geq \cdots \geq \nu_n > 0$ denote the eigenvalues of r. Since $p \ll r$, we deduce that $0 < \lambda_i < \nu_i$, and hence $\|\nu_i\| \geq \|\lambda_i\|$. This implies, as above, that $\|\nu_i\| = \|\lambda_i\|$. Since $L_{\mathcal{B}}(\rho(\eta)) = 0$, Lemma 7.12 implies that

$$\|\det R(\phi(\eta^+),\phi(y),\phi(\gamma^+),\phi(\eta^-))\| = 1;$$

that is, $\|\det(\mathrm{Id}+r)\| = \|\det(r-p)\|$. Since $0 \ll r-p \ll r$, we obtain $\|\det(r-p)\| \leq \|\det r\|$. On the other hand, $0 \ll r \ll \mathrm{Id}+r$, and hence $\|\det(\mathrm{Id}+r)\| = \|\det(r)\|$, or equivalently, $\prod_{i=1}^{n}\|1+(1/\nu_i)\| = 1$. This, together with the information that $\nu_i > 0$ and the ultrametric inequality, implies $\|\nu_i\| \geq 1$, and thus $\|\lambda_i\| = \|\nu_i\| \geq 1$.

To conclude the proof, we observe that $R(\phi(\eta^-),\phi(\gamma^-),\phi(x),\phi(\gamma^+))$ is conjugate to $R(-\mathrm{Id},0,q,l_\infty) = \mathrm{Id}+q$ and hence has as all eigenvalues of the form $1+f$ with f positive satisfying $\|f\| = 1$. □

Remark 7.14 Recall from Definition 4.14 that $\mathcal{Y}_\gamma$ and $\mathcal{Y}_\eta$ are orthogonal if and only if $R(\phi(\eta^-),\phi(\gamma^-),\phi(\eta^+),\phi(\gamma^+)) = 2\,\mathrm{Id}$. Lemma 7.13 should be interpreted as a weaker form of orthogonality for the projections $\mathbb{Y}_\gamma$ and $\mathbb{Y}_\eta$.

Lemma 7.15 *Let $(a,c,b,d) \in \mathcal{L}(V)^4$ be a maximal quadruple, and assume that all the eigenvalues of $R(a,c,b,d)$ have the form $1+f$ for some $f \in \mathbb{F}^{>0}$ with $\|f\| = 1$. Then the points*

$$\mathrm{pr}_{\mathcal{Y}_{c,d}}(a), \quad \mathrm{pr}_{\mathcal{Y}_{c,d}}(b), \quad \mathrm{pr}_{\mathcal{Y}_{a,b}}(c), \quad \mathrm{pr}_{\mathcal{Y}_{a,b}}(d), \quad \mathcal{Y}_{a,b} \cap \mathcal{Y}_{c,d}$$

have pairwise pseudodistance zero.

Proof Pick $g \in \mathrm{Sp}(V)$ such that $g(a,c,b,d) = (-\mathrm{Id}, 0, D, l_\infty)$ where D is diagonal with strictly positive entries. Then a computation gives $\mathrm{pr}_{\mathcal{Y}_{0,l_\infty}}(-\mathrm{Id}) = i\,\mathrm{Id}$, $\mathrm{pr}_{\mathcal{Y}_{0,l_\infty}}(D) = iD$ and $\mathcal{Y}_{0,l_\infty} \cap \mathcal{Y}_{-\mathrm{Id},D} = i\sqrt{D}$.

Now since $D = \mathrm{diag}(d_1, \dots, d_n)$, the assumption on the eigenvalues implies $\|d_i\| = 1$, and the explicit formula for the distance gives the desired statement. □

Proof of Proposition 7.3 We may assume that $(\eta^-, \gamma^-, \eta^+, \gamma^+)$ is positively oriented. Applying Lemma 7.13 to $x = \eta^+$, we obtain that the pseudodistances of the points $\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^+))$, $\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^-))$, $\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^+))$, $\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^-))$, $\mathcal{Y}_\gamma \cap \mathcal{Y}_\eta$ are all zero. This concludes the proof once one notices that (see Proposition 7.1)

$$\begin{aligned} b_\gamma^+ &= \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^+))), \\ b_\gamma^- &= \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^-))), \\ b_\eta^+ &= \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^-))), \\ b_\eta^- &= \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^+))). \end{aligned}$$ □

8 Decomposition theorem

Let $\rho\colon \pi_1(\Sigma, x) \to \mathrm{Sp}(V)$ be a representation into a symplectic group over a real closed field $\mathbb{F}$ with valuation, and let $\pi_{\mathcal{B}}\colon \mathcal{T}_V \to \mathcal{B}_V$ denote the projection to the building. Recall from the introduction that if $\Sigma = \bigcup_{v \in \mathcal{V}} \Sigma_v$ is a decomposition of the surface Σ into subsurfaces with geodesic boundary, we consider the associated presentation of Γ as fundamental group of a graph of groups with vertex set $\mathcal{V}$ and vertex groups $\pi_1(\Sigma_v)$. We denote by $\widetilde{\mathcal{V}}$ the vertex set of the associated Bass–Serre tree $\mathcal{T}$. For every $v \in \mathcal{V}$ and $w \in \widetilde{\mathcal{V}}$ lying above v, the stabilizer Γ_w of w in Γ is isomorphic to $\pi_1(\Sigma_v)$. In this section, we prove the result mentioned in the introduction as Theorem 1.8:

Theorem 8.1 *Assume that $\rho\colon \Gamma \to \mathrm{Sp}(V)$ admits a maximal framing. Then there is a decomposition $\Sigma = \bigcup_{v \in \mathcal{V}} \Sigma_v$ of Σ into subsurfaces with geodesic boundary such that*

(1) *for every $\gamma \in \Gamma$ whose associated closed geodesic is not contained in any subsurface, $L_{\mathcal{B}}(\rho(\gamma)) > 0$;*

(2) *for every $v \in \mathcal{V}$, there is the following dichotomy:*

(PT) *for every $w \in \widetilde{\mathcal{V}}$ lying above v, and any $\gamma \in \Gamma_w$ which is not boundary parallel, $\rho(\gamma)$ has positive translation length;*

(FP) *for every $w \in \widetilde{\mathcal{V}}$ lying above v, there is a point $b_w \in \mathcal{B}_V$ which is fixed by Γ_w.*

The proof of the theorem is based on the analysis of the incidence structure of the set

$$\mathcal{L}_\rho = \{\gamma \in \Gamma \mid \gamma \neq e,\ \gamma \text{ hyperbolic},\ L_{\mathcal{B}}(\rho(\gamma)) = 0\}.$$

Let

$$\mathbb{P}\mathcal{L}_\rho = \{\gamma \in \mathcal{L}_\rho \mid \gamma \text{ is primitive}\}/_{\gamma \sim \gamma^{-1}},$$

and denote by $\overline{\gamma} \in \mathbb{P}\mathcal{L}_\rho$ the equivalence class of γ. Let

$$\mathcal{A}_\rho = \{\mathrm{ax}(\gamma) \mid \gamma \in \mathcal{L}_\rho\}$$

denote the set of axes of elements in $\mathcal{L}_\rho$, so there is a bijective correspondence $\mathcal{A}_\rho \cong \mathbb{P}\mathcal{L}_\rho$.

On $\mathbb{P}\mathcal{L}_\rho$ we put a graph structure by requiring that $\overline{\gamma}$ is adjacent to $\overline{\eta}$ if they are distinct and their axes intersect. We denote by $\mathcal{G}_\rho$ this graph and proceed to study its connected components. Let $\mathfrak{C} \subset \mathcal{G}_\rho$ be a connected component with vertex set $V(\mathfrak{C})$. We observe that if the component consists of a single vertex $\overline{\gamma}$, then the closed geodesic associated to γ is simple. Indeed, for each η in Γ, the conjugate $\eta\gamma\eta^{-1}$ belongs to $\mathcal{L}_\rho$ and if $\overline{\eta\gamma\eta^{-1}} \neq \overline{\gamma}$, the corresponding axes do not intersect.

Let us assume from now on that $|V(\mathfrak{C})| \geq 2$, and let

$$\Gamma_{\mathfrak{C}} = \{\gamma \in \Gamma \mid \gamma \text{ stabilizes } \mathfrak{C}\}$$

and

$$\Delta_{\mathfrak{C}} = \bigcup_{\overline{\gamma} \in V(\mathfrak{C})} \{\gamma^-, \gamma^+\}.$$

Then we clearly have that if $\overline{\gamma}$ belongs to $V(\mathfrak{C})$, then γ is an element of $\Gamma_{\mathfrak{C}}$ and $\Delta_{\mathfrak{C}}$ is a subset of the limit set $\Lambda(\Gamma_{\mathfrak{C}}) \subset \partial\mathbb{H}^2$ of $\Gamma_{\mathfrak{C}}$. In particular, since $\Delta_{\mathfrak{C}}$ is $\Gamma_{\mathfrak{C}}$–invariant, we get $\overline{\Delta}_{\mathfrak{C}} = \Lambda(\Gamma_{\mathfrak{C}})$.

Lemma 8.2 *There is a point $p_{\mathfrak{C}} \in \mathcal{B}_V$ with $b_\gamma^\pm = p_{\mathfrak{C}}$ for all γ such that $\overline{\gamma} \in V(\mathfrak{C})$.*

Proof Indeed, if $\overline{\gamma}$ is adjacent to $\overline{\eta}$, we have $b_\gamma^+ = b_\gamma^- = b_\eta^+ = b_\eta^-$; see Lemma 7.15. The lemma follows from the assumption that $\mathfrak{C}$ is connected. □

Lemma 8.3 *For every $\gamma \in \Gamma_{\mathfrak{C}}$, we have $\rho(\gamma) p_{\mathfrak{C}} = p_{\mathfrak{C}}$.*

Proof For every $\gamma \in \Gamma_{\mathfrak{C}}$, if η gives a vertex of $V(\mathfrak{C})$, the same holds for $\gamma\eta\gamma^{-1}$. Hence we get

$$b_\eta^\pm = b_{\gamma\eta\gamma^{-1}}^\pm = \rho(\gamma) b_\eta^\pm. \qquad \square$$

Lemma 8.4 *Let g be an oriented geodesic with endpoints g^- and g^+. Assume that $\Delta_{\mathfrak{C}} \cap ((g^-, g^+)) \neq \varnothing$ and $\Delta_{\mathfrak{C}} \cap ((g^+, g^-)) \neq \varnothing$. Then there exists $\overline{\gamma} \in \mathfrak{C}$ with $\mathrm{ax}(\gamma) \cap g \neq \varnothing$.*

Proof Let us choose a class $\overline{\eta} \in \mathfrak{C}$ with $\eta^+ \in ((g^-, g^+))$ and a class $\overline{\tau} \in \mathfrak{C}$ with $\tau^- \in ((g^+, g^-))$. Since $\mathfrak{C}$ is connected, there is a sequence $\overline{\alpha}_1 = \overline{\eta}, \overline{\alpha}_2, \dots, \overline{\alpha}_n = \overline{\tau}$ of classes in $\mathfrak{C}$ such that, for every i, the axis $\mathrm{ax}(\alpha_i)$ intersects $\mathrm{ax}(\alpha_{i+1})$. But then clearly there is an index j such that $\mathrm{ax}(\alpha_j)$ intersects the geodesic g. $\square$

If X is a subset of $\overline{\mathbb{H}}^2 = \mathbb{H}^2 \cup \partial\mathbb{H}^2$, we denote by $\overline{\mathrm{Co}(X)}$ the closed convex hull of X in $\mathbb{H}^2$. To any component $\mathfrak{C}$ we associate the closed convex subset $Y_{\mathfrak{C}}$ of $\mathbb{H}^2$ defined by

$$Y_{\mathfrak{C}} = \overline{\mathrm{Co}(\Lambda(\Gamma_{\mathfrak{C}}))} = \overline{\mathrm{Co}(\Delta_{\mathfrak{C}})}.$$

We say that an element $\gamma \in \Gamma_{\mathfrak{C}}$ is a boundary component if the axis of γ is a boundary component of $Y_{\mathfrak{C}}$.

Proposition 8.5 *For every primitive, hyperbolic element $\gamma \in \Gamma_{\mathfrak{C}}$ which is not a boundary component, we have*

$$\overline{\gamma} \in V(\mathfrak{C}).$$

Proof Since γ stabilizes $\mathfrak{C}$ and is not a boundary component, we have that the intersection $\Delta_{\mathfrak{C}} \cap ((\gamma^-, \gamma^+))$ is not empty, and similarly, $\Delta_{\mathfrak{C}} \cap ((\gamma^+, \gamma^-))$ is not empty. Thus we conclude by Lemma 8.4. $\square$

Our next aim is to show that the image $p(Y_{\mathfrak{C}})$ of $Y_{\mathfrak{C}}$ under the universal covering map $p\colon \mathbb{H}^2 \to \Sigma$ is a compact subsurface of Σ with geodesic boundary.

Proposition 8.6 *Let $\mathfrak{C} \subset \mathcal{G}_\rho$ be a connected component with more than one vertex. For every $\gamma \in \Gamma$, one of the following holds:*

(1) $\gamma Y_{\mathfrak{C}} = Y_{\mathfrak{C}}$;

(2) $\gamma Y_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$ *is a boundary component of* $Y_{\mathfrak{C}}$;

(3) *the intersection* $\gamma Y_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$ *is empty.*

Proof First we show that if the intersection $\gamma \mathring{Y}_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$ is not empty, then $\gamma\mathfrak{C} = \mathfrak{C}$, and hence $\gamma Y_{\mathfrak{C}} = Y_{\mathfrak{C}}$. Let $x \in \gamma \mathring{Y}_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$, and assume by contradiction that $\gamma\mathfrak{C} \neq \mathfrak{C}$, which implies that $\gamma V(\mathfrak{C}) \cap V(\mathfrak{C}) = \varnothing$.

Claim 1 *The point x does not belong to* $\mathrm{ax}(\eta)$ *for any* $\overline{\eta} \in \mathfrak{C}$.

Proof Assume, instead, that x belongs to $\mathrm{ax}(\eta)$ for some element η with $\overline{\eta} \in \mathfrak{C}$. If the intersection $\Delta_{\gamma\mathfrak{C}} \cap ((\eta^-, \eta^+))$ is empty, then $\Delta_{\gamma\mathfrak{C}}$ is contained in the closed interval $[\![\eta^+, \eta^-]\!]$, and hence $Y_{\gamma\mathfrak{C}}$ is contained in one of the closed halfplanes determined by $\mathrm{ax}(\eta)$. This contradicts the hypothesis that x belongs to the interior of $Y_{\gamma\mathfrak{C}}$. Thus we have that both intersections $\Delta_{\gamma\mathfrak{C}} \cap ((\eta^-, \eta^+))$ and $\Delta_{\gamma\mathfrak{C}} \cap ((\eta^+, \eta^-))$ are not empty. But then, by Lemma 8.4, there is an element $\xi \in \gamma\mathfrak{C}$ whose axis $\mathrm{ax}(\xi)$ intersects $\mathrm{ax}(\eta)$. This implies that either $\overline{\xi} = \overline{\eta}$, or the elements $\overline{\xi}$ and $\overline{\eta}$ are adjacent in the graph $\mathcal{G}_\rho$. Both contradict the fact that $\gamma V(\mathfrak{C}) \cap V(\mathfrak{C}) = \varnothing$, and this proves Claim 1. □

Now we can define $B_{\overline{g}}$, for every $\overline{g} \in \mathfrak{C}$, to be the unique closed interval in S^1 with endpoints $\{g^-, g^+\}$ and such that x does not belong to the convex hull $\overline{\mathrm{Co}(B_g)}$. According to Claim 1, this is well defined.

Claim 2 *For every $\overline{g}$ in $\mathfrak{C}$, the intersection $\Delta_{\gamma\mathfrak{C}} \cap B_{\overline{g}}$ is empty.*

Proof Indeed, assume that the intersection is not empty for some $\overline{g} \in \mathfrak{C}$. Since $\overline{g}$ does not belong to $\gamma\mathfrak{C}$, this implies that the intersection $\Delta_{\gamma\mathfrak{C}} \cap \mathring{B}_g$ is not empty. Since x belongs to $\gamma\mathring{Y}_{\mathfrak{C}}$, we get that the intersection $\Delta_{\gamma\mathfrak{C}} \cap (S^1 \setminus B_g)$ is not empty, and hence, by Lemma 8.4, there is $\overline{\xi} \in \gamma\mathfrak{C}$ whose axis $\mathrm{ax}(\xi)$ intersects $\mathrm{ax}(g)$ nontrivially. This again contradicts the assumption $\gamma V(\mathfrak{C}) \cap V(\mathfrak{C}) = \varnothing$. □

Claim 3 *The union $\bigcup_{\overline{g} \in \mathfrak{C}} B_{\overline{g}}$ is connected.*

Proof Indeed, for any pair of adjacent elements $\overline{\gamma}$ and $\overline{\eta}$ in $\mathfrak{C}$, we have that the intersection $B_{\overline{\gamma}} \cap B_{\overline{\eta}}$ is not empty. Now enumerate $\mathfrak{C}$ by a possibly redundant sequence $\overline{\gamma}_1, \overline{\gamma}_2, \ldots$ of consecutive adjacent vertices. Then the union $\bigcup_{i=1}^{\infty} B_{\overline{\gamma}_i}$ is connected. □

Since the union $\bigcup_{\overline{g} \in \mathfrak{C}} B_{\overline{g}}$ is connected, it is an interval of S^1 say with endpoints α_1, α_2, numbered such that

$$((\alpha_1, \alpha_2)) \subset \bigcup_{\overline{g} \in \mathfrak{C}} B_{\overline{g}} \subset [\![\alpha_1, \alpha_2]\!].$$

It follows then from Claim 2 that the intersection $\Delta_{\gamma\mathfrak{C}} \cap ((\alpha_1, \alpha_2))$ is empty; on the other hand, $\Delta_{\mathfrak{C}} \subseteq \bigcup_{\overline{g} \in \mathfrak{C}} B_{\overline{g}} \subset [\![\alpha_1, \alpha_2]\!]$. This implies that $Y_{\mathfrak{C}}$ and $Y_{\gamma\mathfrak{C}}$ lie in different half-planes determined by the geodesic joining α_1 to α_2 and hence the intersection $\gamma\mathring{Y}_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$ is empty. This gives a contradiction.

Assume now that $\gamma Y_{\mathfrak{C}}$ is different from $Y_{\mathfrak{C}}$ and that the intersection $\gamma Y_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$ is not empty. Let x be a point in the intersection $\gamma Y_{\mathfrak{C}} \cap Y_{\mathfrak{C}}$; then x belongs to the boundary of $\gamma Y_{\mathfrak{C}}$ and also to the boundary of $Y_{\mathfrak{C}}$. Let g and g' be the geodesics giving respectively the connected components of $\partial(\gamma Y_{\mathfrak{C}})$ and $\partial(Y_{\mathfrak{C}})$ containing x.

If $g \cap g' = \{x\}$, then the intersection of the interiors $\gamma \mathring{Y}_{\mathfrak{C}} \cap \mathring{Y}_{\mathfrak{C}}$ is not empty which, together with what we proved, implies that the $\gamma \mathring{Y}_{\mathfrak{C}}$ is equal to $\mathring{Y}_{\mathfrak{C}}$ and leads to a contradiction. Thus $g = g' \subseteq \partial(\gamma Y_{\mathfrak{C}}) \cap \partial Y_{\mathfrak{C}}$. Since the intersection $\gamma \mathring{Y}_{\mathfrak{C}} \cap \mathring{Y}_{\mathfrak{C}}$ is empty, we deduce that $\gamma Y_{\mathfrak{C}}$ and $Y_{\mathfrak{C}}$ lie on different sides of g, and hence $(\gamma Y_{\mathfrak{C}}) \cap Y_{\mathfrak{C}} = g$. □

Proposition 8.7 *Let $\mathfrak{C} \subset \mathcal{G}_\rho$ be a component with more than one vertex. Let $\Gamma_{\mathfrak{C}}$ be the stabilizer of $\mathfrak{C}$ in Γ and $Y_{\mathfrak{C}} \subset \mathbb{H}^2$ the closed convex hull of the limit set of $\Gamma_{\mathfrak{C}}$. Then the map*

$$\Gamma_{\mathfrak{C}} \backslash Y_{\mathfrak{C}} \hookrightarrow \Gamma \backslash \mathbb{H}^2$$

induces an embedding with image a compact surface with geodesic boundary.

Proof Let us enumerate the vertices $\{\overline{\gamma}_1, \overline{\gamma}_2, \dots\}$ of $V(\mathfrak{C})$ in such a way that for each i, we have $\overline{\gamma}_i$ adjacent to $\overline{\gamma}_{i+1}$. Let $\widetilde{x}_0$ be the intersection $\operatorname{ax}(\gamma_1) \cap \operatorname{ax}(\gamma_2)$, and define $X_n = \bigcup_{i=1}^{n} \operatorname{ax}(\gamma_i)$. By construction, X_n is connected. Let furthermore x_0 denote the projection $x_0 = p(\widetilde{x}_0)$.

Let $\Gamma_n < \Gamma$ be the image of the natural map $\pi_1(p(X_n), x_0) \to \pi_1(\Sigma, x_0)$ induced by the inclusion $p(X_n) \hookrightarrow \Sigma$. Then Γ_n is the fundamental group of the surface $\Sigma_n \subseteq \Sigma$ obtained by taking an appropriate tubular neighborhood of $p(X_n) \subseteq \Sigma$ and adding to it all components of the complement which are either simply connected or whose fundamental group is generated by a parabolic element of Γ. Then Σ_n is a subsurface with smooth boundary and of finite topological type. Since $\Gamma_n < \Gamma_{n+1}$, there exists $N \geq 1$ with $\Gamma_n = \Gamma_N$ for all $n \geq N$.

We will finish the proof by showing that $\Gamma_{\mathfrak{C}} = \Gamma_N$. Since $\Gamma_n \widetilde{x}_0 \subset \widetilde{p(X_n)}$, we have $\Gamma_n < \Gamma_{\mathfrak{C}}$. Conversely, let us take $\gamma \in \Gamma_{\mathfrak{C}}$; then $\gamma \widetilde{x}_0 = \operatorname{ax}(\gamma\gamma_1\gamma^{-1}) \cap \operatorname{ax}(\gamma\gamma_2\gamma^{-1})$, and since $\Gamma_{\mathfrak{C}}$ preserves $V(\mathfrak{C})$, we have that $\overline{\gamma\gamma_1\gamma^{-1}}$ and $\overline{\gamma\gamma_2\gamma^{-1}}$ are in $V(\mathfrak{C})$. Thus $\gamma \widetilde{x}_0 \in X_n$ for n large enough, which implies $\gamma \in \Gamma_n$. As a conclusion, we get $\Gamma_N = \Gamma_{\mathfrak{C}}$, which implies that $\Gamma_{\mathfrak{C}} \backslash Y_{\mathfrak{C}}$ in Σ is isotopic to Σ_N. □

Proof of Theorem 8.1 The set of isolated components of $\mathcal{G}_\rho$ is a Γ–invariant subset. Since we know that each isolated component of $\mathcal{G}_\rho$ corresponds to a geodesic of $\mathbb{H}^2$ that projects to a simple closed curve, we have that the projection of all the isolated components is a collection $\mathcal{C}$ of pairwise disjoint simple closed curves which cut the surface Σ in subsurfaces $\{\Sigma_v\}_{v \in \mathcal{V}}$ for some index set $\mathcal{V}$.

Moreover, for any component $\mathfrak{C}$ consisting of more than one element, we have that

$$Y_{\mathfrak{C}} = \overline{\mathrm{Co}\Big(\bigcup_{\overline{\gamma}\in\mathfrak{C}} \mathrm{ax}(\gamma)\Big)}$$

is a subsurface in $\mathbb{H}^2$ which projects to a subsurface of Σ whose boundary consists of elements of $\mathcal{C}$. In particular, there exists $v \in V$ with $p(Y_{\mathfrak{C}}) = \Sigma_v$. □

9 Quasi-isometric embeddings

Let $\rho\colon \pi_1(\Sigma, x) \to \mathrm{Sp}(V)$ be a representation admitting a maximal framing and $\Sigma = \bigcup_{v\in\mathcal{V}} \Sigma_v$ be the corresponding decomposition given by Theorem 8.1. We assume, as usual, that Σ is equipped with a hyperbolic metric of finite area and denote by $p\colon \mathbb{H}^2 \to \Sigma$ the canonical projection, so $\Sigma = \Gamma\backslash\mathbb{H}^2$.

As we have seen in Section 8, the decomposition of the surface Σ comes from a Γ–invariant decomposition

$$\mathbb{H}^2 = \bigcup_{w\in\widetilde{\mathcal{V}}} S_w$$

into subsurfaces with totally geodesic boundary. The Bass–Serre tree $\mathcal{T} = (\widetilde{\mathcal{V}}, E)$ can be identified with the incidence tree of the set $\{S_w \mid w \in \widetilde{\mathcal{V}}\}$. Recall that a pair $\{w_1, w_2\}$ forms an edge if the intersection $S_{w_1} \cap S_{w_2}$ is not empty. In this case, the intersection corresponds to the axis of an element of Γ that acts on the building $\mathcal{B}_V$ with zero translation length and determines an isolated component of the graph $\mathcal{G}_\rho$.

Assume now that for every subsurface Σ_v we are in the second case of the dichotomy in the decomposition theorem. Then for every $w \in \widetilde{\mathcal{V}}$, the stabilizer Γ_w of w in Γ has a canonical fixed point $b_w \in \mathcal{B}_V^0$ which equals $b_\gamma^\pm$ for each $\gamma \in \Gamma_w$.

Theorem 9.1 *The map*

$$\widetilde{\mathcal{V}} \to \mathcal{B}_V^0, \quad w \mapsto b_w,$$

is a Γ–equivariant quasi-isometry.

Let $\eta \in \Gamma$ be an element whose corresponding geodesic is not contained in a subsurface. The axis $\mathrm{ax}(\eta)$ determines a sequence $(w_n)_{n\in\mathbb{Z}}$ of vertices in $\mathcal{T}$, namely the consecutive sequence of surfaces S_{w_n} crossed by $\mathrm{ax}(\eta)$. This gives a geodesic path in $\mathcal{T}$, which is the axis of the isometry of $\mathcal{T}$ induced by η.

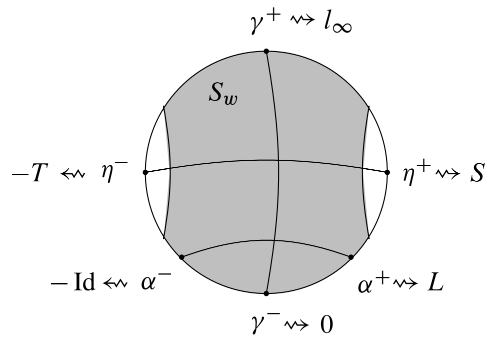


Figure 3: The setting in the proof of Lemma 9.2

Lemma 9.2 *Let us assume that the axis* $\mathrm{ax}(\eta)$ *crosses the surface* S_w. *Let* $\gamma \in \Gamma_w$ *be an element which is not boundary parallel and such that* $\mathrm{ax}(\gamma)$ *intersects* $\mathrm{ax}(\eta)$. *Then*

$$b_w = \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^+))) = \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^-))).$$

In particular, b_w *belongs to* $\mathbb{Y}_\eta$.

Proof Without loss of generality, we assume that the 4–tuple $(\eta^-, \gamma^-, \eta^+, \gamma^+)$ is positively oriented. We will show that all of the eigenvalues of the cross-ratio $R(\phi(\eta^-), \phi(\gamma^-), \phi(\eta^+), \phi(\gamma^+))$ have the form $1+f$ for a positive f satisfying $\|f\| = 1$.

Since γ is not boundary parallel, we can find $\alpha \in \Gamma_w$ such that α^- belongs to $((\eta^-, \gamma^-))$ and α^+ belongs to $((\gamma^-, \eta^+))$; see Figure 3. Since $(\alpha^-, \gamma^-, \alpha^+, \gamma^+)$ is positively oriented, we can pick an element $g \in \mathrm{Sp}(V)$ with $g(\alpha^-, \gamma^-, \gamma^+) = (-\mathrm{Id}, 0, l_\infty)$. For such g, we set $g\phi(\eta^-) = -T$, $g\phi(\eta^+) = S$ and $g\phi(\alpha^+) = L$. With this notation we have $T \gg \mathrm{Id}$ and $S \gg L \gg 0$; moreover, the cross-ratio $R(\phi(\eta^-), \phi(\gamma^-), \phi(\eta^+), \phi(\gamma^+))$ is conjugate to $R(-T, 0, S, l_\infty) = \mathrm{Id} + T^{-1}S$.

First observe that all the eigenvalues of $R(-T, 0, S, l_\infty)$ are smaller than the corresponding eigenvalues of $R(-\mathrm{Id}, 0, S, l_\infty)$. Indeed, the first matrix is conjugate to $\mathrm{Id} + S^{1/2}T^{-1}S^{1/2}$ and the second equals $\mathrm{Id} + S$, moreover all the eigenvalues of T are by assumption greater than 1. Now α and γ cross and have zero translation length since they both belong to Γ_w. Since η^+ belongs to $((\alpha^+, \gamma^+))$, it follows from Lemma 7.13 that all the eigenvalues of $R(\phi(\alpha^-), \phi(\gamma^-), \phi(\eta^+), \phi(\gamma^+))$ have the form $1+\lambda$ for a positive λ satisfying $\|\lambda\| = 1$. This implies that for each eigenvalue ν_i of $\mathrm{Id} + T^{-1}S$, we have $\|\nu_i - 1\| \leq 1$.

On the other hand, all the eigenvalues of $R(-T, 0, S, l_\infty)$ are bigger than the corresponding eigenvalues of $R(-T, 0, L, l_\infty)$: indeed the first matrix is conjugate to $\mathrm{Id} + T^{-1/2} S T^{-1/2}$ and the second is conjugate to $\mathrm{Id} + T^{-1/2} L T^{-1/2}$. This implies that, denoting by μ_i the eigenvalues of $R(-T, 0, L, l_\infty)$ we have that $\|\nu_i - 1\| \geq \|\mu_i - 1\|$. This is enough to conclude: we have by Lemma 4.4 that $R(-T, 0, L, l_\infty) \cong R(L, l_\infty, -T, 0)$, and as a consequence of Lemma 7.13, this latter cross-ratio has all its eigenvalues of the form $1 + f$ for some positive f of norm one.

Now we exploit that b_w is in particular equal to $b_\gamma^\pm$. This latter point is, in view of Proposition 7.1, equal to $\pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^-)))$. Moreover, we deduce from Lemma 7.15 that

$$\pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\gamma}(\phi(\eta^-))) = \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^-))) = \pi_{\mathcal{B}}(\mathrm{pr}_{\mathcal{Y}_\eta}(\phi(\gamma^+))),$$

and this concludes. □

Lemma 9.3 *Let $a, b \in \mathcal{L}(V)$ be transverse subspaces, and fix $x_1, \dots, x_k \in ((a, b))$ such that (a, x_i, x_{i+1}, b) is maximal for all i. Then*

$$\sum_{i=1}^{k-1} d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x_i), \mathrm{pr}_{\mathcal{Y}_{a,b}}(x_{i+1})) \leq \sqrt{n}\, d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x_1), \mathrm{pr}_{\mathcal{Y}_{a,b}}(x_k)).$$

Proof Since for each pair of symmetric matrices S, T we have $\det R(0, S, T, l_\infty) = \det S^{-1} \det T$, we deduce

$$\det R(a, x_1, x_k, b) = \prod_{j=1}^{k-1} \det R(a, x_j, x_{j+1}, b).$$

Thus we get

$$\ln \|\det R(a, x_1, x_k, b)\| = \sum_{j=1}^{k-1} \ln \|\det R(a, x_j, x_{j+1}, b)\|.$$

From Lemma 7.11, we deduce immediately

$$\sum_{i=1}^{k-1} d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x_i), \mathrm{pr}_{\mathcal{Y}_{a,b}}(x_{i+1})) \leq \sqrt{n}\, d(\mathrm{pr}_{\mathcal{Y}_{a,b}}(x_1), \mathrm{pr}_{\mathcal{Y}_{a,b}}(x_k)).$$ □

Proof of Theorem 9.1 Let v, w be vertices of $\mathcal{T}$, and pick an element $\eta \in \Gamma$ whose associated axis in $\mathcal{T}$ contains the geodesic path between v and w. Let us name $v_0 = v, v_1, \dots, v_k = w$ the vertices in such path.

We choose, for every i an element $\gamma_i \in S_{v_i}$ whose axis $\mathrm{ax}(\gamma_i)$ intersects the axis $\mathrm{ax}(\eta)$ nontrivially, and with the property that $\gamma_i^+ \in ((\eta^-, \eta^+))$. Then we have that, for every i, the 4–tuple

$$(\phi(\eta^-), \phi(\gamma_i^+), \phi(\gamma_{i+1}^+), \phi(\eta^+))$$

is maximal, and hence by Lemma 9.2 and 9.3, we have

$$\sum_{i=0}^{k-1} d_{\mathcal{B}}(b_{v_i}, b_{v_{i+1}}) \leq \sqrt{n}\, d_{\mathcal{B}}(b_{v_0}, b_{v_k}).$$

Notice that for any pair of adjacent vertices l, r in $\mathcal{T}$, the distance $d_{\mathcal{B}}(b_l, b_r)$ is positive: otherwise it is easy to verify that for each pair of hyperbolic elements $\gamma_l \in \Gamma_l$ and $\gamma_r \in \Gamma_r$, the composition $\gamma_l \gamma_r$ fixes $b_l = b_r$ and corresponds to an element of Γ whose axis crosses the common boundary component of S_l and S_r, contradicting the decomposition of Theorem 8.1.

Now, since the number of Γ–orbits on the set of edges of $\mathcal{T}$ is finite, there are positive constants C_1, C_2 with

$$C_1 \leq d_{\mathcal{B}}(b_l, b_r) \leq C_2$$

for every pair (l, r) of adjacent vertices. Thus we get

$$k C_1 \leq \sqrt{n}\, d_{\mathcal{B}}(b_{v_0}, b_{v_k}),$$

which implies

$$d_{\mathcal{T}}(v_0, v_k) \leq \frac{\sqrt{n}}{C_1} d_{\mathcal{B}}(b_{v_0}, b_{v_k}).$$

The inequality

$$d_{\mathcal{B}}(b_{v_0}, b_{v_k}) \leq C_2 k = C_2 d_{\mathcal{T}}(v_0, v_k)$$

is immediate. □

10 Ultralimits of maximal representations

In this section, we apply the general theory developed so far to the field of hyperreals and the Robinson field in order to deduce the decomposition theorem for ultralimits of maximal representations.

10.1 Hyperreals and Robinson fields

Let $\omega\colon \mathcal{P}(\mathbb{N}) \to \{0,1\}$ be a nonprincipal ultrafilter on the set of natural numbers. Recall that the ultraproduct $\prod_\omega X_i$ of a sequence $(X_i)_{i\in\mathbb{N}}$, of sets is the quotient of $\prod_{i\in\mathbb{N}} X_i$ by the equivalence relation $(x_i) \sim (y_i)$ if $\omega(\{i \mid x_i = y_i\}) = 1$. We denote by $\lambda_\omega\colon \prod_{i\in\mathbb{N}} X_i \to \prod_\omega X_i$ the quotient map and write X_ω for $\prod_\omega X$. In particular, $\mathbb{R}_\omega$ is the field of hyperreals, and if X_i are vector spaces over $\mathbb{R}$ (resp. $\mathbb{R}$–algebras, groups), then $\prod_\omega X$ is a $\mathbb{R}_\omega$–vector space (resp. an $\mathbb{R}_\omega$–algebra, a group) and λ_ω is a morphism in the appropriate category. For a $\mathbb{R}$–vector space V, the map

$$V \times \mathbb{R}_\omega \to V_\omega, \quad (v, [(l_i)]) \mapsto [(l_i v)],$$

induces an $\mathbb{R}_\omega$–isomorphism $V \otimes_{\mathbb{R}} \mathbb{R}_\omega \to V_\omega$. For V finite-dimensional at least, we deduce from the isomorphism $\mathrm{End}_{\mathbb{R}_\omega}(V \otimes_{\mathbb{R}} \mathbb{R}_\omega) \cong (\mathrm{End}\, V) \otimes_{\mathbb{R}} \mathbb{R}_\omega$ that the map

$$\prod_{i \in \mathbb{N}} \mathrm{End}(V) \to \mathrm{End}(V_\omega), \quad (T_i)_i \mapsto T,$$

where $T([v_i]) = [T_i(v_i)]$ induces an algebra isomorphism $(\mathrm{End}(V))_\omega \cong \mathrm{End}(V_\omega)$ which restricts to a group isomorphism $(\mathrm{GL}(V))_\omega \cong \mathrm{GL}(V_\omega)$. By abuse of notation, we will also denote by $\lambda_\omega\colon \prod_{\mathbb{N}} \mathrm{GL}(V) \to \mathrm{GL}(V_\omega)$ the induced map. Given a symplectic form $\langle\cdot,\cdot\rangle$ on V, let $\langle\cdot,\cdot\rangle_\omega$ denote the symplectic form on V_ω obtained by extending the scalars from $\mathbb{R}$ to $\mathbb{R}_\omega$. Given a sequence of representations $\rho_i\colon \Gamma \to \mathrm{Sp}(V)$, we will denote by ρ_ω the representation of Γ into $\mathrm{Sp}(V_\omega)$ obtained by composing $\prod_{i\in\mathbb{N}} \rho_i$ with λ_ω.

Proposition 10.1 *Assume that $\rho_i\colon \Gamma \to \mathrm{Sp}(V)$ is a sequence of maximal representations. Then $\rho_\omega\colon \Gamma \to \mathrm{Sp}(V_\omega)$ admits a maximal framing.*

The proof uses the following lemma, which is a straightforward verification:

Lemma 10.2 (1) *The map $\prod_{\mathbb{N}} \mathrm{Gr}_k(V) \to \mathrm{Gr}_k(V_\omega)$ defined by $(L_i)_{i\in\mathbb{N}} \mapsto \prod_\omega L_i$ induces a $(\mathrm{GL}(V))_\omega \cong \mathrm{GL}(V_\omega)$–equivariant bijection $(\mathrm{Gr}_k(V))_\omega \cong \mathrm{Gr}_k(V_\omega)$ and restricts to a $(\mathrm{Sp}(V))_\omega \cong \mathrm{Sp}(V_\omega)$–equivariant bijection $\mathcal{L}(V)_\omega \cong \mathcal{L}(V_\omega)$.*

(2) *Let $f_i\colon W_i \to \mathbb{R}$ be quadratic forms with signature $n_i \in \mathbb{Z}$. Assume that the sequence $\dim W_i$ is bounded, and let $f_\omega\colon \prod_\omega W_i \to \mathbb{R}_\omega$ be the quadratic form given by $f_\omega([(v_i)]) = [(f_i(v_i))]$. Then f_ω has signature n where n is defined by $\omega(\{i \mid n_i = n\}) = 1$.*

Proof of Proposition 10.1 Since each ρ_i is maximal, there exists a maximal framing $\phi_i\colon \partial\mathbb{H}^2 \to \mathcal{L}(V)$. Define then $\phi_\omega\colon \partial\mathbb{H}^2 \to \mathcal{L}(V_\omega)$ by composing $\prod \phi_i\colon \partial\mathbb{H}^2 \to \prod_{\mathbb{N}} \mathcal{L}(V)$ with the quotient map $\prod_{\mathbb{N}} \mathcal{L}(V) \to \mathcal{L}(V_\omega)$. The maximality of the so obtained framing follows then from Lemma 10.2(2). □

Let now $\sigma \in \mathbb{R}_\omega$ be an infinitesimal and recall the definition of the local ring

$$\mathcal{O}_\sigma = \{x \in \mathbb{R}_\omega : |x| < \sigma^{-k} \text{ for some } k \in \mathbb{N}\}$$

with associated maximal ideal

$$\mathcal{I}_\sigma = \{x \in \mathbb{R}_\omega : |x| < \sigma^{k} \text{ for all } k \in \mathbb{N}\}.$$

The quotient is the Robinson field $\mathbb{R}_{\omega,\sigma} = \mathcal{O}_\sigma / \mathcal{I}_\sigma$ associated to σ [24; 16].

Remark 10.3 Assuming the continuum hypothesis, a deep result of Erdös, Gillman and Henriksen [8] implies that the field $\mathbb{R}_\omega$ does not depend on the choice of the ultrafilter. And under the same hypothesis, Thornton showed that the normed field $\mathbb{R}_{\omega,\sigma}$ does not depend on the choice of the ultrafilter ω nor on the infinitesimal σ [28, Theorem 2.34].

If instead we assume the negation of the continuum hypothesis, it was shown by Kramer, Shelah, Tent and Thomas [12, Theorem 1.8] that there exists an uncountable set of nonprincipal ultrafilters such that the associated Robinson fields are pairwise nonisomorphic.

If (λ_i) is a divergent sequence of real numbers and we set $\sigma = [(e^{-\lambda_i})] \in \mathbb{R}_\omega$ we have that the field $\mathbb{R}_{\omega,\sigma}$ is the field denoted by $\mathbb{R}_{\omega,\lambda}$ in [22].

Now let ρ_ω be a representation into $\mathrm{Sp}(V_\omega)$ admitting the maximal framing (S, ϕ_ω). Choose a compatible complex structure J_ω and an infinitesimal $\sigma \in \mathbb{R}_\omega$ such that $\rho_\omega(\Gamma) \subseteq \mathrm{Sp}(V_\omega)(\mathcal{O}_\sigma)$, and denote by $V_{\omega,\sigma}$ the vector space $V_\omega(\mathcal{O}_\sigma)/V_\omega(\mathcal{I}_\sigma)$. According to Theorem 5.9, composing ρ_ω with $\pi_\sigma\colon \mathrm{Sp}(V_\omega)(\mathcal{O}_\sigma) \to \mathrm{Sp}(V_{\omega,\sigma})$ we obtain a representation which admits $q_\sigma \circ \phi_\omega\colon S \to \mathcal{L}(V_{\omega,\sigma})$ as maximal framing.

Thus we obtain in particular:

Corollary 10.4 *If $(\rho_i)_{i\in\mathbb{N}}\colon \Gamma \to \mathrm{Sp}(V)$ is a sequence of maximal representations where V is a real symplectic vector space, $\rho_\omega\colon \Gamma \to \mathrm{Sp}(V_\omega)$ the corresponding representation over the field of hyperreals, J_ω a choice of compatible complex structure and σ an infinitesimal such that $\rho_\omega(\Gamma) \subset \mathrm{Sp}(V_\omega)(\mathcal{O}_\sigma)$, then the representation $\rho_{\omega,\sigma}\colon \Gamma \to \mathrm{Sp}(V_{\omega,\sigma})$ admits a maximal framing defined on $\partial\mathbb{H}^2$.*

In the compact case we obtain a converse:

Theorem 10.5 *Assume that the surface $\Gamma\backslash\mathbb{H}^2$ is compact. Then a representation $\rho\colon \Gamma \to \mathrm{Sp}(V_{\omega,\sigma})$ admits a maximal framing if and only if there is a sequence $\rho_i\colon \Gamma \to \mathrm{Sp}(V)$ of maximal representations such that $\rho_{\omega,\sigma} = \rho$.*

Proof Let

$$\mathrm{Rep}_g := \left\{ (A_1, B_1, \dots, A_g, B_g) \in \mathrm{Sp}(V)^{2g} \;\middle|\; \prod_{i=1}^{n} [A_i, B_i] = \mathrm{Id} \right\}$$

be the $\mathbb{R}$–variety of representations of Γ in $\mathrm{Sp}(V)$. Then it follows from [28] that the reduction modulo $\mathcal{I}_\sigma$ induces a surjection $\mathrm{Rep}_g(\mathcal{O}_\sigma) \to \mathrm{Rep}_g(\mathbb{R}_{\omega,\sigma})$. Thus we can lift ρ to a representation $\rho_\omega\colon \Gamma \to \mathrm{Sp}(V_\omega)(\mathcal{O}_\sigma)$ which we represent by a sequence $(\rho_i)_{i\in\mathbb{N}}$ of representations of Γ into $\mathrm{Sp}(V)$.

Let $\phi\colon S \to \mathcal{L}(V_{\omega,\sigma})$ be a maximal framing for ρ. It follows from the collar lemma that for every hyperbolic element $\gamma \in \Gamma$, the image $\rho(\gamma)$ is Shilov hyperbolic. Then $\rho_\omega(\gamma)$ needs also to be Shilov hyperbolic and we have $q_\sigma(L^+_{\rho_\omega(\gamma)}) = L^+_{\rho(\gamma)}$ because of uniqueness of attractive fixed Lagrangians.

Fix a decomposition of $\Sigma = \Gamma\backslash\mathbb{H}^2$ into pairs of pants, let $P \subseteq \Sigma$ denote any such pair of pants and let $\{c_1, c_2, c_3\}$ be standard generators of $\pi_1(P)$; in particular, $c_1c_2c_3 = e$. Let ξ_1, ξ_2, ξ_3 be the attractive fixed points in $\partial\mathbb{H}^2$ of c_1, c_2, c_3. Then (ξ_1, ξ_2, ξ_3) and $(\xi_1, c_1 \cdot \xi_3, \xi_2)$ are positively oriented. Thus the images under ϕ of the two triples are maximal, and hence the triples

$$(L^+_{\rho_\omega(c_1)}, L^+_{\rho_\omega(c_2)}, L^+_{\rho_\omega(c_3)}) \quad \text{and} \quad (L^+_{\rho_\omega(c_1)}, \rho_\omega(c_1)L^+_{\rho_\omega(c_3)}, L^+_{\rho_\omega(c_2)})$$

are maximal. It follows that there is a set $E_P \subset \mathbb{N}$ of full ω–measure such that for each i in E_P, $\rho_i(c_1), \rho_i(c_2), \rho_i(c_3)$ are Shilov hyperbolic and both

$$(L^+_{\rho_i(c_1)}, L^+_{\rho_i(c_2)}, L^+_{\rho_i(c_3)}) \quad \text{and} \quad (L^+_{\rho_i(c_1)}, \rho_i(c_1)L^+_{\rho_i(c_3)}, L^+_{\rho_i(c_2)})$$

are maximal. It follows then from [27, Theorem 5] that $\rho_i|_{\pi_1(P)} \to \mathrm{Sp}(V)$ is maximal for each i in E_P. Thus if $P_1, \dots, P_{2g-2}$ is the pair of pants decomposition, we have that for all $i \in \bigcap_{j=1}^{2g-2} E_{P_j}$, the restriction $\rho_i|_{\pi_1(P_j)}$ is maximal. By additivity of the Toledo invariant (see [6, Theorem 1]), we deduce that ρ_i is maximal. Since $\bigcap_{j=1}^{2g-2} E_{P_j}$ is of full ω–measure, this concludes the proof. □

10.2 Asymptotic cones

We finish the paper deducing the statements about ultralimits of maximal representations from the general theory of representations admitting a maximal framing.

Proof of Theorem 1.1 Let $\rho_k\colon \Gamma \to \mathrm{Sp}(V)$ be a sequence of maximal representations, $J_k \in \mathbb{X}_V$ a sequence of basepoints, namely a sequence of compatible complex structures, and $\lambda = (\lambda_k)_{k\in\mathbb{N}}$ an adapted sequence of scales. If the sequence λ is bounded on a set of full ω–measure, then we may assume

$$\sup_{k\in\mathbb{N}} \max_{\gamma\in S} d(\rho_k(\gamma)J_k, J_k) < \infty,$$

and hence, if we conjugate ρ_k by $g_k \in \mathrm{Sp}(V)$ with $g_k J_k = x$ a fixed basepoint, it follows that the sequence $(\pi_k = g_k\rho_k g_k^{-1})_{k\in\mathbb{N}}$ is relatively compact in the space of representations. In this case, ${}^\omega\mathcal{X}_\lambda$ is just the Siegel space $\mathcal{X}_\mathbb{R}$ with rescaled distance, and ${}^\omega\rho_\lambda$ is an ordinary accumulation point of the sequence $(\pi_k)_{k\in\mathbb{N}}$.

If the sequence λ is unbounded, let $\sigma := (e^{-\lambda_k})_{k\in\mathbb{N}}$, which is an infinitesimal in $\mathbb{R}_\omega$, and let $J_\omega := [(J_k)_{k\in\mathbb{N}}] \in \mathrm{End}(V_\omega)$ which is a compatible complex structure. Then we conclude from the fact that λ is adapted to $(\rho_k, J_k)_{k\in\mathbb{N}}$ that $\rho_\omega(\Gamma) \subset \mathrm{Sp}(V_\omega)(\mathcal{O}_\sigma)$.

Furthermore, it follows from [22] that the action on the Bruhat–Tits building of $\mathrm{Sp}(V_{\omega,\sigma})$ coming from the representation $\rho_{\omega,\sigma}\colon \Gamma \to \mathrm{Sp}(V_{\omega,\sigma})$ coincides with the ultralimit ${}^{\omega}\rho_{\lambda}\colon \Gamma \to \mathrm{Iso}({}^{\omega}\mathcal{X}_{\lambda})$ under the identification of ${}^{\omega}\mathcal{X}_{\lambda}$ with the Bruhat–Tits building $\mathcal{B}_{V_{\omega,\sigma}}$. Theorem 1.1 follows then from Corollary 10.4 and Theorem 8.1. $\square$

We now characterize the cases which lead to actions without a global fixed point. Recall from the introduction that when S is a finite generating set for Γ, and ρ is a maximal representation we denote by $D_S(\rho)(x)$ the displacement function.

The function $D_S(\rho)$ is convex and, since $\rho(\Gamma)$ is not contained in any proper parabolic subgroup of $\mathrm{Sp}(V)$, we have that for every $C>0$, the convex set $\{x \mid D_S(\rho)(x) \leq C\}$ must be compact; in particular, $D_S(\rho)(x)$ achieves its minimum that we will denote by $\mu_S(\rho) = \min_{x\in\mathcal{X}} D_S(\rho)(x)$.

The function $\rho \mapsto \mu_S(\rho)$ descends then to a proper function

$$\mathrm{Hom}_{\max}(\Gamma, \mathrm{Sp}(V))/\,\mathrm{Sp}(V) \to (0,\infty)$$

on the character variety of maximal representations. Let now $(\rho_k)_{k\in\mathbb{N}}$ be a sequence of maximal representations, $(x_k)_{k\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}}$ a sequence of basepoints and λ an adapted sequence of scales. Furthermore, let $y_k \in \mathcal{X}$ be such that $\mu_S(\rho_k) = D_S(\rho_k)(y_k)$.

Proposition 10.6 *The representation ${}^{\omega}\rho_{\lambda}$ on ${}^{\omega}\mathcal{X}_{\lambda}$ has no global fixed point if and only if*

$$\lim_{\omega} \frac{\lambda_k}{\mu_S(\rho_k)} < \infty \quad \text{and} \quad \lim_{\omega} \frac{d(y_k,x_k)}{\lambda_k} < \infty,$$

in which case ${}^{\omega}\mathcal{X}_{\lambda} = {}^{\omega}\mathcal{X}_{\mu}$, the distances on the asymptotic cones are homothetic and the actions ${}^{\omega}\rho_{\lambda}$ and ${}^{\omega}\rho_{\mu}$ coincide.

Remark 10.7 We can also deduce the fact that if ${}^{\omega}\rho_{\lambda}$ has no global fixed point then the limit $\lim_{\omega} \lambda_k/\mu_S(\rho_k)$ is finite by combining [22, Proposition 4.4] and [21, Corollary 3].

Proof of Proposition 10.6 For the "if" part: changing the sequence on a set of ω–measure zero, we may assume that for some constant $C>0$, we have $\mu_S(\rho_k)/C \leq \lambda_k \leq C\mu_S(\rho_k)$ and $d(y_k,x_k) \leq C\lambda_k$ for all $k\in\mathbb{N}$. This readily implies that the asymptotic cones ${}^{\omega}\mathcal{X}_{\lambda}$ and ${}^{\omega}\mathcal{X}_{\mu}$ are equal, that the induced distances are homothetic with factor $\lim_{\omega} \lambda_k/\mu_S(\rho_k)$ and that the actions ${}^{\omega}\rho_{\lambda}$ and ${}^{\omega}\rho_{\mu}$ coincide. Thus we have to verify that ${}^{\omega}\rho_{\mu}$ does not have a global fixed point. But this follows immediately from the fact that

$$\max_{\gamma\in S} \frac{d(\rho_k(\gamma)x,x)}{\mu_S(\rho_k)} \geq 1 \quad \text{for all } x\in\mathcal{X}.$$

We next show the "only if" part. Let T be a finite connected generating set, and let us denote by K the maximal length of an element of T with respect to the generating set S. Since ${}^{\omega}\rho_{\lambda}$ does not have a global fixed point, it follows from Corollary 7.6 that there is $\gamma_0 \in T$ with $L({}^{\omega}\rho_{\lambda}(\gamma_0)) = \lim_{\omega} L(\rho_k(\gamma_0))/\lambda_k > 0$. Since

$$L(\rho_k(\gamma_0)) \le d(\rho_k(\gamma_0)y_k, y_k) \le K\mu_S(\rho_k) \le KD_S(\rho_k)(x_k)$$

and $\lim_{\omega} D_S(\rho_k)(x_k)/\lambda_k < \infty$, we may assume that the sequences $(\lambda_k)_{k\in\mathbb{N}}$ and $(\mu_S(\rho_k))_{k\in\mathbb{N}}$ are equivalent, namely that there are positive constants C_1, C_2 such that $C_1\mu_S(\rho_k) \le \lambda_k \le C_2\mu_S(\rho_k)$ for all $k \in \mathbb{N}$.

Pick now two hyperbolic elements γ, η in Γ with intersecting axes. If $\phi_k\colon S^1 \to \mathcal{L}(V)$ denotes the boundary map associated to ρ_k, we have

$$\mathcal{Y}_{\phi_k(\gamma^+),\phi_k(\gamma^-)} \cap \mathcal{Y}_{\phi_k(\eta^+),\phi_k(\eta^-)} = \{z_k\},$$

and the sequence $(z_k)_{k\in\mathbb{N}}$ in ${}^{\omega}\mathcal{X}_{\lambda}$ represents a point in the intersection $\mathbb{Y}^{\lambda}_{\gamma} \cap \mathbb{Y}^{\lambda}_{\eta}$; see Section 7. Thus we get $\lim_{\omega} d(x_k, z_k)/\lambda_k < \infty$. The same applies to ${}^{\omega}\mathcal{X}_{\mu}$ and hence $\lim_{\omega} d(y_k, z_k)/\mu_S(\rho_k) < \infty$. Using the triangle inequality and taking into account that the sequences $(\lambda_k)_{k\in\mathbb{N}}$ and $(\mu_S(\rho_k))_{k\in\mathbb{N}}$ are equivalent, we deduce

$$\lim_{\omega} \frac{d(x_k, y_k)}{\lambda_k} < \infty. \qquad \square$$

Proof of Corollary 1.3 The first inequality follows from the collar lemma, while the last follows by contradiction from Proposition 10.6. $\square$

Proof of Corollary 1.4 Applying iteratively Theorem 1.1, it is possible to obtain a canonical decomposition of the surface in subsurfaces with geodesic boundary with the property that all curves strictly contained in a subsurface have the same growth rate. The set $\mathcal{C}$ of curves defining this decomposition is the union of the curves given by Theorem 1.1 and all the curves contained in subsurfaces of type (FP) selected by applying Theorem 1.1 to the restrictions of the representations to those subsurfaces. One can apply Theorem 1.1 at most $3g-3+p$ times corresponding to the case when at each step precisely one curve is added and all the complementary pieces are of type (FP). Hence there are at most $3g-3+p$ distinct growth rates among curves having nontrivial intersection with $\mathcal{C}$. There are three possibilities for the remaining curves: either a curve is contained in a subsurface defined by the decomposition $\mathcal{C}$, or it is one of the curves in $\mathcal{C}$ or it corresponds to a puncture in the surface. The claim follows since there are at most $2g-2+p$ complementary components. $\square$

Appendix

Proposition A.1 *Let $\mathbb{F}$ be a real closed field. Let n be a positive integer and assume that $a_1, \dots, a_n \geq 1$. Then we have*

$$(a_1 a_2 \cdots a_n - 1)^n \geq (a_1^n - 1)(a_2^n - 1) \cdots (a_n^n - 1),$$

with equality if and only if $a_1 = \cdots = a_n$.

For $\mathbb{F} = \mathbb{R}$, this follows easily from the convexity of the function $e^x/(e^x - 1)$; here we reproduce the proof due to Thomas Huber for general real closed fields. We start with a key lemma:

Lemma A.2 *Let n be a positive integer, and let $c, x \geq 1$. Then we have*

$$(cx - 1)^n \geq (c^n x - 1)(x - 1)^{n-1}, \tag{10}$$

with equality if and only if $n = 1$ or $c = 1$.

Proof We use induction. For $n = 1$, the inequality is in fact an equality. By induction,

$$(cx - 1)^{n+1} = (cx - 1)(cx - 1)^n \geq (cx - 1)(c^n x - 1)(x - 1)^{n-1}$$

(observe that all factors are nonnegative), and it suffices to show that

$$(cx - 1)(c^n x - 1) \geq (c^{n+1} x - 1)(x - 1)$$

holds. But the difference of the left and the right hand side factors as

$$x(c - 1)^2 (c^{n-1} + \cdots + c + 1)$$

and is clearly nonnegative. □

Now we turn to the proof of the main result and proceed again by induction. For $n = 1$, there is nothing to show; hence let $n \geq 2$. By symmetry, we may assume that $a = a_1 \geq a_i$ for all $i \geq 2$. By the induction hypothesis, the right hand side of the inequality does not decrease when we replace $a_2, \dots, a_n$ by their geometric mean $b = (a_2 \cdots a_n)^{1/(n-1)}$; notice that in a real closed field, positive numbers admit k^{th} roots for any natural number $k \geq 1$. Therefore, it suffices to show the inequality

$$(ab^{n-1} - 1)^n \geq (a^n - 1)(b^n - 1)^{n-1},$$

where $a \geq b \geq 1$. But this is a direct consequence of our lemma: just set $c = a/b \geq 1$ and $x = b^n \geq 1$ in (10). Equality only holds for $c = 1$, that is, for $a = b$. But this implies $a_1 = \cdots = a_n$ by the maximal choice of a_1.

References

[1] **D Alessandrini**, *Tropicalization of group representations*, Algebr. Geom. Topol. 8 (2008) 279–307 MR

[2] **F Bruhat**, **J Tits**, *Groupes réductifs sur un corps local*, Inst. Hautes Études Sci. Publ. Math. 41 (1972) 5–251 MR

[3] **G W Brumfiel**, *The real spectrum compactification of Teichmüller space*, from "Geometry of group representations" (W M Goldman, A R Magid, editors), Contemp. Math. 74, Amer. Math. Soc., Providence, RI (1988) 51–75 MR

[4] **G W Brumfiel**, *The tree of a non-Archimedean hyperbolic plane*, from "Geometry of group representations" (W M Goldman, A R Magid, editors), Contemp. Math. 74, Amer. Math. Soc., Providence, RI (1988) 83–106 MR

[5] **M Burger**, **A Iozzi**, **F Labourie**, **A Wienhard**, *Maximal representations of surface groups: symplectic Anosov structures*, Pure Appl. Math. Q. 1 (2005) 543–590 MR

[6] **M Burger**, **A Iozzi**, **A Wienhard**, *Surface group representations with maximal Toledo invariant*, Ann. of Math. 172 (2010) 517–566 MR

[7] **B Collier**, **Q Li**, *Asymptotics of Higgs bundles in the Hitchin component*, Adv. Math. 307 (2017) 488–558 MR

[8] **P Erdös**, **L Gillman**, **M Henriksen**, *An isomorphism theorem for real-closed fields*, Ann. of Math. 61 (1955) 542–554 MR

[9] **I Kaplansky**, *Linear algebra and geometry: a second course*, revised edition, Dover, Mineola, NY (2003) MR

[10] **L Katzarkov**, **A Noll**, **P Pandit**, **C Simpson**, *Constructing buildings and harmonic maps*, preprint (2015) arXiv

[11] **B Kleiner**, **B Leeb**, *Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings*, Inst. Hautes Études Sci. Publ. Math. 86 (1997) 115–197 MR

[12] **L Kramer**, **S Shelah**, **K Tent**, **S Thomas**, *Asymptotic cones of finitely presented groups*, Adv. Math. 193 (2005) 142–173 MR

[13] **L Kramer**, **K Tent**, *Asymptotic cones and ultrapowers of Lie groups*, Bull. Symbolic Logic 10 (2004) 175–185 MR

[14] **I Le**, *Higher laminations and affine buildings*, Geom. Topol. 20 (2016) 1673–1735 MR

[15] **G-S Lee**, **T Zhang**, *Collar lemma for Hitchin representations*, Geom. Topol. 21 (2017) 2243–2280 MR

[16] **A H Lightstone**, **A Robinson**, *Nonarchimedean fields and asymptotic expansions*, North-Holland, Amsterdam (1975) MR

[17] **G Lion**, **M Vergne**, *The Weil representation, Maslov index and theta series*, Progress in Mathematics 6, Birkhäuser, Boston (1980) MR

[18] **J Loftin**, *Convex $\mathbb{R}P^2$ structures and cubic differentials under neck separation*, preprint (2015) arXiv

[19] **R Mazzeo**, **J Swoboda**, **H Weiss**, **F Witt**, *Ends of the moduli space of Higgs bundles*, Duke Math. J. 165 (2016) 2227–2271 MR

[20] **A Parreau**, *Immeubles affines: construction par les normes et étude des isométries*, from "Crystallographic groups and their generalizations" (P Igodt, H Abels, Y Félix, F Grunewald, editors), Contemp. Math. 262, Amer. Math. Soc., Providence, RI (2000) 263–302 MR

[21] **A Parreau**, *Sous-groupes elliptiques de groupes linéaires sur un corps valué*, J. Lie Theory 13 (2003) 271–278 MR

[22] **A Parreau**, *Compactification d'espaces de représentations de groupes de type fini*, Math. Z. 272 (2012) 51–86 MR

[23] **A Parreau**, *Invariant subspaces for some surface groups acting on A2–Euclidean buildings*, preprint (2015) arXiv

[24] **A Robinson**, *Function theory on some nonarchimedean fields*, Amer. Math. Monthly 80 (1973) 87–109 MR

[25] **I Satake**, *Algebraic structures of symmetric domains*, Kanô Memorial Lectures 4, Iwanami Shoten, Tokyo (1980) MR

[26] **C L Siegel**, *Symplectic geometry*, Amer. J. Math. 65 (1943) 1–86 MR

[27] **T Strubel**, *Fenchel–Nielsen coordinates for maximal representations*, Geom. Dedicata 176 (2015) 45–86 MR

[28] **B Thornton**, *Asymptotic cones of symmetric spaces*, PhD thesis, The University of Utah (2002) MR Available at `https://search.proquest.com/docview/305520472`

[29] **T Zhang**, *The degeneration of convex $\mathbb{RP}^2$ structures on surfaces*, Proc. Lond. Math. Soc. 111 (2015) 967–1012 MR

[30] **T Zhang**, *Degeneration of Hitchin representations along internal sequences*, Geom. Funct. Anal. 25 (2015) 1588–1645 MR

Department of Mathematics, ETH Zentrum
Zürich, Switzerland

Mathematics Institute, University of Warwick
Coventry, United Kingdom

`burger@math.ethz.ch`, `b.pozzetti@warwick.ac.uk`

Proposed: Ian Agol Received: 4 November 2015
Seconded: Leonid Polterovich, Yasha Eliashberg Revised: 15 October 2016

msp

Geometry & Topology 21 (2017) 3601–3657

C^0 approximations of foliations

WILLIAM H KAZEZ
RACHEL ROBERTS

Suppose that $\mathcal{F}$ is a transversely oriented, codimension-one foliation of a connected, closed, oriented 3–manifold. Suppose also that $\mathcal{F}$ has continuous tangent plane field and is *taut*; that is, closed smooth transversals to $\mathcal{F}$ pass through every point of M. We show that if $\mathcal{F}$ is not the product foliation $S^1 \times S^2$, then $\mathcal{F}$ can be C^0 approximated by weakly symplectically fillable, universally tight contact structures. This extends work of Eliashberg and Thurston on approximations of taut, transversely oriented C^2 foliations to the class of foliations that often arise in branched surface constructions of foliations. This allows applications of contact topology and Floer theory beyond the category of C^2 foliated spaces.

57M50; 53D10

1 Introduction

In [8], Eliashberg and Thurston introduce the notion of *confoliation* and prove that when $k \geq 2$, a transversely oriented C^k foliation $\mathcal{F}$ of a closed, oriented 3–manifold not equal to $S^1 \times S^2$ can be C^0 approximated by a pair of C^∞ contact structures, one positive and one negative. They also prove that when $\mathcal{F}$ is also taut, any contact structure sufficiently close to the tangent plane field of $\mathcal{F}$ is weakly symplectically fillable and universally tight.

The focus of this paper is $C^{k,0}$ foliations (see Section 2 for definitions). These foliations are less smooth than those studied by Eliashberg and Thurston. In the context of applications of $C^{k,0}$ foliations, the natural definition of taut is a foliation for which closed smooth transversals pass through every point of the manifold.

Definition 1.1 A foliation $\mathcal{F}$ of M is *taut* if for each $p \in M$, there is a simple closed curve everywhere transverse to $\mathcal{F}$ and passing through p.

Given sufficient smoothness, this is equivalent to the usual definition, which stipulates the existence of transversals through every leaf of the foliation; see Kazez and Roberts [26]. In addition, when dealing with $C^{k,0}$ foliations, it is important to specify whether smooth or topological transversals are being used [26]. We use the terms *transverse, transversal* and *transversely* in the smooth sense; that is, they refer to

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3601

smooth objects intersecting so that the associated tangent spaces intersect minimally. In contrast, a curve is *topologically transverse* to $\mathcal{F}$ if no nondegenerate subarc is isotopic, relative to its endpoints, into a leaf of $\mathcal{F}$.

In this paper, we complete the project started in Kazez and Roberts [24] and prove that the requirement that $\mathcal{F}$ be C^2 can be weakened to the condition that $\mathcal{F}$ be $C^{1,0}$; equivalently, that the tangent plane field of $\mathcal{F}$ be defined and continuous. Our main theorem is the following.

Theorem 1.2 *Let M be a closed, connected, oriented 3–manifold, and let $\mathcal{F}$ be a transversely oriented $C^{1,0}$ foliation on M. Then $\mathcal{F}$ can be C^0 approximated by a positive (resp. negative) contact structure ξ_+ (resp. ξ_-) if and only if $\mathcal{F}$ is not a foliation of $S^1 \times S^2$ by spheres. When $\mathcal{F}$ is taut, these contact structures, (M, ξ^+) and $(-M, \xi^-)$, are weakly symplectically fillable and universally tight.*

In addition, we take the proof of this theorem as an opportunity to revisit classical foliation results in the topological setting, and attempt to state as clearly as possibly what is true in the TOP category. Proofs are given when we could not find them in the literature.

Tautness of a $C^{1,0}$ foliation $\mathcal{F}$ guarantees the existence of a transverse, smooth, volume-preserving flow [26, Theorem 6.1]. The existence of a volume-preserving flow transverse to $\mathcal{F}$ is used in Proposition 3.2.2 of Eliashberg and Thurston [8] to conclude weak symplectic fillability of the approximating contact structure. Tightness then follows from Theorem 3.2.4, and universal tightness by Proposition 3.5.5, of [8]. Theorem 1.2 therefore implies the following.

Corollary 1.3 *Let $\mathcal{F}$ be a transversely oriented $C^{1,0}$ foliation on a closed, connected, oriented 3–manifold $M \neq S^1 \times S^2$. If Φ is a smooth volume-preserving flow transverse to $\mathcal{F}$, then any positive contact structure transverse to Φ is weakly symplectically fillable and universally tight.*

With the additional assumptions that $\mathcal{F}$ is $C^{\infty,0}$ and is not a foliation of T^3 by planes, Jonathan Bowden [1] has also proved Theorem 1.2, and hence any accompanying corollaries. His approach is similar to ours but uses different propagation techniques.

In this paper, M will denote a connected, closed, oriented 3–manifold, and $\mathcal{F}$ will be a transversely oriented (not necessarily taut) foliation in M.

There are two useful notions that we use in describing approximations of a given foliation $\mathcal{F}$ by either another foliation, a contact structure or indeed any 2–plane field ξ

on M. The first is C^0 approximation. This can be defined using the standard topology, or a compatible metric, on the Grassmannian bundle of continuous tangent 2–planes over M. Pick a compatible metric. If for all $\epsilon > 0$ there exists a 2–plane field ξ on M of a particular type (eg corresponding to a foliation in Theorem 1.4, or a contact structure in Theorem 1.2) that is within ϵ of $T\mathcal{F}$, then we say $\mathcal{F}$ is C^0 close to a 2–plane bundle of the type of ξ.

The starting point for the second notion of approximation is a foliation $\mathcal{F}$ together with a transverse flow Φ. A tangent 2–plane distribution ξ on M is Φ*–close* to $\mathcal{F}$ if it is also transverse to Φ. Since the notion of Φ–closeness is purely topological, it is very well suited to the study of continuous plane fields. When $\mathcal{F}$ is taut, there exists a volume-preserving flow Φ transverse to $\mathcal{F}$, and this notion of Φ approximation is particularly useful, and is the focus of Kazez and Roberts [24], because it is sufficient that a contact structure ξ be Φ–close to $\mathcal{F}$ in order to conclude that it is weakly symplectically fillable and universally tight. Clearly Φ–close is implied by C^0 close, and it is the latter notion that is the focus of this paper.

The construction in Eliashberg and Thurston [8] of a contact structure approximating a foliation consists of two steps. First, a contact structure is constructed in neighborhoods of curves in leaves of the foliation about which there is contracting holonomy. Next the foliation is used to propagate the contact structure to the remainder of the manifold.

To carry out this strategy for $C^{1,0}$ foliations, extra care is required. The preferred neighborhoods of curves must be chosen to be particularly thin, and it may be that a foliation has no such thin neighborhoods. In such a case, a new foliation that C^0 approximates the first and has thin, contracting holonomy is constructed. This is done by the method of generalized Denjoy blow up and is described in Section 5.

The key issue in attempting to propagate a contact structure defined only on a subset V of the manifold to the entire manifold is the possibility that not every point outside of V is connected to V by a leaf of the foliation. The main ideas in finding such a set V are discussed in Section 4. Since we are changing the foliation and studying a possibly changing minimal set, arguments have to be made that this is a finite procedure, and that the desired V is just a finite union of thin contracting neighborhoods.

The culmination of this strategy is Theorem 6.2, which can be roughly stated as:

Theorem 1.4 *A transversely oriented $C^{1,0}$ foliation $\mathcal{F}$ not equal to $S^1 \times S^2$ is C^0 close to a $C^{\infty,0}$ foliation $\mathcal{F}'$ for which there exists a subset V of M with the following properties. First, V is a disjoint union of finitely many thin solid tori each of which is a standard neighborhood of a closed curve in a leaf for which the holonomy has a contracting interval. Second, M is V–transitive with respect to $\mathcal{F}'$.*

Much of the focus of the paper is on creating and approximating nontrivial holonomy in foliations. Section 8 covers the case in which no nontrivial holonomy exists, including the case that all leaves of the foliation are planes. Throughout the paper, *plane* is used in the topological sense; that is, a plane is a surface homeomorphic to $\mathbb{R}^2$. We use I to mean a nondegenerate closed interval in places, and to mean $I = [0, 1]$ in other places, and the meaning should be clear from context.

In Section 9 we briefly recall the techniques introduced in [24] to propagate a contact structure along a $C^{\infty,0}$ foliation and thereby conclude our main result, Theorem 1.2, from Theorem 1.4.

In [2], Calegari proves the theorem, proposed in Gabai [18] as folklore in need of proof, that any C^0 foliation can be isotoped to a $C^{\infty,0}$ foliation. This leads to an existence, as opposed to an approximation, result. That is, the existence of a taut, oriented C^0 foliation is sufficient to guarantee the existence of a pair of contact structures, one positive and one negative.

Corollary 1.5 *Suppose M is a closed, oriented 3–manifold that contains a oriented C^0 foliation $\mathcal{F}$ not equal to $S^1 \times S^2$. Then M contains a pair of contact structures ξ^+ and ξ^-, one positive and one negative, that may be chosen arbitrarily C^0 close to each other. When $\mathcal{F}$ is taut, these contact structures, (M, ξ^+) and $(-M, \xi^-)$, are weakly symplectically fillable and universally tight.*

In [35], Ozsváth and Szabó use the Eliashberg–Thurston existence theorem to show that L–spaces do not admit transversely orientable, taut C^2 foliations. Since many constructions of foliations — see Dasbach and Li [5], Delman and Roberts [6], Gabai [10; 11; 12; 13; 14; 15; 17], Kalelkar and Roberts [23], Kazez and Roberts [24], Li [29; 30], Li and Roberts [31] and Roberts [36; 37; 38] — are not C^2, it is useful to be able to remove the smoothness assumption from their theorem.

A C^0 foliation is *topologically taut* if there is a topological transversal through every leaf of the foliation. By Corollary 5.6 of Kazez and Roberts [26], a topologically taut C^0 foliation is isotopic to a taut $C^{\infty,0}$ foliation, and hence the differences between the versions of tautness are unimportant when working with foliations up to topological conjugacy.

Corollary 1.6 *An L–space does not contain a transversely orientable, topologically taut C^0 foliation.*

One of the motivations for our work is the uniqueness theorem proved by Vogel [43]. He shows that with mild assumptions on the leaves of a C^2 foliation, which are necessarily satisfied in an atoroidal, irreducible manifold, sufficiently close approximating contact structures must be isotopic to each other. Although our work does not address the uniqueness question for approximations of $C^{1,0}$ foliations, one of our main tools, flow box decompositions, seems well suited for comparing a pair of close contact structures.

This suggests questions related to uniqueness that can be asked in an atoroidal, irreducible manifold.

(1) If ξ_1 and ξ_2 are sufficiently close to a taut, transversely oriented $C^{1,0}$ foliation $\mathcal{F}$, are they isotopic?

(2) Can the contact invariant of an approximating contact structure be computed, and shown not to vanish, directly from the foliation?

(3) If ξ_1 and ξ_2 are sufficiently close to a taut, transversely oriented $C^{1,0}$ foliation $\mathcal{F}$, are the contact invariants determined by ξ_1 and ξ_2 necessarily equal?

Acknowledgements We thank Jonathan Bowden for helpful feedback on a preliminary version of this paper. Our focus was originally on taut foliations, as our goal was to establish the tightness of certain contact structures, and our theorem statements reflected this, but Bowden made the interesting observation that the condition that the foliation be taut is not necessary for the existence of a pair of approximating contact structures. Tautness is necessary only to conclude that there is a volume-preserving transverse flow, and hence that the approximating contact structures are weakly symplectically fillable and universally tight. Subsequently, we incorporated this observation in the statement of Theorem 1.2, and removed superfluous references to tautness throughout the paper. Proofs remained unchanged, except for the proof of Corollary 8.16, which was modified to use Lemma 2.12 of [1].

We also thank Vincent Colin, Larry Conlon and John Etnyre for many helpful conversations. We would also like to thank the referee for many helpful comments.

This work was partially supported by grants from the Simons Foundation (#244855 to Kazez, #317884 to Roberts) and from the National Science Foundation (DMS-1612036 to Kazez, DMS-1612475 to Roberts).

2 Background

We begin by defining foliations in 3–manifolds with empty boundary. Near the end of this section, we extend these definitions to 3–manifolds with nonempty boundary that are smooth or smooth with corners; namely, manifolds locally modeled by open sets in $[0,\infty)^3$.

Definition 2.1 Let M be a smooth 3–manifold with empty boundary. Let k and l be nonnegative integers or infinity with $l \le k$. Both C^k and $C^{k,l}$ *codimension-one* *foliations* $\mathcal{F}$ are decompositions of M into a disjoint union of C^k immersed connected surfaces, called the *leaves* of $\mathcal{F}$, together with a collection of charts U_i covering M, with $\phi_i\colon \mathbb{R}^2 \times \mathbb{R} \to U_i$ a homeomorphism, such that the preimage of each component of a leaf intersected with U_i is a horizontal plane.

The foliation $\mathcal{F}$ is C^k if the charts (U_i, ϕ_i) can be chosen so that each ϕ_i is a C^k diffeomorphism.

The foliation $\mathcal{F}$ is $C^{k,l}$ if for all i and j,

(1) the derivatives $\partial_x^a \partial_y^b \partial_z^c$, taken in any order, on the domain of each ϕ_i and each transition function $\phi_j^{-1}\phi_i$ are continuous for all $a+b \leq k$, and $c \leq l$, and

(2) if $l \geq 1$, then ϕ_i is a C^1 diffeomorphism.

Remark 2.2 The smoothness conditions on both the charts and the transition functions are to ensure that the smooth structure on the leaves is compatible with the smooth structure on M.

In particular, $T\mathcal{F}$ exists and is continuous if and only if $\mathcal{F}$ is $C^{1,0}$. Also notice that $C^{k,l}$ foliations are C^l, but not conversely.

Two $C^{k,0}$ foliations $\mathcal{F}$ and $\mathcal{G}$ of M are called $C^{k,0}$*–equivalent* if there is a self-homeomorphism of M that maps the leaves of $\mathcal{F}$ to the leaves $\mathcal{G}$, and is C^k when restricted to any leaf of $\mathcal{F}$.

Remark 2.3 Definition 2.1 is an amalgamation of two definitions due to Candel and Conlon [4, 1.2.22, 1.2.24]. In one of their definitions, they allow for the possibility that the ambient manifold is not given a differentiable structure or that it may have a differentiable structure that does not contain $T\mathcal{F}$ as a subbundle. Since a topological 3–manifold admits a unique smoothness structure [32], we forgo this generality and require leaves of $\mathcal{F}$ to be C^k immersed in M.

Given a codimension-one foliation $\mathcal{F}$, it is useful to fix a flow Φ transverse to $\mathcal{F}$. Even when the leaves of the foliation are only topologically immersed, Siebenmann shows [41, Theorem 6.26] — see also Chapter IV of [21] — that there is a 1–dimensional transverse foliation. For foliations with smoother leaves it is much easier to construct a transverse flow. We next recall some basic facts about flows.

Definition 2.4 (see, for example, Section 17 of [27]) A *global flow* on M is an action of $\mathbb{R}$ on M; that is, a continuous map

$$\phi\colon M \times \mathbb{R} \to M$$

such that

(1) $\phi(\phi(p,s),t) = \phi(p,t+s)$, and

(2) $\phi(p,0) = p$.

For each $p \in M$, there is a curve $\phi_p\colon \mathbb{R} \to M$ given by $\phi_p(t) = \phi(p,t)$.

Proposition 2.5 [27, Proposition 17.3] *Let $\phi\colon M \times \mathbb{R} \to M$ be a smooth global flow. The vector field $V\colon M \to TM$ given by $V(p) = \phi_p'(0)$ is smooth, and each curve ϕ_p is an integral curve of V.* □

We call V the vector field *determined by* the flow ϕ. Let Φ denote the 1–dimensional smooth foliation with leaves the integral curves ϕ_p of V. Conversely, since M is compact, any smooth vector field determines a smooth flow.

Theorem 2.6 [27, Theorems 17.8 and 17.11] *Given a smooth vector field V on M, there is a unique global flow $\phi\colon M \times \mathbb{R} \to M$ such that V is the vector field determined by ϕ.* □

So any choice of nowhere-zero tangent vector field to a smooth 1–dimensional foliation determines a smooth global flow on M.

If there is a continuously varying vector field transverse to the leaves of a foliation $\mathcal{F}$, then $\mathcal{F}$ is *transversely orientable*.

Theorem 2.7 [41, Theorem 6.26; 21, Theorems 1.1.2 and 1.3.2] *Let $\mathcal{F}$ be a codimension-one, transversely oriented C^0 foliation. There is a continuous flow Φ transverse to $\mathcal{F}$.*

When M is $C^{1,0}$, this flow can be chosen to be smooth. In fact, there is a smooth nowhere-zero vector field $V\colon M \to TM$ everywhere positively transverse to $\mathcal{F}$, and if M is given a Riemannian metric, then V can be chosen to lie arbitrarily C^0 close to the continuous vector field of vectors perpendicular to $T\mathcal{F}$.

Proof We are primarily interested in the case that $\mathcal{F}$ is $C^{1,0}$. Since in this case the proof is both immediate and enlightening, we reproduce it here. Fix a Riemannian metric on M. Since $\mathcal{F}$ is transversely oriented, each tangent plane $T_p\mathcal{F}$ has a preferred orientation. For each $p \in M$, let $V_\perp(p)$ denote the positive unit normal to the tangent plane $T_p\mathcal{F}$. Approximate $V_\perp(p)$ by a smooth vector field V. If the approximation is taken close enough it will be nonzero and transverse to $T\mathcal{F}$. □

When a foliation $\mathcal{F}$ and a transverse flow Φ are understood, a submanifold of positive codimension in M is called *horizontal* if each component is a submanifold of a leaf of $\mathcal{F}$ and *vertical* if it can be expressed as a union of subsegments of the flow Φ. A codimension-0 submanifold X of M is called $(\mathcal{F},\Phi)$*–compatible* if its boundary is piecewise horizontal and vertical, and hence $\mathcal{F}$ and Φ restrict naturally to foliation and flow on X. If X is $(\mathcal{F},\Phi)$–compatible, let $\partial_v X$ denote its vertical boundary and let $\partial_h V$ denote its horizontal boundary.

Definition 2.8 [24] Let $\mathcal{F}$ be either a C^k or $C^{k,l}$ foliation, and let Φ be a transverse flow. A *flow box*, F, is an $(\mathcal{F},\Phi)$–compatible closed chart, possibly with corners. That is, it is a submanifold diffeomorphic to $D\times I$, where D is either a closed C^k disk or a polygon (a closed disk with at least three corners), Φ intersects F in the arcs $\{(x,y)\}\times I$, and each component of $D\times\partial I$ is embedded in a leaf of $\mathcal{F}$. The components of $\mathcal{F}\cap F$ give a family of graphs over D.

In the case that D is a polygon, it is often useful to view the disk D as a 2–cell with ∂D the cell complex obtained by letting the vertices correspond exactly to the corners of D. Similarly, it is useful to view the flow box F as a 3–cell possessing the product cell complex structure of $D\times I$. Then $\partial_h F$ is a union of two (horizontal) 2–cells and $\partial_v F$ is a union of c (vertical) 2–cells, where c is the number of corners of D. In the case that D has no corners, we abuse language slightly and consider $\partial_h F$ to be a union of two (horizontal) 2–cells and $\partial_v F$ to be a single vertical face, where the face is the entire vertical annulus $\partial D\times I$.

Suppose V is either empty or else a compact, $(\mathcal{F},\Phi)$–compatible, codimension-0 submanifold of M. A *flow box decomposition of M rel V*, or simply *flow box decomposition of M* if $V=\varnothing$, is a decomposition of $M\setminus\operatorname{int}V$ as a finite union $M=V\cup\left(\bigcup_i F_i\right)$, where

(1) each F_i is a flow box,

(2) the interiors of F_i and F_j are disjoint if $i\neq j$, and

(3) if $i\neq j$ and $F_i\cap F_j$ is nonempty, it must be homeomorphic to a point, an interval or a disk that is wholly contained either in $\partial_h F_i\cap\partial_h F_j$ or in a single face in each of $\partial_v F_i$ and $\partial_v F_j$.

In [25], we develop a theory of flow box decompositions and show that they are particularly well suited to the study of codimension-one foliations. Their role is similar to that played by triangulations and branched surfaces, but they are perhaps better suited to the consideration of differentiability properties.

Lemma 2.9 [24] *Let $\mathcal{F}$ be a C^0 foliation and Φ a C^0 transverse flow. There is a flow box decomposition for $(M,\mathcal{F},\Phi)$. When $\mathcal{F}$ is $C^{k,0}$ with $k\geq 1$, and Φ is smooth, this flow box decomposition can be chosen to be C^k.* □

Two 2–plane bundles, for instance two contact structures, are said to be C^0 *close*, if at each point, the associated 2–planes are close in the associated Grassmann bundle of 2–planes. Two $C^{1,0}$ foliations, $\mathcal{F}$ and $\mathcal{G}$, are C^0 *close* if the associated 2–plane bundles $T\mathcal{F}$ and $T\mathcal{G}$ are C^0 close. A 2–plane bundle, for instance a contact structure, is C^0 *close* to a $C^{1,0}$ foliation $\mathcal{F}$ if it is C^0 close to $T\mathcal{F}$. A diffeomorphism C^1 close to the identity map will preserve C^0 proximity of foliations and contact structures.

As the next theorem and its corollary show, there is often no loss of generality in restricting attention to foliations with smooth leaves.

Theorem 2.10 [2; 25] *Suppose $\mathcal{F}$ is a $C^{1,0}$ foliation in M. Then there is an isotopy of M taking $\mathcal{F}$ to a $C^{\infty,0}$ foliation $\mathcal{G}$ which is C^0 close to $\mathcal{F}$. If Φ is a smooth flow transverse to $\mathcal{F}$, the isotopy may be taken to map each flow line of Φ to itself.*

Corollary 2.11 *If every $C^{\infty,0}$ transversely oriented foliation can be C^0 approximated by positive and negative contact structures, then the same is true for every $C^{1,0}$ transversely oriented foliation.* □

3 Holonomy neighborhoods

Let γ be an *oriented* simple closed curve in a leaf L of $\mathcal{F}$, and let p be a point in γ. We are interested in the behavior of $\mathcal{F}$ in a neighborhood of γ. Let h_γ be a holonomy map for $\mathcal{F}$ along γ, and let σ and τ be small closed segments of the flow Φ which contain p in their interiors and satisfy $h_\gamma(\tau) = \sigma$. Choose τ small enough that $\sigma \cup \tau$ is a closed segment and not a loop. Notice that $\sigma \cap \tau$ is necessarily a closed segment containing p in its interior. There are three possibilities:

(1) $\sigma = \tau$,

(2) one of σ and τ is properly contained in the other, or

(3) $\sigma \cap \tau$ is properly contained in each of σ and τ.

We will need to consider very carefully a regular neighborhood of γ which lies nicely with respect to both $\mathcal{F}$ and Φ. To this end, restrict attention to foliations $\mathcal{F}$ which are $C^{\infty,0}$ and transverse flows Φ which are smooth, and suppose that γ is smoothly embedded in L. Let A be the closure of a smooth regular neighborhood of γ in L; so A is a smoothly embedded annulus in L.

Lemma 3.1 *Suppose $\mathcal{F}$ is $C^{\infty,0}$ and Φ is smooth. If τ and A are chosen to be small enough, there is a compact submanifold V of M, smoothly embedded with corners, which satisfies the following:*

(1) *V is homeomorphic to a solid torus.*

(2) *∂V is piecewise vertical and horizontal; namely, ∂V decomposes as a union of subsurfaces $\partial_v V \cup \partial_h V$, where $\partial_v V$ is a union of flow segments of Φ and $\partial_h V$ is a union of two surfaces L_- and L_+, each of which is either a disk or an annulus, contained in leaves of $\mathcal{F}$.*

(3) *each flow segment of $\Phi \cap V$ runs from L_- to L_+.*

(4) *τ is a component of the flow segments of $\Phi \cap V$.*

(5) *A is a leaf of the foliation $\mathcal{F} \cap V$.*

Proof Cover a small open neighborhood of γ by finitely many smooth flow boxes. By passing to a smaller τ and A as necessary, we may suppose that A is covered by two flow boxes with union, V, satisfying the properties (1)–(5). □

Notation 3.2 Denote the neighborhood V of Lemma 3.1 by $V_\gamma(\tau, A)$.

Notice that if $\tau = \sigma$, then $V_\gamma(\tau, A)$ is diffeomorphic to $A \times I$, where I is a non-degenerate closed interval. Otherwise, there is a unique smooth vertical rectangle, R say, such that the result of cutting $V_\gamma(\tau, A)$ open along R, and taking the metric closure, is diffeomorphic to a solid cube.

Notation 3.3 Write $R_\gamma(\tau, A)$ for any smooth vertical rectangle embedded in $V_\gamma(\tau, A)$ such that the result of cutting $V_\gamma(\tau, A)$ open along R, and taking the metric closure, is diffeomorphic to a solid cube. When $V_\gamma(\tau, A)$ is not diffeomorphic to a product, $R_\gamma(\tau, A)$ is uniquely determined. Let $Q_\gamma(\tau, A)$ denote the solid cube obtained by splitting $V_\gamma(\tau, A)$ along $R_\gamma(\tau, A)$.

Note that if γ is essential, then $Q_\gamma(\tau, A)$ can be viewed as an $(\widetilde{\mathcal{F}}, \widetilde{\Phi})$ flow box, where $(\widetilde{\mathcal{F}}, \widetilde{\Phi})$ is the lift of $(\mathcal{F}, \Phi)$ to the universal cover of M.

Definition 3.4 The neighborhood $V_\gamma(\tau, A)$ is called the *holonomy neighborhood determined by* (τ, A), and is called an *attracting neighborhood* if $h_\gamma(\tau)$ is contained in the interior of σ.

Notice that at most one of $V_\gamma(\tau)$ and $V_{-\gamma}(\tau)$ can be attracting. More generally, consider the fixed point set $\mathrm{Fix}(h_\gamma) \subset \tau$ of h_γ. The set $\mathrm{Fix}(h_\gamma)$ is closed and cuts τ into open intervals on which h_γ is either strictly increasing or strictly decreasing. Identify (τ, p) with $(I, 0)$ for some closed interval I containing 0 in its interior. If h_γ is decreasing on $(0, t)$ and increasing on $(-s, 0)$ for some $s, t > 0$, then there is a choice of $\tau' \subset \tau$ determining an attracting neighborhood $V_\gamma(\tau', A)$. Symmetrically, if h_γ is increasing on $(0, t)$ and decreasing on $(-s, 0)$, then there is a choice of $\tau' \subset \tau$ determining an attracting neighborhood $V_{-\gamma}(\tau', A)$.

Definition 3.5 Let $\mathcal{F}$ be a $C^{\infty,0}$ foliation. A set of holonomy neighborhoods $V_{\gamma_1}(\tau_1, A_1), \ldots, V_{\gamma_n}(\tau_n, A_n)$ for $\mathcal{F}$ is *spanning* if each leaf of $\mathcal{F}$ has nonempty intersection with the interior of at least one $V_{\gamma_i}(\tau_i, A_i)$.

Definition 3.6 Let V be the union of pairwise disjoint holonomy neighborhoods $V_{\gamma_1}(\tau_1, A_1), \ldots, V_{\gamma_n}(\tau_n, A_n)$ for $\mathcal{F}$. A $C^{\infty,0}$ foliation $\mathcal{G}$ in M is called *V–compatible with $\mathcal{F}$* (or simply *V–compatible* if $\mathcal{F}$ is clear from context) if each $V_{\gamma_i}(\tau_i, A_i)$ is a holonomy neighborhood for $\mathcal{G}$, with V spanning for $\mathcal{G}$ if it is spanning for $\mathcal{F}$.

Fix a set of pairwise disjoint holonomy neighborhoods $V_{\gamma_1}(\tau_1, A_1), \ldots, V_{\gamma_n}(\tau_n, A_n)$ for $\mathcal{F}$, and let V denote their union. Let $R_i = R_{\gamma_i}(\tau_i, A_i)$ for $1 \le i \le n$ and let R denote the union of the R_i. For each i with $1 \le i \le n$ fix a smooth open neighborhood N_{R_i} of R_i in V_i. Choose each N_{R_i} small enough that its closure, $\overline{N}_{R_i}$, is a closed regular neighborhood of R_i. Let N_R denote the union of the N_{R_i}.

Now, given V, R and N_R, we further constrain the set of foliations $\mathcal{F}$ (that we need to approximate by smooth contact structures) to $C^{\infty,0}$ foliations which are smooth on N_R. The following lemma establishes that we can do this with no loss of generality.

Lemma 3.7 [25] *Let $\mathcal{F}$ be a $C^{\infty,0}$ foliation and let Φ be a smooth flow transverse to $\mathcal{F}$. Let V denote the union of a set of pairwise disjoint holonomy neighborhoods for $\mathcal{F}$ and fix N_R as above. There is an isotopy of M taking $\mathcal{F}$ to a $C^{\infty,0}$ foliation which is both C^0 close to $\mathcal{F}$ and smooth on N_R. This isotopy may be taken to preserve both V and the flow lines of Φ setwise.*

Next we describe a preferred product parametrization on a closed set containing V. In this paper, we express S^1 as the quotient $S^1 = [-1,1]/\sim$, where $\sim$ is the equivalence relation on $[-1,1]$ which identifies -1 and 1.

Lemma 3.8 *Let $\mathcal{F}$ be a $C^{\infty,0}$ foliation and let Φ be a smooth flow transverse to $\mathcal{F}$. Let V denote the union of pairwise disjoint holonomy neighborhoods $V_i = V_{\gamma_i}(\tau_i, A_i)$ for $\mathcal{F}$ for $1 \le i \le n$, and fix N_R as above. Suppose $\mathcal{F}$ is smooth on N_R. Then for each i with $1 \le i \le n$ there is a pairwise disjoint collection of closed solid tori P_i such that P_i contains V_i and there is a diffeomorphism $P_i \to [-1,1] \times S^1 \times [-1,1]$ which satisfies the following:*

(1) *The flow segments $\Phi \cap P_i$ are identified with the segments $\{(x,y)\} \times [-1,1]$.*

(2) *A_i is identified with $[-1,1] \times S^1 \times \{0\}$.*

(3) *γ_i is identified with $\{0\} \times S^1 \times \{0\}$.*

(4) *R_i is identified with $[-1,1] \times \{1 \sim -1\} \times [-1,1]$.*

(5) *The restriction of the diffeomorphism to N_{R_i} maps leaves of $\mathcal{F}$ to horizontal level surfaces $D_z \times \{z\}$, where D_z is either $[-1,1]\times\left[\frac{1}{2},1\right]$, $[-1,1]\times\left[-1,-\frac{1}{2}\right]$, or it is the union of both with identifications along R_i, that is,*

$$[-1,1]\times\left(\left(\left[\tfrac{1}{2},1\right]\cup\left[-1,-\tfrac{1}{2}\right]\right)/\sim\right).$$

Proof Since V_i is homeomorphic to a solid torus, it is contained in a solid torus which is diffeomorphic to $[-1,1]\times S^1\times[-1,1]$, where the diffeomorphism can be chosen to identify A with $[-1,1]\times S^1\times\{0\}$ and the flow segments $\Phi\cap P$ with the segments $\{(x,y)\}\times[-1,1]$. Moreover, since the restriction of $\mathcal{F}$ to $V_i\cap N_R$ is a smooth product foliation transverse to vertical fibers, and there is a unique such up to diffeomorphism, this diffeomorphism $P_i\to[-1,1]\times S^1\times[-1,1]$ can also be chosen so that the restriction of the diffeomorphism to N_R maps leaves of $\mathcal{F}\cap N_R$ to horizontal level surfaces $D_z\times\{z\}$, where $D_z=[-1,1]\times\left(\left(\left[\frac{1}{2},1\right]\cup\left[-1,-\frac{1}{2}\right]\right)/\sim\right)$. $\square$

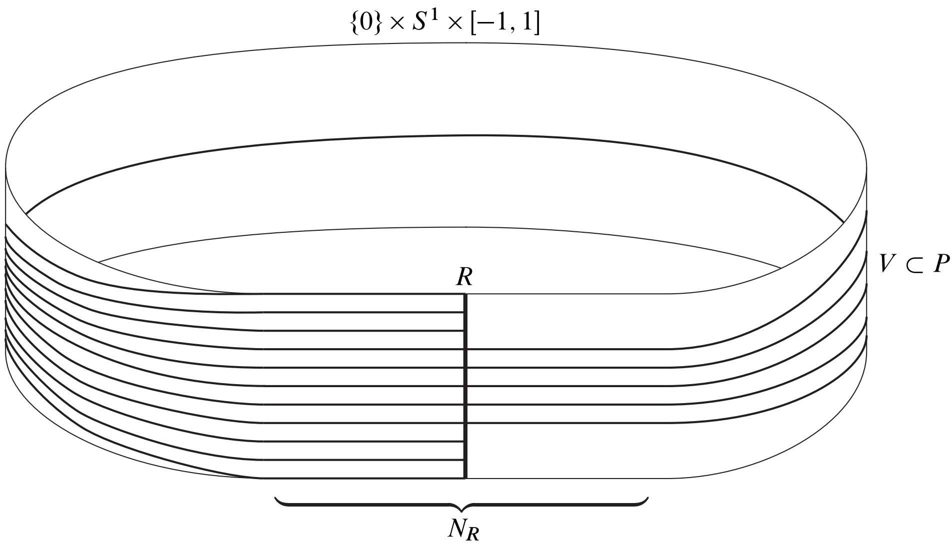


Figure 1: A vertical slice of the image of V under the diffeomorphism from P to $[-1,1]\times S^1\times[-1,1]$

Definition 3.9 Fix V and N_R as above. Let P_i and $P_i\to[-1,1]\times S^1\times[-1,1]$ be as given in Lemma 3.8. Abuse notation and use the diffeomorphism to identify P_i with $[-1,1]\times S^1\times[-1,1]$. Let $\mathcal{P}_i$ be the product foliation of P_i with leaves $([-1,1]\times S^1)\times\{t\}$, and call such a foliated solid torus, $(P_i,\mathcal{P}_i)$, a *product neighborhood of* $(V_i;N_{R_i})$. Letting P denote the union of the P_i and $\mathcal{P}$ denote the union of the $\mathcal{P}_i$, call $(P,\mathcal{P})$ a *product neighborhood of* $(V;N_R)$.

Definition 3.10 Let $\mathcal{F}$ be a $C^{\infty,0}$ foliation and V the union of pairwise disjoint, holonomy neighborhoods $V_{\gamma_i}(\tau_i, A_i)$ for $\mathcal{F}$ for $1 \le i \le k$. Let R denote the union of the $R_{\gamma_i}(\tau_i, A_i)$ for $1 \le i \le k$, and let N_R be an open regular neighborhood of R in V. Let $(P, \mathcal{P})$ be a product neighborhood of $(V; N_R)$. The foliation $\mathcal{F}$ is *strongly* (V, P)*–compatible* if

(1) $\mathcal{F} \cap N_R = \mathcal{P} \cap N_R$, and

(2) in the coordinates inherited from P, the intersection $\mathcal{F} \cap V$ is a product foliation $[-1, 1] \times \mathcal{F}_0$, where $\mathcal{F}_0$ is a $C^{\infty,0}$ foliation of $V \cap (\{0\} \times S^1 \times [-1, 1])$ (ie $\mathcal{F} \cap V$ is x–invariant).

Given V, R and N_R, we will further constrain the set of foliations $\mathcal{F}$ (that we need to approximate by smooth contact structures) to $C^{\infty,0}$ foliations which are strongly (V, P)–compatible for some choice of product neighborhood $(P, \mathcal{P})$. The following lemma establishes that we can do this with no loss of generality; namely, after a small perturbation of $\mathcal{F}$, it is possible to rechoose the diffeomorphisms $P_i \to [-1, 1] \times S^1 \times [-1, 1]$ so that $\mathcal{F} = \mathcal{P}$ on $\overline{N}_R$ and $\mathcal{F}$ is invariant under translation in the first coordinate.

Lemma 3.11 [25] *Let $\mathcal{F}$ be a $C^{\infty,0}$ foliation and let Φ be a smooth flow transverse to $\mathcal{F}$. Let V denote the union of a set of pairwise disjoint holonomy neighborhoods for $\mathcal{F}$ and fix N_R as above. There is an isotopy of M which takes $\mathcal{F}$ to a $C^{\infty,0}$ foliation which is C^0 close to $\mathcal{F}$ and strongly (V, P)–compatible for some choice of product neighborhood $(P, \mathcal{P})$ of $(V; N_R)$. This isotopy may be taken to preserve both V and the flow lines of Φ setwise.*

For simplicity of exposition, it is sometimes useful to fix a product metric on P. We do so as follows.

Definition 3.12 Let $V_\gamma(\tau, A)$ be a holonomy neighborhood and let $(P, \mathcal{P})$ be a product neighborhood of $V_\gamma(\tau, A)$.

Put the standard metric on intervals and let $S^1 = [-1, 1]/\{-1 \sim 1\}$ inherit its metric from $[-1, 1]$. Then put the product metric on $P = [-1, 1] \times S^1 \times [-1, 1]$. When P comes equipped with the identification $P = [-1, 1] \times S^1 \times [-1, 1]$ as given in Lemma 3.8 together with this product metric, the product neighborhood $(P, \mathcal{P})$ is called a *metric product neighborhood of* $V_\gamma(\tau, A)$.

A metric product neighborhood induces a metric on both $V_\gamma(\tau, A) \subset P$ and, via lengths of paths, on $Q_\gamma(\tau, A)$.

Proposition 3.13 *Suppose $V_\gamma(\tau, A)$ is a holonomy neighborhood with metric product neighborhood $(P, \mathcal{P})$. Choose $\epsilon > 0$. There is a closed transversal $\tau' \subset \tau$ containing p in its interior such that $T(\mathcal{F} \cap V_\gamma(\tau', A))$ is ϵ C^0 close to $T(\mathcal{P} \cap V_\gamma(\tau', A))$.*

Proof At each point x of γ, choose a foliation chart of the form $U_x \times (-\delta_x, \delta_x)$ such that U_x is a neighborhood of x in A and $\{x\} \times (-\delta_x, \delta_x)$ is transverse to both $\mathcal{P}$ and $\mathcal{F}$. Since both foliations are at least $C^{\infty,0}$ and contain A as a leaf, δ_x may be chosen small enough that $T\mathcal{P}$ and $T\mathcal{F}$ are within ϵ at all points of $U_x \times (-\delta_x, \delta_x)$. Passing to a finite cover, there exists a short enough transversal $\tau' \subset \tau$ about p such that all its leaf-preserving translates sweep out the desired neighborhood $V_\gamma(\tau', A)$. □

Definition 3.14 Suppose $V_\gamma(\tau, A)$ is a holonomy neighborhood with product neighborhood $(P, \mathcal{P})$. Choose a metric on P. The restriction of $\mathcal{F}$ to $V_\gamma(\tau, A)$ is *ϵ–horizontal* if $T(\mathcal{F} \cap V_\gamma(\tau', A))$ is ϵ C^0 close to $T(\mathcal{P} \cap V_\gamma(\tau', A))$.

Definition 3.15 Suppose $V_\gamma(\tau, A)$ is a holonomy neighborhood for $\mathcal{F}$ with metric product neighborhood $(P, \mathcal{P})$. Fix $\epsilon > 0$. $V_\gamma(\tau, A)$ is *ϵ–flat* if the length of the longest segment of the restriction of Φ to $V_\gamma(\tau, A)$ is bounded above by ϵ.

In other words, $V_\gamma(\tau, A)$ is ϵ–flat if its maximum "height" is small relative to both its minimum "width" and its minimum "length". The next lemma follows immediately from the continuity of $T\mathcal{F}$.

Lemma 3.16 *Fix $\epsilon > 0$. There is a $\delta > 0$ so that if $V_\gamma(\tau, A)$ is δ–flat, then the restriction of $\mathcal{F}$ to $V_\gamma(\tau, A)$ is ϵ–horizontal.* □

4 Minimal sets

Holonomy neighborhoods which are attracting play a particularly important role in the construction of approximating contact structures. Results from the last section will eventually be used to make C^0 approximations in a single (attracting) holonomy neighborhood. Results in this section will be used to choose or create enough attracting holonomy neighborhoods that these approximations can be extended to all of M. This is accomplished through the use of minimal sets.

Definition 4.1 Let $\mathcal{F}$ be a C^0 foliation. A *minimal set* of $\mathcal{F}$ is a nonempty closed subset of M that is a union of leaves and is minimal with respect to inclusion among such sets.

Equivalently, a subset Λ of M is a minimal set if and only if both $\Lambda = \overline{L}$ for some leaf L and Λ properly contains no leaf closure. In particular, if Λ is a minimal set, then $\Lambda = \overline{L}$ for all leaves L contained in Λ. More generally, if L is any leaf of $\mathcal{F}$, then by Zorn's lemma, $\overline{L}$ contains at least one minimal set.

The next two results follow from elementary point set topology. See, for example, [3, Chapter 3, Theorem 1].

Proposition 4.2 *Let Λ be a minimal set of $\mathcal{F}$. Then every leaf of $\mathcal{F}$ contained in Λ is dense in Λ.* □

Proposition 4.3 *Let Λ be a minimal set of $\mathcal{F}$ and let τ be a transversal to $\mathcal{F}$ satisfying $\Lambda \cap \operatorname{int} \tau \neq \varnothing$. Then the following properties hold:*

(1) *The intersection $\Lambda \cap \tau$ is either discrete, all of τ or a Cantor set.*

(2) *$\Lambda \cap \tau = \tau$ if and only if $\mathcal{F} = \Lambda$ is minimal.*

(3) *$\Lambda \cap \tau$ is discrete if and only if Λ consists of a single compact leaf.*

Moreover, if σ is any other transversal to $\mathcal{F}$ satisfying $\Lambda \cap \operatorname{int} \sigma \neq \varnothing$, then both $\Lambda \cap \tau$ and $\Lambda \cap \sigma$ are of the same type, namely, discrete, all of the transverse interval, or Cantor. □

A minimal set Λ is called *exceptional* if $\Lambda \cap \tau$ is a Cantor set for some transversal τ. A foliation is called *minimal* if it is itself a minimal set.

Definition 4.4 Suppose that Λ is a minimal set and that τ is a transversal to $\mathcal{F}$. When $\Lambda \cap \tau$ is a Cantor set, we differentiate between those points of $\Lambda \cap \tau$ which are endpoints of "removed intervals" and those which are not. The leaves of Λ which correspond to points which are endpoints of "removed intervals" are isolated on one side in Λ and so we refer to them as *boundary leaves*. The leaves L of Λ which do not correspond to endpoints of "removed intervals" are approached arbitrarily closely on both sides by leaves in Λ (and hence by leaves of L). Call these leaves *nonboundary leaves*.

When $\Lambda \cap \tau = \tau$, Proposition 4.3 implies $\Lambda = M$ and all leaves of Λ are nonboundary leaves. When $\Lambda \cap \tau$ is discrete, the single compact leaf of Λ is a boundary leaf.

Definition 4.5 Let $p \in M$. A closed segment τ is a *transversal about* p if τ is a transversal to $\mathcal{F}$ which contains p in its interior.

Lemma 4.6 *Let $S \times \{0, 1\}$ be leaves of $\mathcal{F}$ which bound an I–bundle $S \times I$ such that each I–fiber is transverse to $\mathcal{F}$. If $\Lambda \subset S \times I$ is a minimal set of $\mathcal{F}$, then necessarily S is compact and Λ is isotopic to $S \times \{0\}$.*

Proof Begin by noting that since any I–fiber $\{x\}\times(0,1)$ for $x\in S$ is disjoint from $S\times\{0,1\}$, the closure $\overline{S\times\{0\}}$ has empty intersection with $S\times(0,1)$. Hence, either S is compact and Λ is a component of $S\times\{0,1\}$, or $S\times\{0,1\}$ is disjoint from Λ.

Restrict attention therefore to the case that $S\times\{0,1\}$ is disjoint from Λ. Since Λ is closed, for each $x\in S$ there is a minimum point m_x of $(\{x\}\times I)\cap\Lambda$. Let $L_{\min}$ be collection of all such points m_x as x ranges over S. In the analogous way define $L_{\max}$ to be the union of points M_x.

Notice that if $B=D\times I\subset S\times I$ is any flow box, each of $L_{\min}\cap B$ and $L_{\max}\cap B$ is a component of $\mathcal{F}\cap B$. It follows that $L_{\min}$ and $L_{\max}$ are leaves of Λ. If $m_x<M_X$ for some x, it follows that $\overline{L}_{\min}\cap\overline{L}_{\max}=\varnothing$, a contradiction. Thus $L_{\min}=L_{\max}=\Lambda$ is compact, and isotopic to $S\times\{0\}$. It follows that S is necessarily compact. □

In Section 8, we will investigate foliations with only trivial holonomy, and the following lemma will prove useful. It is included here as its proof uses flow box decompositions, but is not needed for any of the results of this section.

Lemma 4.7 *Let $S\times\{0,1\}$ be leaves of $\mathcal{F}$ which bound an I–bundle $S\times I$ in M. Suppose that $\mathcal{F}$ lies transverse to the I–fibers and has only trivial holonomy. Then the restriction of $\mathcal{F}$ to $S\times I$ is, up to a Φ–preserving isotopy, the product foliation $S\times I$.*

Proof Let L be any leaf of the restriction of $\mathcal{F}$ to $S\times(0,1)$. We claim that L intersects each flow box $D\times I$ in exactly one component, and hence is isotopic, via a Φ–preserving isotopy, to $L\times\{0\}$. Consider a component $\Delta=D\times\{a\}$ of the intersection of L with flow boxes $B=D\times I$. Suppose by way of contradiction that $L\cap B$ contains a second component $\Delta'=D\times\{a'\}$.

Let $x\in D$. Since L is connected, there is a path ρ in L from (x,a) to (x,a'). Express ρ as a concatenation of finitely many intervals, each of which lies in a single flow box, and consider the immersed cylinder $\rho\times I$. A standard cut and paste argument reveals that we may assume that this cylinder is embedded. But the existence of the path ρ implies the existence of nontrivial holonomy, a contradiction. Hence, the intersection of L with any flow box is connected. In addition, since S is path-connected, and therefore any two flow boxes are connected by a path in $S\times[0,1]$, the intersection of L with any flow box is also nonempty. Hence, the restriction of $\mathcal{F}$ to $S\times[0,1]$ is, up to a Φ–preserving isotopy, the product foliation $S\times I$. □

Corollary 4.8 *Suppose $\mathcal{F}$ is a C^0 foliation of a compact 3–manifold M.*

(1) *At most finitely many isotopy classes of compact surfaces can be realized as leaves of $\mathcal{F}$.*

(2) *$\mathcal{F}$ can contain only finitely many exceptional minimal sets.*

Proof If $\mathcal{F}$ is not transversely oriented, work instead with $(\widehat{M}, \widehat{\mathcal{F}})$, where $\pi\colon \widehat{M} \to M$ is a double cover such that $\widehat{\mathcal{F}} = \pi^{-1}(\mathcal{F})$ is transversely oriented. (See, for example, Proposition 3.5.1 of [4].) In this case, Λ is a minimal set for $\mathcal{F}$ only if $\pi^{-1}(\Lambda)$ contains a minimal set for $\widehat{\mathcal{F}}$, and hence if the claimed result holds true for $\widehat{\mathcal{F}}$, it holds true for $\mathcal{F}$ also.

So restrict attention to the case that $\mathcal{F}$ is transversely oriented, and let Φ be an oriented flow transverse to $\mathcal{F}$. Choose a flow box decomposition $\mathcal{B}$ for $(M, \mathcal{F}, \Phi)$. Denote a flow box in $\mathcal{B}$ by $B = D \times I$. For each leaf L of $\mathcal{F}$ and flow box B of $\mathcal{B}$, call any component of $L \cap B$ a *plaque* of L.

Let $\mathcal{L} = \{\Lambda_\alpha\}$ denote any finite set of distinct minimal sets of $\mathcal{F}$, and let $\Lambda = \bigcup_\alpha \Lambda_\alpha$. Let B_i for $i = 1, \dots, n$ be those flow boxes such that $B_i \cap \Lambda \neq \varnothing$. Let Y denote the open manifold $M \setminus \Lambda$. Let $\mathcal{S} = \{D_i \times \{0\} \mid 1 \le i \le n\} \cup \{D_i \times \{1\} \mid 1 \le i \le n\}$. Define a map $f\colon \mathcal{S} \to \mathcal{L}$ by $f(D_i \times \{j\}) = \Lambda_{\alpha_0}$ if the plaque of Λ in B_i closest to $D_i \times \{j\}$ lies in a leaf of Λ_{α_0}. Since $\mathcal{S}$ is finite, so is the image of f.

Consider a minimal set Λ_β in $\mathcal{L}$ which is not in the image of f. We claim that Λ_β is a compact leaf L_β and lies as a section of a trivial $\mathbb{R}$–bundle component of $Y \cup \Lambda_\beta$, where the $\mathbb{R}$–fibers are subsegments of the flow Φ.

To see that this is true, proceed as follows. Let Σ denote the component of the open manifold $Y \cup \Lambda_\beta = M \setminus \bigcup_{\alpha \neq \beta} \Lambda_\alpha$ which contains Λ_β. Choose any $p \in L_\beta$, a leaf of Λ_β. This point p lies in a plaque $D_i \times \{b\}$ of Λ_β for some $0 < b < 1$. Consider the two plaques of the closed set $\Lambda \setminus \Lambda_\beta$ which lie closest to $D_i \times \{b\}$. They cobound a flow box $\Sigma_i(a, c) = D_i \times (a, c) \subset \Sigma$, for some $0 \le a < b < c \le 1$. Let L_a be the leaf containing $D_i \times \{a\}$, and let L_c be the leaf containing $D_i \times \{c\}$.

Next consider any $B_j \in \mathcal{B}$ satisfying $\partial_v B_j \cap \Sigma_i(a, c) \neq \varnothing$. Since Λ_β is not in the image of f, we have $\Sigma_i(a, c) \cap \partial_h B_j = \varnothing$, and hence $\Sigma_i(a, c)$ naturally extends to an I–bundle of the form $(D_i \cup D_j) \times \mathbb{R} \subset \Sigma$. Repeat this process of moving through adjacent flow boxes to obtain an exhaustion of L_a by an increasing union of plaques, and an exhaustion of Σ by an increasing union of $\mathbb{R}$–bundles, where the $\mathbb{R}$–fibers are subsegments of the flow Φ and these $\mathbb{R}$–bundles are trivial since the flow is oriented. To see that all of Σ is realized by this exhaustion, note that Σ is path-connected, and any closed path can be decomposed as a piecewise union of finitely many intervals, each of which lies in a single flow box. It follows that Σ is an a trivial $\mathbb{R}$–bundle over L_a, with all $\mathbb{R}$–fibers subsegments of the flow Φ. So the leaves L_a, L_β and L_c are isotopic. Since L_a and L_β bound an I–bundle and $\overline{L}_a \neq \overline{L}_\beta$, it follows from Lemma 4.6 that L_a, L_β and L_c are compact.

Hence, any exceptional minimal set lies in the image of f, and thus, there can be at most $2|\mathcal{B}|$ exceptional minimal sets. Moreover, any compact leaf has an isotopy representative that lies in the image of f, and thus, at most finitely many isotopy classes of compact surface can be realized as leaves of $\mathcal{F}$. □

Remark 4.9 Recall that any compact 3–manifold admits a triangulation [32]. Moreover, given a foliation $\mathcal{F}$, there is a triangulation $\mathcal{T}$ of M which is compatible with $\mathcal{F}$; namely, $\mathcal{F}$ is in Haken normal form with respect to $\mathcal{T}$ [19, Lemma 1.3]. Hence there is a similar proof of Corollary 4.8 which uses triangulations instead of flow box decompositions. Our proof doesn't depend on first finding a good triangulation.

Corollary 4.10 *Suppose Λ is a minimal set and $p \in \Lambda$. Let τ be any transversal about p. If Λ is a compact leaf, then there is a transversal $\tau_0 \subset \tau$ about p such that if a minimal set intersects τ_0, it is a compact leaf isotopic to Λ. If Λ is not a compact leaf, there is a transversal $\tau_0 \subset \tau$ about p which does not intersect any minimal set other than Λ.*

Proof Since minimal sets are closed and there are only finitely many exceptional minimal sets by Corollary 4.8, it is possible to choose τ_0 about p that misses all exceptional minimal sets that are not equal to Λ. Furthermore, Corollary 4.8 implies that if there is no transversal τ_0 about p that is disjoint from all minimal sets not equal to Λ, there must exist a sequence of compact leaves L_i all mutually isotopic that limit on Λ. However, referring to the proof of Corollary 4.8, at most finitely many of the compact leaves L_i lie in the image of f and any other, $S = L_j$ say, lies as a surface fiber in one of finitely many I–bundles of the form $S \times I$, where the surfaces $S \times \{0, 1\}$ are leaves of $\mathcal{F}$. Hence, the nonexistence of τ_0 implies that Λ is a minimal set embedded in an I–bundle $S \times [0, 1]$ transverse to the I–fibers, and therefore by Lemma 4.6 is a compact leaf isotopic to $S = L_j$. □

Notation 4.11 Fix a C^0 foliation $\mathcal{F}$. If $\mathcal{F}$ is minimal, set $\Lambda_1 = \mathcal{F}$. Otherwise, let $\Lambda_1, \dots, \Lambda_r$ denote the exceptional minimal sets of $\mathcal{F}$ and let $[L_1], \dots, [L_s]$ denote the isotopy classes of compact leaves of $\mathcal{F}$.

For each compact leaf of $\mathcal{F}$, let $X(L)$ denote the minimal $\mathcal{F}$–saturated closed subset of M containing all leaves of $\mathcal{F}$ which are isotopic to L.

Corollary 4.12 *Let $\mathcal{F}$ be a C^0 foliation that is not a fibering of M over S^1. Then there are finitely many pairwise disjoint holonomy neighborhoods*

$$V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_{r+s}}(\tau_{r+s}, A_{r+s})$$

such that

(1) *γ_i is an essential simple closed curve in a leaf of Λ_i for $1 \le i \le r$,*

(2) *γ_{r+i} is an essential simple closed curve in a leaf in the isotopy class $[L_i]$ and each leaf isotopic to L_i lies in the interior of $V_{\gamma_{r+i}}(\tau_{r+i}, A_{r+i})$ for $1 \le i \le s$, and*

(3) *each minimal set of $\mathcal{F}$ has nonempty intersection with the interior of exactly one $V_{\gamma_i}(\tau_i, A_i)$.*

In particular, the set of holonomy neighborhoods $\{V_{\gamma_i}(\tau_i, A_i)\}_i$ is spanning.

Conditions (1) and (2) guarantee that each minimal set of $\mathcal{F}$ has nonempty intersection with the interior of at least one $V_{\gamma_i}(\tau_i, A_i)$. So condition (3) guarantees that each $V_{\gamma_i}(\tau_i, A_i)$ for $1 \le i \le r$ has nonempty intersection with exactly one minimal set, Λ_i, and each $V_{\gamma_{r+i}}(\tau_{r+i}, A_{r+i})$ for $1 \le i \le s$ has nonempty intersection with exactly one isotopy class of minimal set, $[L_i]$.

Proof For each i with $1 \le i \le r$ let γ_i be an essential simple closed curve in a leaf of Λ_i, and choose a holonomy neighborhood $V_{\gamma_i}(\tau_i, A_i)$. Choose the (τ_i, A_i) so that the neighborhoods $V_{\gamma_i}(\tau_i, A_i)$ are pairwise disjoint and disjoint from any compact leaf of $\mathcal{F}$. This is possible by Corollary 4.10.

Let L be a compact leaf of $\mathcal{F}$. Recall that $X(L)$ denotes the minimal $\mathcal{F}$–saturated closed subset of M containing all leaves of $\mathcal{F}$ which are isotopic to L. Since $\mathcal{F}$ is not a fibering of M over S^1, either $X(L) = L$ or $X(L) \cong L \times [0,1]$, with the identification given by a homeomorphism. Note that if L and F are compact leaves of $\mathcal{F}$, then either L and F are isotopic and $X(L) = X(F)$, or L and F are not isotopic and $X(L) \cap X(F) = \varnothing$. Rechoose the isotopy class representatives as necessary so that $X(L_1), \dots, X(L_s)$ is a listing of the sets $X(L)$, where either $X(L_i) = L_i$ or L_i is identified with $L_i \times \{0\}$ under the identification $X(L_i) = L_i \times I$.

Set $n = r + s$.

For each j, let γ_{r+j} be an essential simple closed curve in $L_j = L_j \times \{0\}$. Choose a holonomy neighborhood $V_{\gamma_{r+j}}(\tau_{r+j}, A_{r+j})$. If $X(L_j) = L_j \times [0,1]$, choose the transversal τ_{r+j} long enough that its interior has nonempty intersection with $L_j \times \{1\}$ (and hence also with all leaves isotopic to L_j), and short enough that the holonomy neighborhoods $V_{\gamma_{r+1}}(\tau_1, A_1), \dots, V_{\gamma_n}(\tau_n, A_n)$ are pairwise disjoint and disjoint from the minimal sets $\Lambda_1, \dots, \Lambda_r$ and their fixed holonomy neighborhoods $V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_r}(\tau_r, A_r)$.

Since the closure of any leaf contains a minimal set, each leaf of $\mathcal{F}$ has nonempty intersection with the interior of $V_{\gamma_i}(\tau_i, A_i)$ for some i; in other words, the collection of holonomy neighborhoods $\{V_{\gamma_i}(\tau_i, A_i)\}_i$ is spanning. □

5 Generalized Denjoy blow up

In this section we define the operation of generalized Denjoy blow up. Informally, this operation consists of thickening a leaf and inserting a new foliation into the thickened region. This will be used to modify a foliation and create attracting neighborhoods of curves in a leaf of a foliation.

Definition 5.1 Let L be a finite (or even countably infinite) union of leaves of a $C^{k,0}$ foliation $\mathcal{F}$ of M with $k \geq 1$, and let Φ be a smooth flow transverse to $\mathcal{F}$. A $C^{k,0}$ foliation, $\mathcal{F}'$, is a *generalized Denjoy blow up* of $\mathcal{F}$ along L if there is an open subset $U \subset M$ and a collapsing map $h\colon M \to M$ satisfying the following properties:

(1) $\mathcal{F}'$ is transverse to Φ.

(2) There is an injective C^k immersion $j\colon L \times (0,1) \to M$ with $j(L \times (0,1)) = U$.

(3) For each $x \in L$, the image $j(\{x\} \times I)$ is contained in a flow line of Φ.

(4) $j(L \times \{0\})$ and $j(L \times \{1\})$ are leaves of $\mathcal{F}'$.

(5) $h^{-1}(x)$ is a point for $x \notin L$ and equals $j(\{x\} \times I)$ for $x \in L$.

(6) h preserves flow lines of Φ and maps leaves of $\mathcal{F}'$ to leaves of $\mathcal{F}$.

(7) h is C^0 on M and C^k when restricted to any leaf of $\mathcal{F}'$.

When the restriction of $\mathcal{F}'$ to $j(L \times [0,1])$ is a product foliation, $\mathcal{F}'$ is also referred to as a *Denjoy blow up* of $\mathcal{F}$ along L.

Theorem 5.2 [7; 25] *Let $\mathcal{F}$ be $C^{k,0}$ foliation with $k \geq 1$ that is transverse to a smooth flow Φ. Let L be a finite or countable collection of leaves of $\mathcal{F}$, and let $\mathcal{F}_1$ be a $C^{k,0}$ foliation of $L \times I$ transverse to the I coordinate that contains $L \times \partial I$ as leaves. Then there exists $\mathcal{F}'$ arbitrarily C^0 close to $\mathcal{F}$ that is a generalized Denjoy blow up of $\mathcal{F}$ along L, and such that the pullback of $\mathcal{F}'$ to $L \times I$ is $C^{k,0}$–equivalent to $\mathcal{F}_1$.*

Moreover, if V is the union of a set of pairwise disjoint holonomy neighborhoods for $\mathcal{F}$, $(P, \mathcal{P})$ is a product neighborhood of V, and $\mathcal{F}$ is strongly (V, P)–compatible, then $\mathcal{F}'$ can be chosen to be both V–compatible with $\mathcal{F}$ and strongly (V, P)–compatible.

Remark 5.3 The main ideas of Theorem 5.2 are due to Dippolito [7]. We require slightly more than is easily extracted from his work, namely, C^0 approximation. We give a proof of the theorem in [25] using flow box decompositions that also allows for C^1, rather than C^∞, leaves. We also realize the additional conditions that $\mathcal{F}'$ be both V–compatible with $\mathcal{F}$ and strongly (V, P)–compatible.

In order to create attracting neighborhoods, we will be interested in inserting foliations into $L\times(0,1)$ of the form described in the following lemma.

Lemma 5.4 *Let γ be an oriented essential simple closed curve in $L = L\times\{0\}$. There is a $C^{\infty,0}$ foliation on $L\times I$, transverse to the I–fibers $\{x\}\times I$ for $x\in L$ and such that the holonomy h_γ along γ is monotone decreasing on the interior of I. Moreover, this foliation of $L\times I$ can be chosen so that $L\times\{0\}$ and $L\times\{1\}$ are its only minimal sets.*

Proof If γ does not separate, let α be a properly embedded curve in L that intersects γ in a point. Let $\mathcal{P}$ be the product foliation on $(L\backslash\alpha)\times I$, and let $h_\gamma\colon I\to I$ be the desired holonomy about γ. Then glueing the leaves of $\mathcal{P}$ at height x to those of height $h_\gamma(x)$ gives a foliation on $L\times I$ with holonomy h_γ around γ.

If γ separates L into components A and B, first consider the case that A is compact with genus $g\geq 1$. In the usual way, A may be thought of a disk D with $4g$ disjoint subarcs glued in pairs. Let $\mathcal{P}$ be the product foliation on $D\times I$. Using $2g$ homeomorphisms h_i of I to glue up the leaves of $\mathcal{P}$ produces a foliation on $A\times I$. With the right choice of pairings of glued subarcs, the holonomy along γ will be the product of g commutators of the h_i. Since any orientation-preserving homeomorphism of I can be written as a single commutator — see for instance, Lemma 3.1 of [29] — this construction can be carried out for any genus and any choice of h_γ.

Next, consider the case that A is not compact. Let α be a properly embedded half-infinite line contained in A and starting on γ. Splitting A along α and glueing leaves with h_γ as in the nonseparating case gives the desired foliation around γ.

The same constructions are used to extend a given choice of holonomy across B. □

Notation 5.5 Any foliation given by a generalized Denjoy blow up of $\mathcal{F}$ along L, with the foliation inserted into $L\times(0,1)$ of the type generated by Lemma 5.4, is denoted by $\mathcal{F}'=\mathcal{F}(L,\gamma)$.

The following two lemmas use the notation of Definition 5.1 to describe the effect of generalized Denjoy blow up on the set of minimal sets of F.

Lemma 5.6 *Let L be a leaf of $\mathcal{F}$ that is not contained in any minimal set of $\mathcal{F}$. Let $\mathcal{F}'$ be a generalized Denjoy blow up of $\mathcal{F}$ along L. There is a bijective correspondence between the minimal sets of $\mathcal{F}$ and those of $\mathcal{F}'$ given by $\Lambda_\beta\leftrightarrow h^{-1}(\Lambda_\beta)$. In particular, neither L_0, L_1, nor any leaf in $j(L\times(0,1))$ is contained in a minimal set of $\mathcal{F}'$. The restriction of the collapsing function, h, gives a homeomorphism from $h^{-1}(\Lambda_\beta)$ to Λ_β for all β.*

Proof Since $\overline{L}$ is not minimal, it properly contains $\overline{S}$ for some leaf S of $\mathcal{F}$. It follows that each of $\overline{L}_0$ and $\overline{L}_1$ properly contains $h^{-1}(\overline{S}) = \overline{h^{-1}(S)}$. So neither $\overline{L}_0$ nor $\overline{L}_1$ is minimal. For each leaf $F \subset j(L \times (0,1))$ the set $\overline{F}$ properly contains each of L_0 and L_1 and hence each of $\overline{L}_0$ and $\overline{L}_1$. Therefore $\overline{F}$ is not minimal.

Finally, notice that S' is a leaf of $\mathcal{F}'$ not equal to L_0, L_1 or a leaf of $j(L \times (0,1))$ if and only if $S = h(S')$ is a leaf of $\mathcal{F}$ not equal to L. Since $h^{-1}(\overline{S}) = \overline{h^{-1}(S)}$, the claimed bijective correspondence of minimal sets follows immediately. □

Lemma 5.7 *Suppose $\overline{L}$ and Λ are minimal sets of $\mathcal{F}$ with $\overline{L} \neq \Lambda$. Let $\mathcal{F}'$ be a generalized Denjoy blow up of $\mathcal{F}$ along L. Then $h^{-1}(\Lambda)$ is a minimal set of $\mathcal{F}'$. Any other minimal set of $\mathcal{F}'$ arises in one of the following ways:*

(1) *If L is compact, let $\mathcal{F}_1$ be the foliation of $L \times I$ that is contained in $\mathcal{F}'$. Minimal sets of $\mathcal{F}_1$ are mapped to minimal sets of $\mathcal{F}'$ by inclusion; in particular L_0 and L_1 are minimal sets.*

(2) *If L is noncompact, with L a nonboundary leaf of $\overline{L}$, then $\overline{L}_0 = \overline{L}_1$ is a minimal set of $\mathcal{F}'$.*

(3) *If L is noncompact, with L a boundary leaf of $\overline{L}$, then there are two possibilities, depending on whether L is isolated in $\overline{L}$ from above or from below. If L is isolated in $\overline{L}$ from below (above), then $\overline{L}_1$ (resp. $\overline{L}_0$) is a minimal set, and $\overline{L}_0$ (resp. $\overline{L}_1$) properly contains this minimal set. In particular, the leaf L_0 (resp. L_1) is not contained in a minimal set of $\mathcal{F}'$.*

Proof Since $\mathcal{F}'$ is not minimal, any minimal set of $\mathcal{F}'$ is either a compact leaf or else an exceptional minimal set. Moreover, the only minimal set impacted by the blow up of L is $\overline{L}$. So we restrict attention to $\overline{L}$ and $h^{-1}(\overline{L})$.

If L is compact, then so is $h^{-1}(\overline{L}) = j(L \times [0,1])$ and (1) follows immediately.

Suppose instead that L is not compact and Λ' is a minimal set of $\mathcal{F}'$ that intersects $j(L \times (0,1))$. It follows from Lemma 4.6 that $\Lambda' \cap (\overline{L}_0 \cup \overline{L}_1) \neq \varnothing$, and hence that $\Lambda' \subset (\overline{L}_0 \cup \overline{L}_1)$.

Since $h(\overline{L}_0 \cup \overline{L}_1) = \overline{L}$, the union $\overline{L}_0 \cup \overline{L}_1$ cannot properly contain $\overline{S}$, for some leaf S of $\mathcal{F}'$, unless $\overline{S} = \overline{L}_i$ for some $i \in \{0,1\}$. It follows that Λ' is equal to one or both of $\overline{L}_0$ and $\overline{L}_1$.

The cases are distinguished as follows. If L is a nonboundary leaf, $\overline{L}_0 = \overline{L}_1$ is minimal, and if L is a boundary leaf, exactly one of $\overline{L}_0$ and $\overline{L}_1$ is minimal. To see this, let τ be a transversal to $\mathcal{F}$ containing a point $p \in \tau \cap L$. There is a sequence of points $p_n \in \tau \cap L$ which limit on p. If there exists a sequence of such points limiting on p from above, then the points $h^{-1}(p_n) \cap L_1$ limit on L_1 from above. Similarly, if there exists a sequence of such points limiting on p from below, then the points $h^{-1}(p_n) \cap L_0$ limit on L_0 from below. It follows that L is isolated in $\overline{L}$ from below if and only if L_0 is isolated, and that L is isolated in $\overline{L}$ from above if and only if L_1 is isolated. So either L is a nonboundary leaf and $\overline{L}_0 = \overline{L}_1$ is minimal, or L is a boundary leaf, isolated from either above or below, and exactly one of $\overline{L}_0$ or $\overline{L}_1$ is minimal. □

6 Creating attracting holonomy

In this section, we restrict attention to the case that every minimal set of $\mathcal{F}$ contains a leaf which is not homeomorphic to $\mathbb{R}^2$. The remaining case, in which $M = T^3$ [16], is considered in Section 8.

Consider a minimal set Λ of a foliation $\mathcal{F}$. Restrict attention to the case that Λ is not a compact leaf. So either $\Lambda = \mathcal{F}$ or Λ is exceptional. When the foliation $\mathcal{F}$ under consideration is C^2, a result of Sacksteder [39] guarantees the existence of a leaf L in Λ and simple closed curve γ in L such that the holonomy h_γ along γ is linear attracting, that is $h'(0) < 1$. As shown by Eliashberg and Thurston, this combination of smoothness and linear attracting holonomy can be used to introduce a contact region in a neighborhood of γ.

When the foliation $\mathcal{F}$ is only $C^{\infty,0}$, it is shown in [24] that it is possible to introduce contact regions about a simple closed loop γ in a leaf L of Λ for which the foliation has a (topologically) attracting neighborhood. In general, however, there might be no curve with such an attracting region.

In this section, we show that by taking advantage of generalized Denjoy blow up, it is possible to C^0 approximate $\mathcal{F}$ by a foliation $\mathcal{F}'$, where each minimal set of $\mathcal{F}'$ has nonempty intersection with one of finitely many attracting neighborhoods, $V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_n}(\tau_n, A_n)$. Moreover, $\mathcal{F}'$ and the $V_{\gamma_i}(\tau_i, A_i)$ can be chosen so that, for each i, τ_i is small and the restriction of $\mathcal{F}'$ to V_{γ_i} is C^0 close to a product foliation $A_{\gamma_i} \times \tau_i$.

In order to make sense of "small" and C^0 close, it is useful to fix a Riemannian metric g on M. We choose a particularly convenient g as follows. Recall that if $\mathcal{F}$ is a $C^{\infty,0}$ foliation of M which is not a fibering of M over S^1, then Corollary 4.12 guarantees the existence of a finite spanning set of pairwise disjoint holonomy neighborhoods for $\mathcal{F}$.

Notation 6.1 Let $\mathcal{F}_0$ be a $C^{\infty,0}$ foliation of M. If $\mathcal{F}_0$ is a fibering of M over S^1, perform a C^0 small Denjoy splitting of $\mathcal{F}_0$ along a fiber and let $\mathcal{F}_1$ denote this new $C^{\infty,0}$ foliation. If $\mathcal{F}_0$ is not a fibering of M over S^1, let $\mathcal{F}_1 = \mathcal{F}_0$.

Let $\{V_1', \dots, V_n'\}$, with $n = r + s$, denote a set of pairwise disjoint holonomy neighborhoods $V_i' = V_{\gamma_i}(\sigma_i, A_i)$ for $\mathcal{F}_1$ satisfying the conditions of Corollary 4.12. Let V' denote the union $V' = \bigcup_i V_i'$. For each i with $1 \le i \le n$, let $R_i' = R_{\gamma_i}(\sigma_i, A_i)$, and set $R' = \bigcup_i R_i$. For each i with $1 \le i \le n$, fix a smooth open neighborhood $N_{R_i'}$ of R_i' in V_i'. Choose each $N_{R_i'}$ small enough that its closure, $\overline{N}_{R_i'}$, is a closed regular neighborhood of R_i'. Let N_R' denote the union of the $N_{R_i'}$.

Apply Lemma 3.11 to isotope $\mathcal{F}_1$ to a $C^{\infty,0}$ foliation $\mathcal{F}_2$ which is C^0 close to $\mathcal{F}_1$ and strongly (V', P)–compatible for some choice of product neighborhood $(P, \mathcal{P})$ of $(V'; N_{R'})$.

Put the product metric on each component $P_i = [-1,1] \times S^1 \times [-1,1]$ of P, as described in Definition 3.12, and let g_0 denote the resulting metric on P. Let $g = g(P)$ be any fixed Riemannian metric on M which restricts to g_0 on P. Since the metric product neighborhoods P_i have pairwise disjoint closures, a partition of unity argument can be used to construct such a metric $g(P)$.

Beginning with a minimal set Λ, a leaf L of Λ and a simple closed curve γ in L which is not homotopically trivial, we show how to introduce an attracting neighborhood, or, sometimes a pair of attracting neighborhoods, about γ via generalized Denjoy blow up. These operations are performed without increasing the number of minimal sets. Since the goal is to produce ϵ–flat holonomy neighborhoods, it may be necessary, as in Theorem 6.2(2) below, to introduce new holonomy neighborhoods to take care of thick collections of parallel compact leaves.

Theorem 6.2 *Let $\mathcal{F}_0$ be a $C^{\infty,0}$ foliation of M. Let $\gamma_1, \dots, \gamma_r, \dots, \gamma_{n=r+s}$, V', $\mathcal{F}_2$, $(P, \mathcal{P})$ and $g = g(P)$ be as given in Notation 6.1. So, in particular, $\mathcal{F}_2$ is not a fibering, V' is spanning for $\mathcal{F}_2$, and $\mathcal{F}_2$ is strongly (V', P)–compatible.*

Fix $\epsilon > 0$. There is a $C^{\infty,0}$ foliation $\mathcal{F}$ that is ϵ C^0 close to $\mathcal{F}_2$, V'–compatible with $\mathcal{F}_2$ and strongly (V', P)–compatible, and a finite set of pairwise disjoint attracting neighborhoods

$$V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_m}(\tau_m, A_m),$$

with $m \ge n$, for $\mathcal{F}$ such that:

(1) $V_{\gamma_i}(\tau_i, A_i) \subset V'_{\gamma_i}(\sigma_i, A_i)$ *for* $1 \le i \le n$.

(2) γ_i *lies in a compact leaf of $\mathcal{F}$ and is isotopic to γ_{j_i} for some $r < j_i \le n$ and* $V_{\gamma_i}(\tau_i, A_i) \subset V'_{\gamma_{j_i}}(\sigma_{j_i}, A_{j_i})$ *for* $n < i \le m$.

(3) *Each $V_{\gamma_i}(\tau_i, A_i)$ is ϵ–flat with respect to $\mathcal{F}$.*

(4) *The restriction of $\mathcal{F}$ to any $V_{\gamma_i}(\tau_i, A_i)$ is ϵ–horizontal.*

(5) *There is a regular neighborhood N_h of $\partial_h V$ such that $\overline{N}_h \cap A = \varnothing$ and the restriction of $\mathcal{F}$ to N_h is a smooth product foliation.*

(6) *Each minimal set of $\mathcal{F}$ has nonempty intersection with the interior of exactly one $V_{\gamma_i}(\tau_i, A_i)$.*

Notice that since $\mathcal{F}$ is strongly (V', P)–compatible, it is strongly (V, P)–compatible. Notice also that condition (6) implies that the collection of attracting neighborhoods $V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_m}(\tau_m, A_m)$ is spanning. Referring back to Notation 6.1, each γ_i is an essential loop in a leaf of a minimal set. Therefore each attracting neighborhood $V_{\gamma_i}(\tau_i, A_i)$ necessarily has nonempty intersection with at least one minimal set. In addition, by the choice of V' and Corollary 4.12, (1) and (2) guarantee that if an attracting set $V_{\gamma_i}(\tau_i, A_i)$, has nonempty intersection with distinct minimal sets, then necessarily $i > r$ and the minimal sets are isotopic compact leaves.

Proof By Lemma 3.16, it is sufficient to prove that there is a $C^{\infty,0}$ foliation $\mathcal{F}$ arbitrarily C^0 close to $\mathcal{F}_2$ such that $\mathcal{F}$ admits a finite spanning set of pairwise disjoint, ϵ–flat attracting neighborhoods $V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_n}(\tau_n, A_n)$ for $\mathcal{F}$ satisfying conditions (1)–(6). Since the holonomy neighborhoods $V'_{\gamma_1}(\sigma_1, A_1), \dots, V'_{\gamma_n}(\sigma_n, A_n)$ are pairwise disjoint, condition (1) will guarantee that the neighborhoods $V_{\gamma_1}(\tau_1, A_1), \dots,$ $V_{\gamma_n}(\tau_n, A_n)$ are pairwise disjoint.

Using Notation 4.11, $\mathcal{F}_2$ has finitely many minimal sets $\Lambda_1, \dots, \Lambda_r$ that have no compact leaves, and at most finitely many isotopy classes $[L_1], \dots, [L_s]$ of compact leaves.

Let L be a compact leaf of $\mathcal{F}_2$. Since $\mathcal{F}_2$ is not a fibering over S^1, either $X(L) = L$ or $X(L) = L \times [0,1]$, with the identification given by a diffeomorphism. Abuse notation and set $\Lambda_{r+j} = X(L_j)$ for $1 \le j \le s$.

Inductively create a V'–compatible, $C^{\infty,0}$ foliation $\mathcal{F}^k$ arbitrarily C^0 close to $\mathcal{F}_2$ and a pairwise disjoint collection of attracting neighborhoods

$$\mathcal{V}_k = \{V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_k}(\tau_{n_k}, A_{n_k})\}$$

for $\mathcal{F}^k$ satisfying conditions (1)–(4). Let $N_0(\mathcal{F}^k)$ be the number of minimal sets Λ_j, with $1 \le j \le r$, that do not intersect the interior of some $V_{\gamma_i}(\tau_i, A_i)$. Let $N_1(\mathcal{F}^k)$ be the number of isotopy classes of compact leaves of $\mathcal{F}^k$ containing leaves that do not intersect the interior of some $V_{\gamma_i}(\tau_i, A_i)$. Set $N(\mathcal{F}^k) = N_0(\mathcal{F}^k) + N_1(\mathcal{F}^k)$.

The construction of the desired set of attracting neighborhoods consists of finding, or creating, attracting ϵ–flat neighborhoods that decrease $N(\mathcal{F}^k)$. These neighborhoods are created by performing, as necessary, generalized Denjoy blow ups very close to leaves L in the sets Λ_i for $1 \le i \le r+s$, and by Theorem 5.2, these blow ups can be chosen arbitrarily C^0 close to $\mathcal{F}$. The construction takes different forms depending on properties of L and is carried out in Propositions 6.5, 6.7, and 6.8.

At the k^{th} stage in the induction, one or more ϵ–flat attracting neighborhoods are added to $\mathcal{V}_k$ to yield $\mathcal{V}_{k+1}$.

To complete the proof of Theorem 6.2, it suffices to establish Propositions 6.5, 6.7, and 6.8. We now do so.

Let γ be an oriented essential simple closed curve in a leaf L contained in a minimal set Λ of $\mathcal{F}^k$ that contributes to $N(\mathcal{F}^k)$. To simplify notation, let $\mathcal{F} = \mathcal{F}^k$ at the k^{th} step. We begin by considering a holonomy neighborhood $V_\gamma(\tau, A) \subset P$. Using the notation of Section 2, $p \in \gamma$ and τ and σ are transversals through p such that $h_\gamma\colon \tau \to \sigma$ is a holonomy map for $\mathcal{F}$ along γ. Notice that if τ is chosen to be sufficiently small, then $V_\gamma(\tau, A)$ is ϵ–flat.

Lemma 6.3 *Let $V_\gamma(\tau, A)$ be a holonomy neighborhood of γ. One of the following is true:*

(1) *There is a choice of $\tau' \subset \tau$ such that one of $V_\gamma(\tau', A)$ and $V_{-\gamma}(\tau', A)$ is an attracting neighborhood, and including the attracting neighborhood decreases $N(\mathcal{F})$.*

(2) *There is a choice of $\tau' \subset \tau$ such that, after performing a generalized Denjoy blow up along L arbitrarily C^0 close to the identity and compatible with $V_\gamma(\tau, A)$, $V_\gamma(\tau, A)$ is the union of two attracting neighborhoods V_{γ_0} and V_{γ_1}, where γ_0 and γ_1 are the copies of γ obtained by the splitting, chosen with opposite orientations as determined by the form of V_γ. Including these attracting neighborhoods decreases $N(\mathcal{F})$.*

(3) *The holonomy map h_γ is the identity when restricted to at least one of the components of $\tau \backslash \{p\}$.*

Proof Consider the holonomy map $h_\gamma\colon \tau \to \sigma$. If there are intervals in each component of $\tau \backslash \{p\}$ on which h_γ is strictly monotonic, then (1) or (2) must hold. Otherwise, (3) holds. □

By the lemma, it is enough to consider the case that γ is essential in L, and h_γ is the identity when restricted to at least one of the components of $\tau \backslash \{p\}$. Notice that since $\mathcal{F}$ is taut, L is π_1–injective, and hence γ is homotopically nontrivial in M.

Identify (τ, p) with $([-u, v], 0)$. For $t \in [-u, v]$, let L_t denote the leaf of $\mathcal{F}$ intersecting τ at height t. Call a compact leaf of $\mathcal{F}$ *isolated* if there is an $\mathcal{F}$ saturated open neighborhood of L containing no other compact leaf. For simplicity of exposition, we will consider first the case that L is either noncompact or else compact but isolated.

As a first step towards building an attracting neighborhood, we show that without increasing $N(\mathcal{F})$, generalized Denjoy blow ups can be used to replace holonomy that is the identity on one side of γ with monotone holonomy.

Lemma 6.4 *Let γ be an oriented essential simple closed curve in an isolated compact leaf L such that h_γ is the identity on $[0, v]$. Then there exists a generalized Denjoy blow up of $\mathcal{F}$ to $\mathcal{F}'$ such that, in $\mathcal{F}'$, h_γ is strictly monotone on a nondegenerate subinterval of $[0, v]$. Moreover, $N(\mathcal{F}') \leq N(\mathcal{F})$, and h_γ may be created to be either monotone increasing or decreasing on this subinterval.*

Proof By Corollary 4.10, there exists $w \in (0, v]$ small enough to guarantee that the transversal $(0, w)$ is disjoint from the minimal sets of $\mathcal{F}$. Let $t \in (0, w)$, and consider the leaf L_t.

Let γ_t be a parallel copy of γ lying in L_t and passing through t. Perform generalized Denjoy blow up $\mathcal{F}(L_t, \gamma_t)$, arbitrarily C^0 close to the identity and compatible with V_γ, to introduce spiraling about $\gamma_t^{\pm 1}$, the two copies of γ_t introduced. This introduces a nondegenerate interval in $(0, v)$ on which h_γ is strictly monotonic. Moreover, the type of monotonicity, increasing or decreasing, can be chosen. Since $\overline{L}_t$ is not minimal, Lemma 5.6 guarantees that no new minimal sets are introduced. □

Proposition 6.5 *Let γ be an oriented essential simple closed curve in an isolated compact leaf L. If h_γ is the identity on at least one of $[-u, 0]$ and $[0, v]$, then there exists a generalized Denjoy blow up $\mathcal{F}'$ of $\mathcal{F}$ that creates no new compact leaves, has $N(\mathcal{F}') \leq N(\mathcal{F})$ and for which there exists an attracting holonomy neighborhood containing L.*

Proof Apply Lemma 6.4 once or twice, as necessary, and let $\mathcal{F}'$ be the result of doing the generalized Denjoy blow up or blow ups required to make h_γ strictly monotone on nondegenerate subintervals of each of $[-u, 0]$ and $[0, v]$. The monotonicity can be chosen so that h_γ is either attracting on both subintervals or repelling on both subintervals. Choose $\tau' \subset \tau$ to be the smallest closed interval containing both subintervals. Including $V_\gamma(\tau')$ with the collection of holonomy neighborhoods may not create intersections with every leaf isotopic to L, that is, it may not decrease $N_1(\mathcal{F})$, but it keeps $N(\mathcal{F}') \leq N(\mathcal{F})$. □

Lemma 6.6 *Let γ be an oriented essential simple closed curve in a noncompact leaf L of a minimal set such that h_γ is the identity on $[0,v]$. Then there exists a generalized Denjoy blow up of $\mathcal{F}$ to $\mathcal{F}'$ such that, in $\mathcal{F}'$, h_γ is strictly monotone on a nondegenerate subinterval of $[0,v]$. Moreover, $N(\mathcal{F}') \le N(\mathcal{F})$, and h_γ may be created to be either monotone increasing or decreasing on this subinterval.*

Proof By Corollary 4.10, there exists $w \in (0,v]$ small enough to guarantee that the transversal $(0,w)$ is disjoint from the minimal sets of $\mathcal{F}$. Hence the proof of Lemma 6.4 works in this case as well. □

Proposition 6.7 *Let γ be an oriented essential simple closed curve in a noncompact leaf L of a minimal set that contributes to $N(\mathcal{F})$. If h_γ is the identity on at least one of $[-u,0]$ and $[0,v]$, then there exists a generalized Denjoy blow up $\mathcal{F}'$ of $\mathcal{F}$ that creates no new compact leaves, satisfies $N(\mathcal{F}') < N(\mathcal{F})$, and for which there exists an attracting holonomy neighborhood containing L.*

Proof The proof is similar to the proof of Proposition 6.5, instead requiring one or two applications of Lemma 6.6. The result is a strict decrease in $N_0(\mathcal{F})$, thereby forcing $N(\mathcal{F}') < N(\mathcal{F})$. □

At this point, attracting neighborhoods have been constructed which intersect every minimal set consisting of noncompact leaves (with at least one non-$\mathbb{R}^2$ leaf). There remain minimal sets that consist of a single compact leaf. Lemma 6.4 shows how to construct a holonomy neighborhood that will contain such a leaf. The next proposition shows how to deal with possibly infinite families of isotopic compact leaves.

Let L be a compact leaf of $\mathcal{F}$. Recall that $X(L)$ denotes the minimal $\mathcal{F}$ saturated closed submanifold of M containing all leaves isotopic to L, and that $\mathcal{F}$ is not a fibering over S^1. So either $X(L) = L$ or $X(L)$ is diffeomorphic to $L \times [0,1]$. We now restrict attention to the remaining case, that $X(L) \cong L \times [0,1]$. Notice that leaves of $\mathcal{F}$ that are contained in $X(L)$ are not required to be homeomorphic to L.

Proposition 6.8 *Let L be a compact leaf such that $X(L) \cong L \times [0,1]$, and let γ be an oriented essential simple closed curve in L. There is a $C^{\infty,0}$ foliation $\mathcal{F}'$ C^0 close to $\mathcal{F}$ such that all surfaces of $\mathcal{F}'$ isotopic to L are covered by finitely many pairwise disjoint, ϵ–flat, attracting neighborhoods that are disjoint from all minimal sets not isotopic to L.*

Moreover, such a foliation $\mathcal{F}'$ can be obtained from $\mathcal{F}$ by performing a finite number of generalized Denjoy blow ups along leaves of $\mathcal{F}$ isotopic to L, and the attracting neighborhoods can be chosen of the form $V_{\gamma_t}(\tau_t, A_t)$, where L_t is a leaf of $\mathcal{F}$ isotopic to L and (L, A, γ) is isotopic to (L_t, A_t, γ_t).

Proof Let A be a smooth regular neighborhood of γ in L and choose $p \in \gamma$. Use a diffeomorphism to identify $X(L)$ with $L \times [0,1]$. Choose the diffeomorphism so that it agrees with the product structure of P restricted to $X(L)$. In particular, flow lines of Φ map to the vertical fibers $\{x\} \times [0,1]$. Let τ_0 denote the vertical fiber $\{p\} \times [0,1]$. Let τ denote a closed flow segment of Φ containing τ_0 in its interior and such that τ is disjoint from all minimal sets of $\mathcal{F}$ not isotopic to L.

Let L_t, $t \in \mathcal{T}$, be a listing of the leaves of $\mathcal{F}$ isotopic to L, where $L_t \cap \tau = \{(p,t)\}$. Let $\gamma_t = L_t \cap (\gamma \times [0,1])$ and $A_t = L_t \cap (A \times [0,1])$. Notice that since lengths can only shorten under projection, the minimum length and width of A_t is bounded below by the minimum length and width of A. For all $t \in \mathcal{T}$, choose an $\frac{\epsilon}{2}$–flat holonomy neighborhood $V_{\gamma_t}(\tau_t, A_t)$, where $\tau_t \subset \tau$.

Since the union $\bigcup_{t \in \mathcal{T}} A_t$ is compact (since closed), the interiors of finitely many of the $V_{\gamma_t}(\tau_t, A_t)$ cover $\bigcup_{t \in \mathcal{T}} A_t$. Choose $t_1, \dots, t_k \in \mathcal{T}$ so that the interiors of the $V_{\gamma_{t_i}}(\tau_{t_i}, A_{t_i})$ for $1 \le i \le k$ cover $\bigcup_{t \in \mathcal{T}} A_t$. Set $\tau' = \tau_{t_1} \cup \cdots \cup \tau_{t_k}$. Choose closed subintervals of the τ_{t_i} and relabel as necessary so that $\tau' = \tau'_{t_1} \cup \cdots \cup \tau'_{t_k}$, where the interiors of the τ'_{t_i} are pairwise disjoint and only successive ones can have nonempty intersection. Then the $V_{\gamma_{t_i}}(\tau'_{t_i}, A_{t_i})$ cover, and after performing finitely many C^0 small blow ups along compact leaves of the form L_a, with $a \in \partial \tau'_{t_i}$, and then extending the τ'_{t_i} slightly as necessary, we have the claimed finite set of pairwise disjoint, ϵ–flat, attracting neighborhoods, disjoint from all minimal sets not isotopic to L. □

By Corollary 4.8, only finitely many applications of Propositions 6.5 and 6.8 generate attracting holonomy neighborhoods that intersect all compact leaves. Proposition 6.7 takes care of the rest of the cases, and therefore completes the proof of Theorem 6.2. □

7 Approximation in an attracting neighborhood

In this section we show (Theorem 7.2) how the tangent plane field of the restriction of a $C^{\infty,0}$ foliation $\mathcal{F}_0$ to a sufficiently small attracting neighborhood $V_\gamma(\tau, A)$ can be C^0 approximated by a contact structure. The idea is that restriction of the foliation $\mathcal{F}_0 \cap V_\gamma(\tau, A)$ can be analyzed by cutting $V_\gamma(\tau, A)$ open along $R_\gamma(\tau, A)$ and considering the resulting foliation on $Q_\gamma(\tau, A)$. Much of the foliation data is then encoded in monodromy maps along the four vertical sides of $Q_\gamma(\tau, A)$.

This portion of the paper is quite different than the more analytic strategy of [8, 2.5.1–2.5.3], in which they approximate forms which define foliations by a contact form in the neighborhood of nontrivial linear monodromy. Our strategy is guided by the relationship between a contact structure on the boundary of a vertical cylinder about the

z–axis in the radial model of the standard contact structure on $\mathbb{R}^3$ and the foliations by horizontal disks. Thus we build foliations on the vertical boundaries of $Q_\gamma(\tau, A)$ in Lemma 7.5 that can serve as the characteristic foliations of a contact structure.

When $V_\gamma(\tau, A)$ is attracting, these characteristic foliations can be chosen both to be compatible with gluing $Q_\gamma(\tau, A)$ to form $V_\gamma(\tau, A)$, and to dominate the given foliation on the complement of $V_\gamma(\tau, A)$. Recall that if two curves intersect in $\partial_v X$ for some codimension-zero submanifold X with piecewise horizontal and vertical boundary, the curve with greater slope, when viewed from outside of X, is said to *strictly* dominate the other curve.

A key result is Proposition 7.6, in which we show that the restriction of $\mathcal{F}$ to $\partial_v Q$ can be approximated by a smooth foliation on $\partial_v Q_\gamma(\tau, A)$ that has decreasing monodromy. By Corollary 7.10, such foliations can be smoothly extended to a disk foliation on $Q_\gamma(\tau, A)$, and hence $V_\gamma(\tau, A)$, using the original $C^{\infty,0}$ foliation as a guide.

To simplify the exposition, we will fix a Riemannian metric on M as described in Notation 6.1.

Notation 7.1 We say one object is $O(\epsilon)$ close to another if for some constant K independent of the two objects, the objects are $K\epsilon$ close.

Theorem 7.2 *Let $\mathcal{F}_0$ be a $C^{\infty,0}$ foliation of M, and let $(P, \mathcal{P})$ and $g = g(P)$ be given as in Notation 6.1. Fix $\epsilon > 0$ and let V, N_h and $\mathcal{F}$ be given as in Theorem 6.2.*

Then there are a regular neighborhood $N_v \subset V$ of the vertical edges of ∂Q in V and smooth plane fields $\xi_V^\pm$ defined on V satisfying

- *ξ_V^+ is positive and ξ_V^- is negative,*
- *$\xi_V^\pm = T\mathcal{F}$ on $\overline{N_h \cup N_v}$ and is contact at all other points of V,*
- *ξ_V^+ dominates $\mathcal{F}$ along $\partial_v V$, with the domination strict outside $\overline{N_h \cup N_v}$,*
- *ξ_V^- is dominated by $\mathcal{F}$ along $\partial_v V$, with the domination strict outside $\overline{N_h \cup N_v}$,*
- *each of $\xi_V^\pm$ is positively transverse to Φ, and*
- *each of $\xi_V^\pm$ is $O(\epsilon)$ C^0 close to $T\mathcal{F}$ on V.*

The proof of this theorem will occupy the rest of this section. By symmetry, it will suffice to establish the existence of ξ^+.

It suffices to consider the case $n = 1$; so, to simplify notation, write $V = V_\gamma(\tau, A)$, with metric product neighborhood $(P, \mathcal{P})$. Recall that the metric product neighborhood

$(P, \mathcal{P})$ has product metric and horizontal product foliation $\mathcal{P}$ induced by the identification $P \cong [-1,1] \times S^1 \times [-1,1]$. Use this identification to view $V = V_\gamma(\tau, A) \subset P$ as a subset $V \subset [-1,1] \times S^1 \times [-1,1]$. Notice that $\partial_v V \subset \partial_v P$. Let N_R denote the open neighborhood of $R_\gamma(\tau, A)$ in V given by the intersection of V with $[-1,1] \times \left(\left[\frac{1}{2}, 1\right] \cup \left[-1, -\frac{1}{2}\right]\right) \times [-1,1]$. Recall that $\mathcal{F} = \mathcal{P}$ on N_R. See Figure 1.

Eventually, we will further constrain N_v, but for now, let $N_v \subset N_R$ be any regular neighborhood of the vertical edges of ∂Q, and set $N = N_h \cup N_v$.

Write $Q = Q_\gamma(\tau, A)$, and let

$$\pi \colon Q \to V$$

denote the quotient map which reverses the splitting of V along $R_\gamma(\tau, A)$. Viewing S^1 as the quotient $[0,1]/\{0 \sim 1\}$, the identification $V \subset [-1,1] \times S^1 \times [-1,1]$ induces an identification

$$Q \subset [-1,1] \times [0,1] \times [-1,1] \subset \mathbb{R}^3,$$

with $\partial_v Q \subset \partial_v([-1,1] \times [0,1] \times [-1,1])$. We will abuse notation and let N_h, N_v, N, and N_R also denote their pullbacks to Q under $\pi \colon Q \to V$. Similarly, we let $\mathcal{F}$ denote the pullback $\pi^{-1}(\mathcal{F} \cap V)$ when this meaning is clear from context.

Our goal is to construct a smooth positive confoliation on V satisfying the conditions of Theorem 7.2. We will do this by defining a smooth positive confoliation ξ^+ on Q which smoothly glues, via $\pi \colon Q \to V$, to a smooth confoliation on V. As a first step, we will define a smooth foliation on $\partial_v Q$ which will serve as the characteristic foliation of the contact structure ξ^+. This characteristic foliation will be closely related to the restriction of $\mathcal{F}$ to $\partial_v Q$. The following proposition will be used to make the transition from continuous to smooth structures.

Proposition 7.3 *Let Φ, N_h, N_v, N and Q be as given above, and let $\kappa \in \{\pm 1\}$.*

Let X denote either a vertical face or a union of three vertical faces of Q, and denote the components of $\partial_v X$ by σ and τ. Let $\mathcal{G}_0$ be a $C^{\infty,0}$ foliation of X which is everywhere transverse to Φ, satisfies $\mathcal{G}_0 = \mathcal{P}$ on $\overline{N}_v \cap X$, and is smooth when restricted to $\overline{N} \cap X$. Let $G_0 \colon \sigma \to \tau$ be the holonomy map defined by following leaves of $\mathcal{G}_0$ across X, beginning in σ and ending in τ. Let $\varepsilon \colon \sigma \to \mathbb{R}$ be a continuous function satisfying $\varepsilon(z) \geq 0$, with equality if and only if $z \in \partial\sigma$.

Then there is a C^∞ foliation $\mathcal{G}$ on X, with holonomy across X given by the holonomy map $G \colon \sigma \to \tau$, such that

(1) *$\mathcal{G}$ is positively transverse to Φ,*

(2) *$\mathcal{G}$ is arbitrarily C^0 close to $\mathcal{G}_0$,*

(3) $\mathcal{G} = \mathcal{G}_0$ on $\overline{N}$,

(4) *$T\mathcal{G}$ dominates (resp. is dominated by) $T\mathcal{G}_0$, with the domination strict outside $\overline{N}$, if $\kappa = 1$ (resp. $\kappa = -1$), and*

(5) $|G(z) - G_0(z)| \leq \varepsilon(z)$.

Proof It suffices to consider the case that $\kappa = 1$. Begin by considering the case that X lies in the vertical plane $x = 1$.

At each point $(1, y, z)$ of X let $f(y,z)$ be the continuous function such that the line field $\partial_y + f(y,z)\partial_z$ is tangent to $\mathcal{G}_0$. Let $g\colon X \to \mathbb{R}$ be a smooth function arbitrarily close to f and such that $g(y,z) \geq f(y,z)$, with equality if and only if $(1, y, z) \in \overline{N}$. Let $\mathcal{G}$ be the foliation given by the integral curves of the flow tangent to $\partial_y + g(y,z)\partial_z$. Denote the foliation determined by g by $\mathcal{G}$. Certainly, conclusions (1)–(4), are satisfied by $\mathcal{G}$.

Let ε_0 be one half of the minimum value of ε on $\sigma - \mathrm{int}(N(\partial_h V))$, where $N(\partial_h V)$ is the portion of N corresponding to a regular neighborhood of $\partial_h V$.

Let $g_n\colon X \to \mathbb{R}$ be a sequence of smooth functions, each determining a foliation satisfying (1)–(4) and with limit f. By the smoothness of solutions to ODEs, there is a number m so that by setting $g = g_n$ for any $n > m$, and letting $\mathcal{G}$ be the foliation determined by g, the corresponding holonomy map $G\colon \sigma \to \tau$ satisfies $G(z) - G_0(z) < \epsilon_0$. Thus (5) holds for $z \notin \sigma - \mathrm{int}(N(\partial_h V))$. On $\mathrm{int}(N(\partial_h V))$, $g = f$, and hence $G(z) = G_0(z)$ for all $z \in \sigma \cap \mathrm{int}(N(\partial_h V))$. Thus (5) holds for all $z \in \sigma$.

Finally, we consider the remaining possibilities for X. Certainly the proof as given applies to any single face of Q. In the case of a union of three faces of Q, isometrically flatten out the union so that it lies in a single plane to see that the proof as given applies. □

Next we introduce some useful notation. Label the vertical faces of Q by B, C, D and E, where $\pi(B) \subset \pi(D) = R_\gamma(\tau, A)$, and the sequence B, C, D, E is a listing of the faces in counterclockwise order about $\partial_v Q$. These faces are illustrated in Figure 2.

Projecting these labels to V, we will abuse notation when convenient by considering B to be a subset of D. Notice that each of $B \cap \overline{N}_h$, $C \cap \overline{N}_h$ and $E \cap \overline{N}_h$ consists of two components, whereas $D \cap \overline{N}_h$ consists of four components.

Also, label the vertical edges of Q

$$\tau_{BC} = B \cap C, \quad \tau_{CD} = C \cap D, \quad \tau_{DE} = D \cap E \quad \text{and} \quad \tau_{EB} = E \cap B.$$

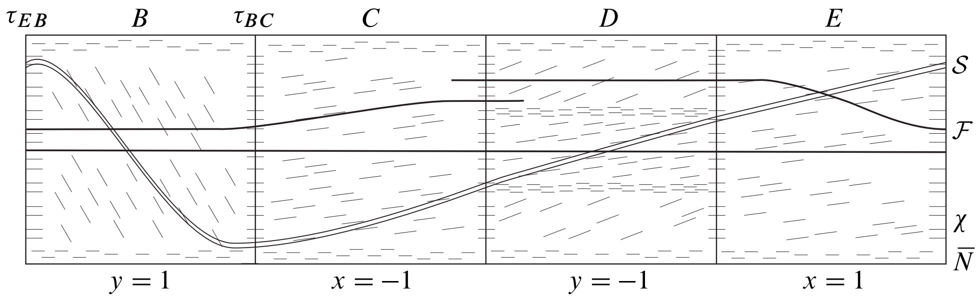


Figure 2: The vector field χ is shown as short dashes. Two leaves of $\mathcal{F}$ are shown, one of which corresponds to the annulus leaf A. A pair of circle leaves of $\mathcal{S}$ is shown dominating χ.

Again, projecting these labels to V, we will abuse notation when convenient by considering τ_{EB} to be a subset of τ_{DE} and τ_{BC} to be a subset of τ_{CD}.

Now consider the (continuous) holonomy maps of $\mathcal{F}$ across the 2–dimensional faces of $\partial_v Q$,

$$\tau_{EB} \xrightarrow{F_B} \tau_{BC} \xrightarrow{F_C} \tau_{CD} \xrightarrow{F_D} \tau_{DE} \xrightarrow{F_E} \tau_{EB}.$$

Since $\mathcal{F}$ is smooth on N_h, the restrictions of F_B, F_C, F_D, and F_E to $\tau_{BC} \cap N_h$, $\tau_{CD} \cap N_h$, $\tau_{DE} \cap N_h$ and $\tau_{EB} \cap N_h$, respectively, are smooth functions.

Recall that $\mathcal{F} = \mathcal{P}$ along $R_\gamma(\tau, A)$. Therefore, using the identifications

$$V \subset [-1,1] \times S^1 \times [-1,1] \quad \text{and} \quad Q \subset [-1,1] \times [0,1] \times [-1,1] \subset \mathbb{R}^3,$$

we have $B \cong (A \cap B) \times \tau_{EB}$ with the leaves of $\mathcal{F} \cap B$ identified with the leaves $(A \cap B) \times \{t\}$ and $D \cong (A \cap D) \times \tau_{CD}$ with the leaves of $\mathcal{F} \cap D$ identified with the leaves $(A \cap D) \times \{t\}$. In particular, $\tau_{EB} \cong \tau_{BC}$ and $\tau_{CD} \cong \tau_{DE}$, and under these smooth identifications, the maps F_B and F_D are automatically identity maps.

In the next corollary, we show that the restrictions of $\mathcal{F}$ to C and E can be approximated by smooth foliations which dominate.

Corollary 7.4 *There are smooth foliations χ_C on C and χ_E on E satisfying:*

(1) *Each of χ_C and χ_E is positively transverse to Φ.*

(2) *χ_C (resp. χ_E) is arbitrarily C^0 close to $\mathcal{F}$ on C (resp. on E).*

(3) *$\chi_C = \mathcal{F}$ and $\chi_E = \mathcal{F}$ on $\overline{N}$.*

(4) *χ_C (resp. χ_E) dominates $\mathcal{F}$ on C (resp. E), with the domination strict outside $\overline{N}$.*

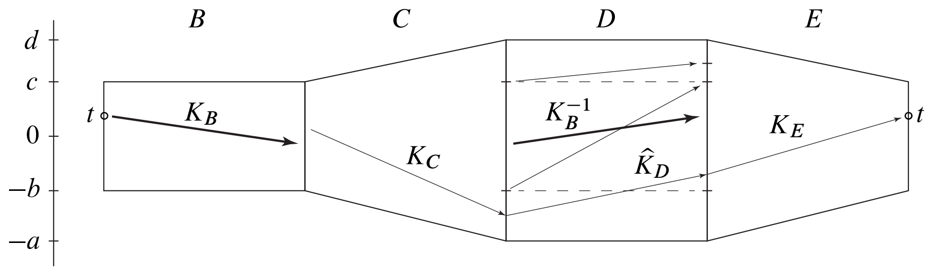


Figure 3

Proof Let $\mathcal{G}_0$ denote the restriction of $\mathcal{F}$ to X, where X is either C or E, and set $\kappa = 1$. Apply Proposition 7.3 with $X = C$ to obtain χ_C, and with $X = E$ to obtain χ_E. □

In the next proposition, we show that the restrictions of $\mathcal{F}$ to B and D can be approximated by smooth foliations which dominate along D and are dominated along B. Moreover, the union of these foliations with the smooth foliations of the preceding corollary gives a foliation realizing decreasing monodromy about $\partial_v Q$.

Lemma 7.5 *Let χ_C and χ_E be as guaranteed in Corollary 7.4, with monodromy maps $K_C\colon \tau_{BC} \to \tau_{CD}$ and $K_E\colon \tau_{DE} \to \tau_{EB}$, respectively. There are maps $K_B\colon \tau_{EB} \to \tau_{BC}$ and $K_D\colon \tau_{CD} \to \tau_{DE}$, respectively, satisfying*

(1) $K_D(t) = K_B^{-1}(t)$ *for all* $t \in \tau_{EB} \subset \tau_{DE}$,

(2) $K_B = F_B$ *and* $K_D = F_D$ *when restricted to* $\overline{N}$,

(3) $K_B(t) \leq F_B(t)$ *for all* $t \in \tau_{EB}$,

(4) $K_D(t) \geq F_D(t)$ *for all* $t \in \tau_{CD}$,

(5) $K_E K_D K_C K_B(t) \leq t$ *for all* $t \in \tau_{EB}$.

where, for each inequality, equality holds if and only if $t \in \overline{N}$.

Proof Recall that the metric product neighborhood $(P, \mathcal{P})$ has product metric and horizontal product foliation $\mathcal{P}$ induced by the identification $P \cong [-1,1]\times S^1\times[-1,1]$. Use this identification to view $V_\gamma(\tau, A) \subset P$ as a subset $V_\gamma(\tau, A) \subset [-1,1]\times S^1\times[-1,1]$ and therefore Q as a subset of $[-1,1]\times[0,1]\times[-1,1]$. This identification induces $\tau_{CD} \cong \tau_{DE} \cong [-a, d]$ and $\tau_{EB} \cong \tau_{BC} \cong [-b, c]$ for some $-1 \leq -a < -b < 0 < c < d \leq 1$. Choose e so that $c < e < d$ and $(c, e]$ is disjoint from $\overline{N}$.

As a first step towards choosing K_D, let $\widehat{K}_D\colon [-a, d] \to [-a, d]$ be any orientation-preserving diffeomorphism that maps $[-b, c] \to [c, e]$ and satisfies $F_D(t) \leq \widehat{K}_D(t)$ for all $t \in \tau_{CD}$, with equality if and only if $t \in \overline{N}$.

Next choose K_B satisfying (2), $K_B(t) \leq (K_E \hat{K}_D K_C)^{-1}(t)$ and (3). Finally, choose K_D satisfying (2) to agree with K_B^{-1} on $[-b, c]$ and such that $F_D(t) \leq K_D(t) \leq \hat{K}_D(t)$ for all $t \in [-a, -b] \cup [c, d]$. Thus, $K_D(t) \leq \hat{K}_D(t)$ for all $t \in [-a, d]$, and K_D satisfies (4). Since $K_B(t) < (K_E \hat{K}_D K_C)^{-1}(t) \leq (K_E K_D K_C)^{-1}(t)$, condition (5) is satisfied. Some of these relationships are shown in Figure 3. □

Proposition 7.6 *Let χ_C and χ_E be as guaranteed in Corollary 7.4, with monodromy maps $K_C\colon \tau_{BC} \to \tau_{CD}$ and $K_E\colon \tau_{DE} \to \tau_{EB}$, respectively. Let $K_B\colon \tau_{EB} \to \tau_{BC}$ and $K_D\colon \tau_{CD} \to \tau_{DE}$ be the maps as constructed in Lemma 7.5. There are smooth foliations χ_B and χ_D, with holonomy maps K_B and K_D, respectively, satisfying*

(1) *$\chi_B = \mathcal{F}$ and $\chi_D = \mathcal{F}$ when restricted to $\overline{N}$,*

(2) *$\mathcal{F}$ is dominated by χ_B along B, with the domination strict outside $\overline{N}$,*

(3) *$\mathcal{F}$ dominates χ_D, with the domination strict outside $\overline{N}$,*

(4) *χ_B and χ_D agree where identified by $\pi\colon Q \to V$,*

(5) *the foliation on $\partial_v Q$ defined by the union*

$$\chi = \chi_B \cup \chi_C \cup \chi_D \cup \chi_E$$

is smooth, and

(6) *each of χ_B and χ_D are $O(\epsilon)$ C^0 close to the restriction of $\mathcal{F}$,*

where, for each inequality, equality holds if and only if $t \in \overline{N}$.

Proof Use K_B and K_D to construct χ_B and χ_D, respectively. By construction, $\mathcal{F} = \mathcal{P}$ in a neighborhood of $R_\gamma(\tau, A)$, and hence $\mathcal{F}$ is equal to $\mathcal{P}$ in a neighborhood of B and D in Q. So $\mathcal{F}$ restricts to foliations of $B = [-1, 1] \times \{1\} \times [-a, d]$ and $D = [-1, 1] \times \{-1\} \times [-a, d]$, respectively, by horizontal straight line segments.

The foliation of B by line segments with endpoints $(1, 1, z)$ and $(-1, 1, K_B(z))$ has leaves given by $(1-\rho)(1, 1, z) + \rho(-1, 1, K_B(z))$ with $\rho \in [0, 1]$ and has the desired monodromy $K_B(z)$. To guarantee (1) holds, replace ρ by $\rho(t)$ where $\rho\colon [-1, 1] \to [0, 1]$ is a smooth damping function chosen so that $\rho^{-1}(0) \cup \rho^{-1}(1)$ corresponds to $\overline{N} \cap B$. Let χ_B be this damped linear foliation. Similarly define χ_D to be the damped linear function with holonomy $K_D(z)$ damped so that leaves are horizontal exactly on $\overline{N} \cap D$.

The foliations χ_B and χ_D are smooth since the corresponding holonomy maps K_B and K_D are smooth. In addition, since the foliations χ_B, χ_C, χ_D and χ_E are horizontal in a neighborhood of $\partial_v Q^{(1)}$, they glue together to give a smooth foliation of $\partial_v Q$. Moreover, since V is ϵ–flat, the straight lines used to create χ_D have slope between $\pm\epsilon$, thus after damping, $T\chi_D$ has slope bounded in absolute value by 2ϵ. Since $\mathcal{F}$ was chosen to satisfy the conclusions of Theorem 6.2, so in particular $\mathcal{F}$ is ϵ–horizontal, it follows that $T\chi_D$ and $T\chi_B$ are $O(\epsilon)$ close to $T\mathcal{F}$. □

Notice that χ is the characteristic foliation of a smooth 2–plane field along $\partial_v Q$ defined as follows. At each point p of $\partial_v Q$, let ξ_p denote the 2–plane perpendicular to $\partial_v Q$ which contains $T_p \chi$. We will show that this 2–plane field extends to a smooth confoliation on V which stays close to $\mathcal{F}$. The first step in constructing this extension is to build a circle foliation dominated by χ which in turn bounds a disk foliation of Q.

Corollary 7.7 *Let K_B, K_C, K_D, K_E and χ be as given in Proposition 7.6. There is a smooth foliation $\mathcal{S}$ of $\partial_v Q$ by circles (with corners along $\partial_v Q^{(1)}$) such that*

(1) *χ dominates $T\mathcal{S}$, with the domination strict outside $\overline{N}_h$,*

(2) *$\mathcal{S} = \chi = \mathcal{F}$ on $\overline{N}_h$, and*

(3) *$\mathcal{S}$ is $O(\epsilon)$ C^0 close to each of $\mathcal{F}$ and χ.*

Proof Begin by constructing $\mathcal{S}$ on X, where X is the union of the faces C, D and E of Q. Let $\mathcal{G}_0$ denote the restriction of χ to X, and let G_0 denote the holonomy map $G_0\colon \tau_{BC} \to \tau_{EB}$ given by the composition $G_0 = K_E K_D K_C$. For each $z \in \tau_{BC} = [-b, c]$, let $\varepsilon(z) = K_B^{-1} z - G_0(z)$. Apply Proposition 7.3 to get a foliation $\mathcal{G}$ on X satisfying

(i) $\mathcal{G}$ is positively transverse to Φ,

(ii) $\mathcal{G}$ is ϵ C^0 close to $\mathcal{G}_0$,

(iii) $\mathcal{G} = \mathcal{G}_0$ on $\overline{N}$,

(iv) $T\mathcal{G}$ dominates $T\mathcal{G}_0$, with the domination strict outside $\overline{N}$, and

(v) $G(z) - G_0(z) \le \varepsilon(z)$.

Since $\mathcal{G}_0$ is $O(\epsilon)$ close to $\mathcal{F}$ by Corollary 7.4 and Proposition 7.6, and $\mathcal{G}$ is ϵ close to $\mathcal{G}_0$, $\mathcal{G}$ is $O(\epsilon)$ close to $\mathcal{F}$. Set $\mathcal{S} = \mathcal{G}$ on X.

Next, construct $\mathcal{S}$ on the remaining vertical face, B. Take advantage of the fact that, as described in the proof of Proposition 7.6, χ restricted to B consists of damped straight line segments. Let $\mathcal{S}$ consist of a smooth family of similarly damped line segments from $G^{-1}(z) \in \tau_{EB}$ to $z \in \tau_{BC}$. Condition (v) guarantees that $G(z) \le K_B^{-1}(z)$, with equality only in the collar of $\{-b, c\}$ in $[-b, c]$ determined by $\overline{N}$, and hence line segments from $G(z)$ to z are steeper than line segments $K_B^{-1}(z)$ to z for z outside this collar of $\{-b, c\}$. This property is preserved under suitably chosen damping. Notice that $\mathcal{S}$ lies within ϵ of χ in B.

The foliation $\mathcal{S}$ is a foliation by circles since it has trivial monodromy about $\partial_v Q$. By construction, $\mathcal{S}$ is dominated by χ, with the domination strict exactly outside $\overline{N}$. Moreover, $\mathcal{S}$ lies within ϵ of χ, and hence $O(\epsilon)$ close to $\mathcal{F}$. □

Next we consider a smooth cylinder Q' which lies nicely in Q. We do this as follows.

Recall that in our preferred coordinates (namely, the ones inherited from P), Q is an ϵ–flat, x–invariant subset of $[-1,1]^3$ which is diffeomorphic to a cube and satisfies $[-1,1]\times[-1,1]\times\{0\}\subset Q$. In particular, $\mathcal{F}$ agrees with the horizontal foliation $\mathcal{P}$ at all points (x,y,z) of Q for which $y\notin\left[-\frac{1}{2},\frac{1}{2}\right]$.

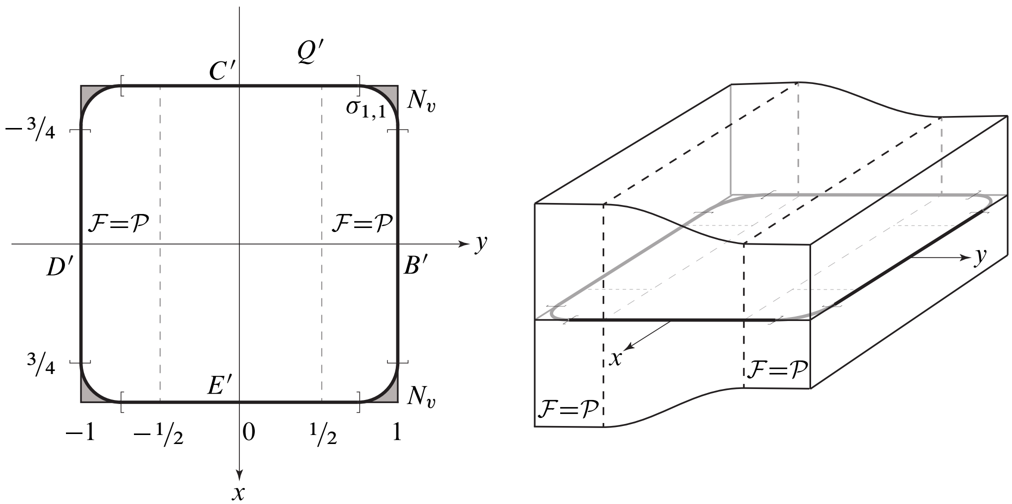


Figure 4: $\Delta\subset[-1,1]^2$ and Q

Let Δ be the smooth disk embedded in the square $[-1,1]^2$ with smooth boundary

$$\partial\Delta = B'\cup\sigma_{-1,1}\cup C'\cup\sigma_{-1,-1}\cup D'\cup\sigma_{1,-1}\cup E'\cup\sigma_{1,1},$$

where $B'=\left[-\frac{3}{4},\frac{3}{4}\right]\times\{1\}$, $C'=\{-1\}\times\left[-\frac{3}{4},\frac{3}{4}\right]$, $D'=\left[-\frac{3}{4},\frac{3}{4}\right]\times\{-1\}$, $E'=\{1\}\times\left[-\frac{3}{4},\frac{3}{4}\right]$ and $\sigma_{i,j}$ is a curve that rounds the corner near (i,j) and smoothly connects the closest pair of line segments just defined. See Figure 4. Set $Q'=Q\cap(\Delta\times[-1,1])$, a smooth closed cylinder. Finally, we specify a particular N_v: set $N_v=Q\setminus Q'$. By the choice of Δ, this N_v is an open regular neighborhood in Q of the vertical edges of Q.

Next smoothly parametrize Δ by polar-like coordinates (r,θ), with $r\in[0,1]$, where θ is the usual polar coordinate, and r will be chosen to facilitate the identification of points $(x,-1)$ and $(x,1)$ in Δ. Let $X_1'=\left[-\frac{7}{8},\frac{7}{8}\right]\times\left[\frac{7}{8},1\right]$ and $X_2'=\left[-\frac{7}{8},\frac{7}{8}\right]\times\left[-1,-\frac{7}{8}\right]$ be rectangular subsets of Δ as shown in Figure 5. Choose r so that

(i) $r=-y$ when $(x,y)\in X_1'$,

(ii) $r=y$ when $(x,y)\in X_2'$,

(iii) $\partial/\partial r=-\partial/\partial x$ when $(x,y)\in\{1\}\times\left[-\frac{3}{4},\frac{3}{4}\right]$,

(iv) $\partial/\partial r = \partial/\partial x$ when $(x,y) \in \{-1\} \times \left[-\frac{3}{4}, \frac{3}{4}\right]$, and

(v) the vector field $\partial/\partial r$ has a single, necessarily elliptic, singularity at $(0,0)$.

Recall that χ is horizontal along $\partial_v Q \cap N_v$. Let χ' denote the smooth extension of the line field χ on $\partial_v Q' \cap \partial_v Q$ to $\partial_v Q'$ obtained by defining χ on $\partial_v Q' \setminus \partial_v Q$ to be the tangent horizontal line field.

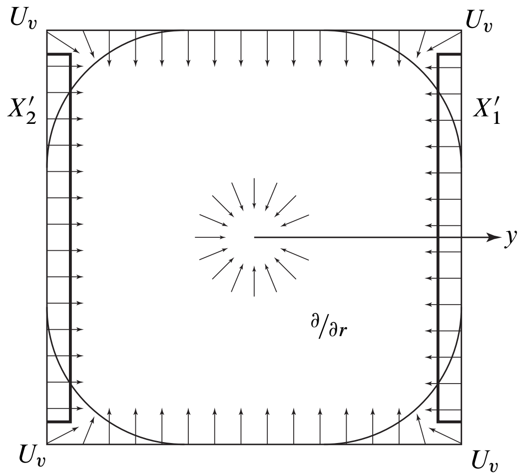


Figure 5: Radial flow on Δ

Let $\mathcal{S}'$ denote a smooth extension of the foliation $\mathcal{S} \cap (\partial_v Q' \cap \partial_v Q)$ to a foliation by circles which satisfies

(1′) χ' dominates $T\mathcal{S}'$, with the domination strict outside $\overline{N}_h$,

(2′) $\mathcal{S}' = \chi' = \mathcal{F}$ on $\overline{N}_h$, and

(3′) $\mathcal{S}'$ is $O(\epsilon)$ C^0 close to each of $\mathcal{F}$ and χ'.

The existence of such an $\mathcal{S}'$ is guaranteed by continuity.

Next we extend the circle foliation $\mathcal{S}'$ to a smooth disk foliation of Q'.

Proposition 7.8 *There is a smooth foliation $\mathcal{D}$ of Q' by disks such that*

(1) *$T\mathcal{D}$ contains $\partial/\partial r$ in a width $\frac{1}{4}$ collar of $\partial_v Q'$,*

(2) *$\mathcal{D} = \mathcal{F}$ on $\overline{N}_h$,*

(3) *$\mathcal{D}$ is everywhere transverse to Φ,*

(4) *$\mathcal{D}$ is $O(\epsilon)$ C^0 close to $\mathcal{F}$, and*

(5) *$\mathcal{S}' = \mathcal{D} \cap \partial_v Q'$.*

Proof Let U denote the intersection of Q' with the width $\frac{1}{2}$ collar of $\partial_v Q$ in Q. In particular, U contains the width $\frac{1}{4}$ collar of $\partial_v Q'$ in Q'.

Let $\mathcal{H}$ be a C^∞ foliation of Q' which is ϵ C^0 close to $\mathcal{F}$, is everywhere transverse to Φ, is x–invariant, and satisfies $\mathcal{H} = \mathcal{F}$ on N_h and $\mathcal{H} = \mathcal{F} = \mathcal{P}$ on $N_R \cap Q'$.

We will choose $\mathcal{D}$ to coincide with $\mathcal{H}$ outside U and to smoothly interpolate between $\mathcal{S}'$ and $\mathcal{H}$ over U. We will take advantage of the polar-like coordinates (r,θ) on Δ. Label the leaves of $\mathcal{H}$ by H_t for $t \in \tau$, where H_t is the (disk) leaf intersecting τ at t, and let $h_t(r,\theta), (r,\theta) \in \Delta$, be the smooth family of functions such that the graph of h_t is the leaf H_t.

Label the leaves of $\mathcal{S}'$ by S'_t for $t \in \tau$, where S'_t is the circle leaf intersecting τ at t. The leaf S'_t can be described as a graph $z = s_t(\theta)$ for $\theta \in S^1$. Since $\mathcal{S}'$ is smooth, s_t defines a smooth family of smooth graphs. Extend $\mathcal{S}'$ to a foliation $\widetilde{\mathcal{S}}'$ of U, by extending the functions s_t to functions $\tilde{s}_t$ defined on U by $\tilde{s}_t(r,\theta) = s_t(\theta)$. Since $\partial/\partial r$ lies in the restriction of $T\mathcal{F}$ to U, the leaves of $\widetilde{\mathcal{S}}'$ lie in Q' and describe a smooth foliation of U.

Let g denote a smooth bump function defined on Δ which is 1 on a width $\frac{1}{4}$ collar of $\partial_v Q'$ and 0 outside U. Since g is smooth and U is compact, g has bounded first partial derivatives, with the bounds independent of ϵ. Finally, define

$$d_t(r,\theta) = g(r,\theta)\tilde{s}_t(r,\theta) + (1 - g(r,\theta))h_t(r,\theta).$$

Let $\mathcal{D}$ be the smooth foliation with leaves given by the graphs of $z = d_t(r,\theta)$. Computing first partial derivatives, we obtain

$$\frac{\partial d_t}{\partial r} = \frac{\partial g}{\partial r} \cdot (\tilde{s}_t - h_t) + (1-g)\frac{\partial h_t}{\partial r}$$

and

$$\frac{\partial d_t}{\partial \theta} = \frac{\partial g}{\partial \theta} \cdot (\tilde{s}_t - h_t) + g \cdot \left(\frac{\partial \tilde{s}_t}{\partial \theta} - \frac{\partial h_t}{\partial \theta} \right) + \frac{\partial h_t}{\partial \theta}.$$

Since Q' is ϵ–flat, $|\tilde{s}_t - h_t| < \epsilon$. The partials of $\tilde{s}_t$ and $\tilde{h}$ are $O(\epsilon)$ small since $\mathcal{S}'$, $\mathcal{H}$ and $\mathcal{F}$ are $O(\epsilon)$ C^0 close to horizontal. It follows that $\mathcal{D}$ is $O(\epsilon)$ close to horizontal, and hence $O(\epsilon)$ close to $\mathcal{F}$. Finally, note that $\mathcal{D} = \widetilde{\mathcal{S}}$ when $g = 1$. Hence, $\partial\mathcal{D} = \mathcal{S}'$, and $T\mathcal{D}$ contains $\partial/\partial r$ in a width $\frac{1}{4}$ collar of $\partial_v Q'$. □

We will use the foliation $\mathcal{D}$ to extend the line field χ' to a contact structure across $Q' \setminus N_h$, and thus to define a smooth confoliation on Q which is contact on the complement of $\overline{N_h \cup N_v}$. First, we establish an elementary glueing lemma.

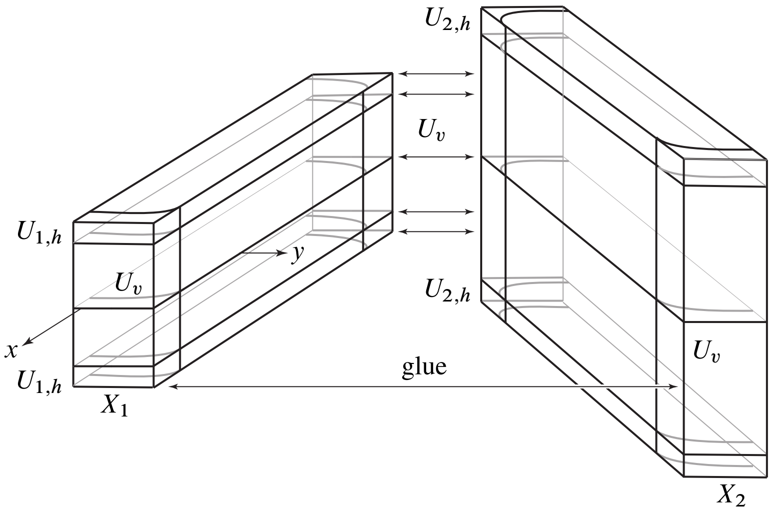


Figure 6: The glueing $X = X_1 \cup X_2$

Lemma 7.9 *Suppose X decomposes as a union of two cubes X_1 and X_2, where $X_1 = [-1,1] \times [u,v] \times J_1$ and $X_2 = [-1,1] \times [v,w] \times J_2$ for some $u < v < w$ and nondegenerate closed intervals $J_1 \subset J_2$.* (See Figure 6.) *Let*

$$\alpha_1 = dz - a(x,y,z)\,dx$$

be a smooth 1–form defining a positive confoliation $\xi_1 = \ker \alpha_1$ on X_1, and let

$$\alpha_2 = dz - b(x,y,z)\,dx$$

be a smooth 1–form defining a positive confoliation on X_2. Let $U_{i,h}$ be a regular open neighborhood of $\partial_h X_i$ in X_i for $i = 1, 2$, and let U_v be a regular open neighborhood of the faces $x = \pm 1$ in X. Suppose that $U_{2,h} \cap X_1 \subset U_{1,h}$ (this allows $U_{2,h} \cap X_1 = \varnothing$).

In addition, suppose that the functions a and b satisfy the following:

(1) $a = b$ *on* $X_1 \cap X_2$,

(2) $a(x,y,z) = 0 \iff a_y(x,y,z) = 0 \iff (x,y,z) \in \overline{U_v \cup U_{1,h}}$, *and*

(3) $b_y(x,y,z) = 0 \iff (x,y,z) \in \overline{U_v \cup U_{2,h}}$.

Then there is a smooth 1–form $\alpha = dz - c(x,y,z)\,dx$ defining a positive confoliation $\xi = \ker \alpha$ on X, where c satisfies

(c1) $c = a$ *on a neighborhood of the $y = u$ face of X_1,*

(c2) $c = b$ *on a neighborhood of the $y = w$ face of X_2, and*

(c3) $c(x,y,z) = 0 \iff c_y(x,y,z) = 0 \iff (x,y,z)$ *is in the closure of* $U_v \cup U_{1,h} \cup U_{2,h}$.

Moreover, if each of a and b is C^0 close to 0, then so is c.

In other words, the continuous 1–form $\alpha_1 \cup \alpha_2$ on X can be C^0 approximated by a smooth 1–form α which agrees with $\alpha_1 \cup \alpha_2$ on a neighborhood of ∂X, and describes a positive confoliation $\xi = \ker(\alpha)$ on X which is a contact structure exactly where ξ_1 or ξ_2 is a contact structure.

Proof For $i = 1$ or 2, ξ_i is a positive confoliation, that is, $\alpha_i \wedge d\alpha_i \geq 0$. Hence $a_y \geq 0$ and $b_y \geq 0$. Also, by hypothesis, $a = b$ on $X_1 \cap X_2$. Therefore, for each $(x_0, z_0) \in [-1, 1] \times J_1$, the one-variable functions $a(x_0, y, z_0)$ for $y \in [u, v]$ and $b(x_0, y, z_0)$ for $y \in [v, w]$ piece together to give a continuous function defined on $[v, w]$ which is smooth on the complement of $\{v\}$.

Notice that $b(x, y, z) = 0$ if and only if $(x, y, z) \in U_{1,h} \cap X_2$ or $b_y(x, y, z) = 0$. So $b(x, y, z) = 0$ if and only if $(x, y, z) \in (U_{1,h} \cap X_2) \cup \overline{U_v \cup U_{2,h}}$.

To facilitate blending a and b into a smooth function c, choose a smooth function $\tilde{b}$ on X_2 such that:

($\tilde{b}1$) $\tilde{b} = b$ in a neighborhood of the $y = w$ face of X_2.

($\tilde{b}2$) $\tilde{b}_y \geq 0$.

($\tilde{b}3$) $\tilde{b} \geq b$ on X_2.

($\tilde{b}4$) $\tilde{b}(x, v, z) > b(x, v, z) \iff (x, v, z) \notin \overline{U_v \cup U_{2,h}}$.

($\tilde{b}5$) $\tilde{b} = 0 \iff \tilde{b}_y = 0 \iff b_y = 0$.

At a point $(x, v, z) \in X_1 \cap X_2$ it might be that $a_y(x, v, z) > b_y(x, v, z) > 0$, but in this case, the choice of $\tilde{b}$ forces $a(x, v, z) < \tilde{b}(x, v, z)$. Thus it is possible to define a smooth $\widetilde{a}$ on $X_1 \cup X_2$ which satisfies all of the following, and in particular ($\widetilde{a}3$):

($\widetilde{a}1$) $\widetilde{a} = a$ on X_1.

($\widetilde{a}2$) $\widetilde{a}_y \geq 0$.

($\widetilde{a}3$) $\widetilde{a} \leq \tilde{b}$.

($\widetilde{a}4$) $\widetilde{a}(x, y, z) = 0 \iff \widetilde{a}_y(x, y, z) = 0 \iff (x, y, z)$ is an element of the closure of $U_v \cup U_{1,h} \cup U_{2,h}$.

To produce such an $\widetilde{a}$, pick a nonnegative smooth extension $e(x, y, z)$ of $a_y(x, y, z)$ to $X_1 \cup X_2$ which is 0 exactly on $\overline{U_v \cup U_{1,h} \cup U_{2,h}}$. With ($\widetilde{a}3$) in mind, such an extension can be modified by multiplying by a smooth nonnegative map which takes the value 1 on X_1 and approaches 0 quickly on X_2 so that

$$\widetilde{a}(x, y, z) = \int_u^y e(x, s, z)\,ds + a(x, u, z)$$

satisfies ($\widetilde{a}1$)–($\widetilde{a}4$).

Now choose $\sigma\colon [u,w] \to [0,1]$ such that $\sigma([u,v]) = 0$, $\sigma\big(\big[\frac{1}{2}(v+w), w\big]\big) = 1$, and σ maps $\big(v, \frac{1}{2}(v+w)\big)$ diffeomorphically to $(0,1)$. Set

$$c(x,y,z) = (1-\sigma(y))\tilde{a}(x,y,z) + \sigma(y)\tilde{b}(x,y,z),$$

so that properties (c1) and (c2) of c are immediate. Since

$$c_y = \sigma_y(\tilde{b} - \tilde{a}) + (1-\sigma)\tilde{a}_y + \sigma\tilde{b}_y \geq 0,$$

is the sum of three nonnegative terms, ξ is a positive confoliation. Moreover, if $c_y = 0$, either $a_y = 0$ and $(x,y,z) \in \overline{U_v \cup U_{1,h}}$, or $b_y = 0$ and $(x,y,z) \in \overline{U_v \cup U_{2,h}}$. Hence, if $c_y(x,y,z) = 0$, necessarily $(x,y,z) \in \overline{U_v \cup U_{1,h} \cup U_{2,h}}$. But this means also that $\tilde{a} = \tilde{b} = 0$, and so $c = 0$.

Conversely, suppose $(x,y,z) \in \overline{U_v \cup U_{1,h} \cup U_{2,h}}$. It suffices to consider the case that $(x,y,z) \in U_v \cup U_{1,h} \cup U_{2,h}$, and hence at least one of the following is true:

(1) $(x,y,z) \in U_v$ means $\tilde{a} = 0$ and $\tilde{b} = b = 0$,

(2) $(x,y,z) \in U_{1,h}$ means $\sigma = 0$ and $\tilde{a} = 0$, and

(3) $(x,y,z) \in U_{2,h}$ means $\tilde{b} = b = 0$ and $\tilde{a} = 0$.

In each of these three cases, $c = 0$ on an open set about (x,y,z) and therefore $c_y(x,y,z) = 0$. Hence, property (c3) of c is satisfied. □

Corollary 7.10 *There exists a smooth confoliation ξ on Q that is $O(\epsilon)$ C^0 close to $T\mathcal{F}$, has characteristic foliation χ on $\partial_v Q$, and satisfies $\xi = T\mathcal{F}$ on $\overline{N_h \cup N_v}$. Moreover, ξ can be chosen so that $\pi(\xi)$ is a smooth confoliation on V, where π is the quotient map $\pi\colon Q \to V$.*

Proof At all points of $\overline{N} = \overline{N_h \cup N_v}$, let ξ be the tangent plane to $\mathcal{F}$. Thus ξ contains χ along $\overline{N} \cap \partial_v Q$, and ξ contains χ' along $\overline{N} \cap \partial_v Q'$.

The foliation by disks, $\mathcal{D}$, given by Proposition 7.8 will be used to extend ξ to all of Q'.

Let ι denote the smooth inward pointing vector field on Q' given by lifting the vector field $-\partial/\partial r$ to the leaves of $\mathcal{D}$, where the lift is the pullback under the projection $(r,\theta,z) \to (r,\theta)$. In particular, in the width $\frac{1}{4}$ collar about $\partial_v Q'$, $\iota = -\partial/\partial r$.

Notice that χ' and ι span a plane at every point of $\partial_v Q'$. Denote this plane by ξ.

At this point we have the start of a smooth confoliation ξ on $\partial_v Q' \cup \overline{N}$, and Q' is foliated by disks of $\mathcal{D}$ which are in turn either everywhere tangent to ξ or foliated by a vector field ι which serves as a candidate for a Legendrian vector field. This is directly analogous to a cylindrical neighborhood of the z–axis in the standard radial model of a tight contact structure.

To extend ξ across Q' it is enough to map Q' to a standard model, use the technique of Lemma 5.14 of [24], and pull back the resulting contact structure to a contact structure ξ on Q'. Roughly speaking, Q' is mapped to a solid cylinder in $\mathbb{R}^3$ centered along the z–axis in such a way that $\mathcal{D}$ and ι are mapped to horizontal planes and radial lines, respectively. The standard radially symmetric contact structure is then pulled back to Q'.

Some extra care is needed so that the confoliation ξ on Q is $O(\epsilon)$ close to $T\mathcal{F}$. By construction, it is $O(\epsilon)$ close along $\partial_v Q$ and *in coordinates* the planes of ξ monotonically approach horizontal planes in $\mathbb{R}^3$ as you move radially towards the z–axis along Legendrian curves. The issue is that the pullback confoliation planes may not monotonically approach $T\mathcal{D}$.

Since Q' is compact, the metric distortion when compared to the standard model is bounded, and so it suffices to show that we can reduce to the case that the line field χ' is arbitrarily close to $T\mathcal{S}'$. We do this by taking advantage of the width $\frac{1}{2}$ collar, U, of $\partial_v Q$ in Q to define a contact structure with planes that rotate from slope χ' to slope close to $T\mathcal{S}'$ as follows.

Since the restriction of $\partial/\partial r$ to U lies in both $T\mathcal{F}$ and $T\mathcal{D}$, the restrictions to U of the flow lines for $\partial/\partial r$ lie in both $\mathcal{D}$ and $\mathcal{F}$.

Recall that $\partial\Delta$ corresponds to $r=1$, and let $s(\theta,z)$ denote the slope of $T\mathcal{S}'$ at $(1,\theta,z)$. Note that $s(\theta,z)$ is also the slope of $T\mathcal{D}\cap\partial_v U$ at (r_0,θ,z), where r_0 is determined by the condition that $(r_0,\theta,z)\in(\partial_v U\setminus\partial_v Q)$. Let $c(\theta,z)$ denote the slope of χ' at $(1,\theta,z)$. Since χ' strictly dominates $\mathcal{S}'$ exactly on the complement of $\overline{N}_h$,

$$s(\theta,z)-c(\theta,z)\geq 0,$$

with equality if and only if $(1,\theta,z)\in\overline{N}_h$.

Fix $\delta\in\left(0,\frac{1}{2}\right)$, and let $f\colon U\to[\delta,1]$ be a smooth function which satisfies

(i) $f(r,\theta,z)=1$ for all $(r,\theta,z)\in\overline{N}\cup\partial_v Q'$,

(ii) $f(r,\theta,z)=\delta$ for all $(r,\theta,z)\in(\partial_v U\setminus(\partial_v Q\cup N_h))$, and

(iii) $f_r(r,\theta,z)\geq 0$ for all $(r,\theta,z)\in U$, with equality if and only if $(r,\theta,z)\in\overline{N}$.

For $(r,\theta,z)\in U$, define

$$\alpha=dz-[(1-f(r,\theta,z))s(\theta,z)+f(r,\theta,z)c(\theta,z)]\,d\theta.$$

Setting $g(r,\theta,z) = (1 - f(r,\theta,z))s(\theta,z) + f(r,\theta,z)c(\theta,z)$, we have

$$\alpha = dz - g(r,\theta,z)\,d\theta.$$

Notice that along $\partial_v Q'$, $\ker(\alpha) = \xi$. In addition,

$$d\alpha = -g_r\,dr\,d\theta - g_z\,dz\,d\theta,$$

and hence

$$\alpha \wedge d\alpha = -g_r\,dr\,d\theta\,dz = -\big(f_r(r,\theta,z)(c(\theta,z) - s(\theta,z))\big)\,dr\,d\theta\,dz$$

is a smooth positive confoliation defined on U. This confoliation agrees with $T\mathcal{F}$ on $U \cap \overline{N}$ and is a contact structure on $U \setminus \overline{N}$. Moreover, by choosing δ as small as is necessary, we may guarantee that at each point $(r,\theta,z) \in (\partial_v U \setminus \partial_v Q)$ the line field given by ξ restricted to $\partial_v U$ dominates and is as close to the line field given by $T\mathcal{D}$ restricted to $\partial_v U$ as is required. Hence there is a smooth confoliation ξ on Q that is $O(\epsilon)$ C^0 close to $T\mathcal{F}$, has characteristic foliation χ on $\partial_v Q$, and satisfies $\xi = T\mathcal{F}$ on $\overline{N_h \cup N_v}$.

To ensure that the contact structure on Q glues smoothly to V, some further care is needed. Extending the 2–plane field from the vertical boundary of a cylinder across a radially foliated disk involves a choice of rate of rotation of contact planes to horizontal. The rates should be chosen to respect the glueing of B to D. Since $\xi = T\mathcal{F}$ along N_v, it suffices to show that the rates can be chosen to respect the glueing of B' to D'. To see that this is possible, proceed as follows.

Recall the rectangles $X_1', X_2' \subset \Delta$, and for $i = 1, 2$ let X_i denote the points $(r,\theta,z) \in Q$ with $(r,\theta) \in X_i'$. Let ξ_i denote the restriction of ξ to X_i. In terms of (x,y,z) coordinates and given the form of ξ in U, we can write $\xi_1 = \ker(dz - a(x,y,z)\,dx)$ and $\xi_2 = \ker(dz - b(x,y,z))$ for smooth functions $a(x,y,z)$ defined on $X_1 = \left[-\frac{7}{8}, \frac{7}{8}\right] \times \left[\frac{7}{8}, 1\right] \times \sigma$ and $b(x,y,z)$ defined on $X_2 = \left[-\frac{7}{8}, \frac{7}{8}\right] \times \left[-1, -\frac{7}{8}\right] \times \tau$.

Without changing notation, regard X_1 and X_2 as subsets of V using the quotient map $\pi\colon Q \to V$ given by identifying $y = \pm 1$. Since $\xi_1 = \xi_2$ along $X_1 \cap X_2$, $a(x,y,z) = b(x,y,z)$ along $X_1 \cap X_2$ in V.

Recall also that χ is dominated by $\mathcal{F}$ along D and dominates $\mathcal{F}$ along B. Hence, $a(x,y,z) = b(x,y,z) \geq 0$ along $X_1 \cap X_2$, with equality if and only if $(x,y,z) \in \overline{N_h \cup N_v}$. Lemma 7.9 therefore applies and produces a smooth contact structure on X which agrees with $\xi_1 \cup \xi_2$ in a neighborhood of ∂X. □

Proof of Theorem 7.2 Let $\epsilon > 0$. By symmetry, it suffices to establish the existence of ξ^+.

Let $\mathcal{F}_0$ be a $C^{\infty,0}$ foliation of M, and let $\mathcal{F}$ be an ϵ C^0 close $C^{\infty,0}$ foliation constructed using Theorem 6.2. Let $V(\tau, A)$ be one of the holonomy neighborhoods constructed for $\mathcal{F}$.

Corollary 7.4 and Proposition 7.6 guarantee the existence of a smooth foliation χ on $\partial_v Q$ which is dominated by and $O(\epsilon)$ C^0 close to $\mathcal{F}$. Corollaries 7.7–7.10 show that χ is the characteristic foliation of a smooth confoliation on Q that glues to a smooth confoliation on V. This confoliation restricts to a contact structure outside $\overline{N}$. Each of $\xi_V^{\pm}$ is $O(\epsilon)$ C^0 close to $\mathcal{F}$ on V, and hence $O(\epsilon)$ C^0 close to $T\mathcal{F}_0$ on V. □

8 Foliations with only trivial holonomy

In this section we investigate taut, transversely oriented C^0 foliations with only trivial holonomy. First we recall some classical results, with a focus on smoothness assumptions. New results then appear as Theorems 8.10 and 8.12 and Corollary 8.16, with Corollary 8.16 giving the conclusion needed for the main result in this paper.

Versions of the next result appear as Theorem 4 of [40], Lemma 3.6 of [22], or, in greatest generality, Corollary 2.6 of [28].

Theorem 8.1 *A transversely oriented C^0 foliation that has trivial holonomy is topologically taut.*

Our focus is on approximations of $C^{1,0}$ and smoother foliations, thus Corollary 5.6 of [26], can be used to approximate a topologically taut C^0 foliation with an isotopic taut $C^{\infty,0}$ foliation when convenient.

Next we recall a theorem found in [22].

Theorem 8.2 [22, Theorem 4.1] *Let $\mathcal{F}$ be a transversely oriented C^0 foliation in M with only trivial holonomy, and let Φ be a C^0 flow transverse to $\mathcal{F}$. Let $\widetilde{M}$ denote the universal cover of M, and let $\widetilde{\mathcal{F}}$ and $\widetilde{\Phi}$ be the lifts of $\mathcal{F}$ and Φ to $\widetilde{M}$. Then $\widetilde{M}$ is homeomorphic to $\mathbb{R}^2 \times \mathbb{R}$, where each $\mathbb{R}^2 \times \{z\}$, $z \in \mathbb{R}$, is a leaf of $\widetilde{\mathcal{F}}$ and each $\{x\} \times \mathbb{R}$, $x \in \mathbb{R}^2$, is an orbit of $\widetilde{\Phi}$.*

This theorem is given in [22] for foliations that are C^∞. However, this smoothness hypothesis in unnecessary. For completeness, we include Imanishi's proof here, reframed using the language of leaf spaces and with careful attention paid to smoothness assumptions.

Lemma 8.3 [22] *Let $\mathcal{F}$ be a transversely oriented C^0 foliation with only trivial holonomy, and let Φ be a C^0 flow transverse to $\mathcal{F}$. Suppose $H\colon [0,1]\times[0,1)\to M$ is a continuous map such that $H([0,1]\times\{t\})$ is a curve in a leaf of $\mathcal{F}$ for all t, and $H(\{s\}\times[0,1))$ is an immersed curve in a flow line of Φ for all s. If H extends continuously to $\{0\}\times[0,1]$, then H extends continuously to $[0,1]\times[0,1]\to M$.*

Proof This follows immediately from Theorem 3.1 of [22]. See also Lemma 9.2.4 of [4]. □

Proof of Theorem 8.2 Let $\widetilde{M}$ be the universal cover of M and let $\widetilde{\mathcal{F}}$ and $\widetilde{\Phi}$ be the lifts of $\mathcal{F}$ and Φ to $\widetilde{M}$. Let T be the leaf space of $\mathcal{F}$, and let $\rho\colon \widetilde{M}\to T$ denote the associated quotient map. Note that it is sufficient to prove that $\rho(\widetilde{C})=T$ for any orbit $\widetilde{C}$ of $\widetilde{\Phi}$.

So suppose that $\rho(\widetilde{C})\neq T$ for some orbit $\widetilde{C}$ of $\widetilde{\Phi}$. Since $\widetilde{C}$ is everywhere transverse to $\widetilde{\Phi}$, $\rho(\widetilde{C})$ is an embedded copy of $\mathbb{R}$ in T. Note that although $\widetilde{C}$ is properly embedded in $\widetilde{M}$, $\rho(\widetilde{C})$ may or may not be properly embedded in T. In either case, there is a leaf $\widetilde{L}$ of $\widetilde{\mathcal{F}}$ such that $\rho(\widetilde{L})$ lies in the closure of $\rho(\widetilde{C})$ but not in $\rho(\widetilde{C})$. Let $\widetilde{C}'$ be an orbit of $\widetilde{\Phi}$ passing through $\widetilde{L}$. There is an interval $[x,y]$ in T such that $[x,y]\cap\rho(\widetilde{C}')=[x,y]$ and $[x,y]\cap\rho(\widetilde{C})=[x,y)$. As illustrated in Figure 7, this is impossible by Lemma 8.3. □

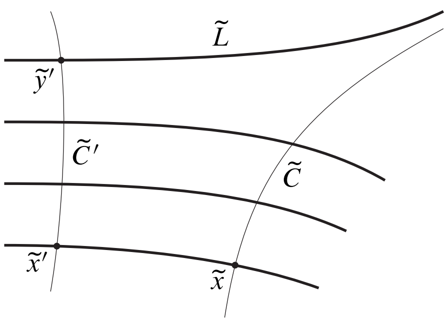


Figure 7: A continuous H which does not continuously extend

Next we show that some of this product structure on $\widetilde{M}$ can be seen also in M.

Corollary 8.4 *Let $\mathcal{F}$ be a transversely oriented C^0 foliation with only trivial holonomy, and let Φ be a C^0 flow transverse to $\mathcal{F}$. Let Λ be a minimal set of $\mathcal{F}$. Then any region complementary to Λ is a product $L\times[0,1]$, with each $\{x\}\times[0,1]$ a segment of the flow Φ. Moreover, the restriction of $\mathcal{F}$ to this complementary region is, up to a Φ–preserving isotopy, the product foliation $L\times[0,1]$.*

Proof Let X denote the metric closure of some complementary region of Λ. By Theorem 8.2, any lift of X to $\widetilde{M}$ has the form $\widetilde{L} \times [0,1]$, where $\widetilde{L}$ is the lift of a boundary leaf L of X, and each $\{\tilde{x}\} \times [0,1]$ is a segment of the flow $\widetilde{\Phi}$. It follows that X is an I–bundle, with each I–fiber a segment of a flow line of Φ. Now apply Lemma 4.7 to conclude that the restriction of $\mathcal{F}$ to X is, up to a Φ–preserving isotopy, the product foliation $L \times [0,1]$. □

Corollary 8.5 *Let $\mathcal{F}$ be a transversely oriented C^0 foliation with only trivial holonomy. Then exactly one of the following is true:*

(1) *$\mathcal{F}$ is a fibering of M over S^1.*

(2) *$\mathcal{F}$ is minimal.*

(3) *$\mathcal{F}$ is a single Denjoy blow up of a minimal foliation; equivalently, $\mathcal{F}$ contains a unique minimal set, and this minimal set is exceptional with complement a product.*

Proof Let Φ be a C^0 flow transverse to $\mathcal{F}$. If $\mathcal{F}$ is not minimal, then it contains a minimal set, Λ say. Let X denote the metric closure of any component of the complement of Λ. It follows from Corollary 8.4 that X is an I–bundle $L \times [0,1]$, and the restriction of $\mathcal{F}$ to X is, up to Φ–preserving isotopy, the product foliation $L \times [0,1]$. If Λ is a compact leaf, conclude that $\mathcal{F}$ is a fibering of M. Otherwise, Λ is exceptional, and $\mathcal{F}$ is a single Denjoy blow up (along at most countably many leaves) of a minimal foliation; in other words, $\mathcal{F}$ contains a unique minimal set, and this minimal set is exceptional with complement a product. □

Theorem 8.6 [22, Theorem 1.3] *Let $\mathcal{F}$ be a transversely oriented C^0 foliation with only trivial holonomy. Suppose $\mathcal{F}$ does not contain an exceptional minimal set. Then there is a topological flow $\psi\colon M \times \mathbb{R} \to M$ such that:*

(1) *ψ preserves $\mathcal{F}$; ie $\psi(-,t)$ sends each leaf of $\mathcal{F}$ into a leaf of $\mathcal{F}$.*

(2) *ψ is topologically transverse to $\mathcal{F}$; ie, $\psi(x,\mathbb{R})$ is topologically transverse to $\mathcal{F}$.*

Recall that a *transverse measure* on a codimension-one foliation $\mathcal{G}$ is an *invariant* measure on each arc transverse to $\mathcal{G}$ that is equivalent to Lebesgue measure on an interval of $\mathbb{R}$. Invariant, in this context, means that the measure of a transverse arc is unchanged under isotopies of the arc that keep each point on the same leaf of $\mathcal{G}$.

Lemma 8.7 *Let $\psi\colon M \times \mathbb{R} \to M$ be the topological flow of Theorem 8.6. If $\psi(x_1,[s_1,t_1])$ and $\psi(x_2,[s_2,t_2])$ are isotopic through an isotopy that keeps each point on the same leaf of $\mathcal{F}$, then $t_1 - s_1 = t_2 - s_2$.*

Proof The isotopy can be lifted to the universal cover of M, so without changing notation we take $M = \widetilde{M}$. The advantage of working in $\widetilde{M}$ is that leaves of the foliation are in bijective correspondence with $\psi(x_2, t)$ for $t \in \mathbb{R}$.

Let $\tau_i = \psi(x_i, [s_i, t_i])$. The given isotopy sweeps out a family of curves α_t each contained in a leaf of $\mathcal{F}$ that starts at a point of τ_1 and ends at a point of τ_2, with $t \in [s_1, t_1]$ so that $\alpha_t(0) = \psi(x_1, t)$. Another family of arcs β_t each contained in a leaf of $\mathcal{F}$ can be generated by using the flow. Define $\beta_t(s) = \psi(\alpha_{s_1}(s), t - s_1)$.

It follows that $\beta_t(0) = \psi(\psi(x_1, s_1), t - s_1) = \psi(x_1, t) = \alpha_t(0)$ for $t \in [s_1, t_1]$, while $\beta_t(1) = \psi(\psi(x_2, s_2), t - s_1) = \psi(x_2, s_2 + t - s_1)$.

The arcs α_t and β_t have the same initial point, are both contained in the same leaf, and terminate in the flow segment $\psi(x_2, [s_2 + t_1 - s_1, t_2])$. It follows that they terminate in the same leaf, that is, $s_2 + t_1 - s_1 = t_2$. □

Corollary 8.8 [22, Corollary 4.1] *Let $\mathcal{F}$ be a transversely oriented C^0 foliation with only trivial holonomy. Suppose $\mathcal{F}$ does not contain an exceptional minimal set. Then $\mathcal{F}$ admits a transverse measure.*

Proof Let τ be a transversal to $\mathcal{F}$. First we show that if the terminal point of τ is $\psi(x, t)$, then there is an $[s, t] \subset \mathbb{R}$ such that τ is isotopic to $\psi(x, [s, t])$ through an isotopy that keeps each point on the same leaf of $\mathcal{F}$.

To see this, fix the initial point x_0 of τ, and use the flow to homotope τ to an arc σ starting at x_0 and contained in a leaf of $\mathcal{F}$. This homotopy can be thought of as the image of a triangle T, with edges τ and σ, and a flow arc $\psi(x, [s, t])$ for some value of s swept out by $\psi(x, t)$. The desired isotopy of τ is given by following horizontal arcs of $T \cap \mathcal{F}$.

Define the length of τ to be $|t - s|$. Lemma 8.7 guarantees that this defines a positive transverse measure on $\mathcal{F}$. □

Branched surfaces were first introduced by Williams in [44]. We refer the reader to [9; 33; 34] for the definitions of branched surface B and an associated I–bundle neighborhood $N(B)$. We recall that a surface S, not necessarily compact, is *carried* by a branched surface B if S is injectively immersed in $N(B)$ so that it is everywhere transverse to the I–fibers. A minimal set of a foliation is carried by B if each of its leaves is carried by B.

It is standard to say that a foliation $\mathcal{F}$ is carried by B if a Denjoy blow up, $\mathcal{F}'$, of $\mathcal{F}$ along a single leaf L results in a lamination $\Lambda = \mathcal{F}' \setminus j(L \times (0,1))$ such that all leaves of Λ are carried by B. (Note that if B carries a foliation in this sense, then the regions complementary to $N(B)$ are necessarily products.) In this paper, however, we instead introduce and use the following definition.

Definition 8.9 A foliation $\mathcal{F}$ is *carried by* a branched surface B if

(1) the restriction of $\mathcal{F}$ to $N(B)$ is a foliation transverse to the I–fibers and tangent to $\partial_h N(B)$, and
(2) the restriction of $\mathcal{F}$ to $M \setminus \operatorname{int} N(B)$ is a product foliation that is transverse to the I–fibers of $N(B)$ along $\partial_v N(B)$ and tangent to $\partial_h N(B)$.

We refer the reader to [20; 29] for a description of the splitting open of a branched surface and the corresponding "splitting open" of the I–bundle neighborhood $N(B)$.

Theorem 8.10 *Let $\mathcal{F}$ be a transversely oriented C^0 foliation of M with only trivial holonomy. Then M fibers over S^1 and either*

(a) *$\mathcal{F}$ is measured, or*
(b) *$\mathcal{F}$ is a single Denjoy blow up of a minimal measured C^0 foliation.*

In the case that $\mathcal{F}$ is measured, either it is minimal or it is a fibering of M over S^1.

Proof The proof proceeds by analyzing the three cases arising in the conclusion of Corollary 8.5.

Case (1) If $\mathcal{F}$ is a fibering, of course M is fibered. It follows immediately that $\mathcal{F}$ satisfies conclusion (1).

Case (2) Since $\mathcal{F}$ does not contain an exceptional minimal set, Corollary 8.8 can be applied. From this it follows that $\mathcal{F}$ is transversely measured. Since $\mathcal{F}$ has a transverse measure, it is fully carried by a measured branched surface B, and since $\mathcal{F}$ is a foliation, the complementary regions of B are products. It remains to show that M fibers over S^1. This will be done following the next case.

Case (3) Then $\mathcal{F}$ is a single Denjoy blow up of a minimal foliation $\mathcal{E}$. Denote the minimal set of $\mathcal{F}$ by Λ and denote the smooth transverse flow used in the blow up by Φ. By Corollary 8.8, $\mathcal{E}$ is measured, thus (2) is satisfied. It remains to show that M fibers over S^1.

This measure on $\mathcal{E}$ determines a transverse measure on Λ in the sense that any segment of a flow line of Φ can be given the measure it has when viewed as a transversal to $\mathcal{E}$. Since Λ has a transverse measure, it is fully carried by a measured branched surface B, and since the complementary regions of Λ are I–bundles, the complementary regions of B are products. So Proposition 4.11 of [33] applies, thereby implying M fibers over S^1.

Hence, in each of the cases (2) and (3), $\mathcal{F}$ is fully carried by a transversely measured branched surface for which every complementary region is a product. The space of positive measures on B is an open cone in a vector space since it is the solution space of a system of homogenous linear equations. Since there is a nontrivial real solution and the coefficients of these equations are integers, there is a nontrivial, positive, rational solution arbitrarily close to the real solution. Any positive rational solution corresponds to an integral measure on B and hence describes a surface S (not necessarily connected) which is fully carried by B.

Since any complementary region to B, and hence to S, is necessarily a product, it follows that M cut open along S is an I–bundle. Letting S_0 be a (possibly the) component of S, it follows that M is a fiber bundle over S^1 with fiber S_0. □

Next we apply a theorem from [25].

Theorem 8.11 [25] *Suppose $\mathcal{F}$ is a transversely orientable $C^{1,0}$ measured foliation in M. Then there is an isotopy of M taking $\mathcal{F}$ to a C^∞ measured foliation which is C^0 close to $\mathcal{F}$. If Φ is a smooth flow transverse to $\mathcal{F}$, the isotopy may be taken to map each flow line of Φ to itself.*

Theorem 8.12 *Fix $\epsilon > 0$. Let $\mathcal{F}$ be a transversely oriented $C^{1,0}$ foliation of M with only trivial holonomy. Then M fibers over S^1, and $\mathcal{F}$ is $O(\epsilon)$ C^0 close to a smooth fibering of M.*

Proof By Theorem 8.10, M fibers over S^1, and $\mathcal{F}$ is either measured or a single Denjoy blow up of a measured foliation. In the case that it is measured and not minimal, it is necessarily a fibering, and so there is nothing to prove. If $\mathcal{F}$ is measured and not a fibering, it is necessarily minimal, and the first step is to perform a single Denjoy blow up.

Thus, it is enough to consider the case when $\mathcal{F}$ is a single Denjoy blow up of a minimal measured $C^{1,0}$ foliation $\mathcal{E}$. By Theorem 8.11, it suffices to show that $\mathcal{F}$ is $O(\epsilon)$ C^0 close to a $C^{1,0}$ fibering of M. Denote the exceptional minimal set of $\mathcal{F}$ by Λ, and denote the smooth transverse flow used in the blow up by Φ. Let B be a smoothly embedded, transversely oriented branched surface which is transverse to Φ and which fully carries $\mathcal{F}$. In particular, Λ lies in a regular I–fibered neighborhood N of B, transverse to the I–fibers, with the I–fibers segments of the flow Φ. In addition, the regions complementary to $\operatorname{int}(N)$ are sutured manifold products, and $\mathcal{F}$ restricts to a product foliation on these complementary regions. Since Λ has a transverse measure, so does B.

Now fix $\epsilon > 0$. Use Λ to split B open as much as is necessary to a smoothly embedded, measured branched surface B_{split} so that the result of splitting open N is a regular I–fibered neighborhood N_{split} that has an ϵ–flat $(\mathcal{F}, \Phi)$ flow box decomposition such that each flow box has horizontal boundary contained in $\partial_h N(B_{\text{split}})$. We say $\mathcal{G}_0$ is a foliation of N_{split} if in addition to the usual local product leaf structure, it is tangent to $\partial_h N_{\text{split}}$, transverse to $\partial_v N_{\text{split}}$, and everywhere transverse to Φ. By ϵ–flatness, any foliation $\mathcal{G}_0$ of N_{split} can be isotoped relative to $\partial N_{\text{split}}$ to be ϵ C^0 close to the restriction $\mathcal{F}_0$ of $\mathcal{F}$ to N_{split}.

Pick a rational measure μ fully carried by B_{split}, and let $\mathcal{G}_0$ be a $C^{1,0}$ fibering of M determined up to isotopy by (B, μ). Choose $\mathcal{G}_0$ to be everywhere transverse to Φ and so that its restriction to N_{split} is a foliation of N_{split}.

Consider a component Σ of the metric closure of the complement of N_{split}. Both $\mathcal{F}$ and $\mathcal{G}_0$ restrict to product foliations on Σ, and hence to foliations by circles on each component A_i of $\partial_v N_{\text{split}}$. Again appealing to the ϵ–flatness of $N(B_{\text{split}})$, there is an $O(\epsilon)$ C^0 small isotopy of M in a neighborhood of $\bigcup_i A_i$ taking $\mathcal{F}$ to $\mathcal{F}_0$ such that if F and G are leaves of the restrictions of $\mathcal{F}_0$ and $\mathcal{G}_0$ respectively to Σ, then either $F = G$ in a neighborhood of $\partial F = \partial G$ or $\partial F \cap \partial G = \varnothing$. Now let $\mathcal{G}$ be the foliation obtained by letting $\mathcal{G}$ coincide with $\mathcal{G}_0$ on N_{split} and coincide with $\mathcal{F}_0$ on each component Σ. By construction, $\mathcal{G}$ is a $C^{1,0}$ fibering of M. Moreover, $\mathcal{G}$ is $O(\epsilon)$ close to $\mathcal{F}$ since $\mathcal{G}_0$ is $O(\epsilon)$ close to $\mathcal{F}$ on N_{split} and coincides with $\mathcal{F}_0$ on each Σ. □

Recall the statement of Tischler's theorem.

Theorem 8.13 [42] *A transversely oriented, C^∞ measured foliation $\mathcal{F}$ of M can be C^∞ approximated by a smooth fibering $\mathcal{G}$ of M over S^1.*

Thus Theorem 8.12 weakens the assumptions of smoothness in Tischler's theorem for 3–manifolds.

Next, we consider the case that all leaves in some minimal set of $\mathcal{F}$ are planes. If $\mathcal{F}$ has only trivial holonomy, then it is Reebless, and the possibilities for $(M, \mathcal{F})$ are very well understood, by work of Imanishi and Gabai:

Lemma 8.14 [16; 22] *Let $\mathcal{F}$ be a Reebless C^0 foliation of a closed 3–manifold M. Suppose $\mathcal{F}$ contains a minimal set all of whose leaves are planes. Then all leaves of $\mathcal{F}$ are planes and $M = T^3$.*

In fact, as shown by Bowden, the condition that $\mathcal{F}$ be Reebless can be removed:

Lemma 8.15 [1, Lemma 2.12] *Let $\mathcal{F}$ be a C^0–foliation on a manifold M that has a minimal set all of whose leaves are planes. Then $M = T^3$, and $\mathcal{F}$ is a foliation by planes.*

The C^0 foliation theory that is used for the main result in this paper is the following corollary.

Corollary 8.16 *Let $\mathcal{F}$ be a transversely oriented $C^{1,0}$ foliation of a closed 3–manifold M. Suppose $\mathcal{F}$ contains a minimal set all of whose leaves are planes. Then $M = T^3$ and $\mathcal{F}$ is C^0 close to a smooth fibering of M by tori.*

Proof By Lemma 8.15, all leaves of $\mathcal{F}$ are planes and $M = T^3$. Necessarily, therefore, $\mathcal{F}$ has only trivial holonomy, and so by Theorem 8.12, $\mathcal{F}$ is C^0 close to a smooth fibering of M. The leaves of this fibering are necessarily tori as they are π_1–injective. □

9 Propagation

In this section, we describe how to extend the smooth confoliation on V to a smooth contact structure on M which is $O(\epsilon)$ close to $\mathcal{F}$, and hence prove our main result, Theorem 1.2.

Begin by recalling the propagation technique introduced in [24]. The starting point is a decomposition of a manifold into two codimension-0 pieces. Roughly speaking, one piece, V, has a contact structure, while the other piece, W, has a foliation. As long as the contact structure dominates the foliation along $\partial_v V = \partial_v W$ and every point of W can be connected to V by a path in a leaf, the contact structure can be propagated throughout W.

For completeness we include the formal definitions of these concepts and the main theorem of [24] stated in the C^0 setting.

Definition 9.1 [24, Definition 6.4] Let M be a closed oriented 3–manifold with smooth flow Φ. Suppose that M can be expressed as a union

$$M = V \cup W,$$

where V and W are smooth 3–manifolds, possibly with corners, such that $\partial V = \partial W$. We say that this decomposition is *compatible with the flow* Φ if ∂V (and hence ∂W) decomposes as a union of compact subsurfaces $\partial_v V \cup \partial_h V$, where $\partial_v V$ is a union of flow segments of Φ and, $\partial_h V$ is transverse to Φ. Let U be a preferred regular neighborhood of the union of the horizontal 2–cells and the vertical 1–cells of ∂V. Suppose this decomposition is compatible with Φ, V admits a smooth confoliation ξ_V, and W admits a $C^{1,0}$ foliation $\mathcal{F}_W$. Suppose that U is smoothly foliated with a foliation $\mathcal{F}_U$ which smoothly agrees with $\mathcal{F}_W$ where they meet. We say that (V, ξ_V) is Φ–*compatible* with $(W, \mathcal{F}_W)$, and that M admits a positive $(\xi_V, \mathcal{F}_W, \Phi)$ decomposition, if the following are satisfied:

(1) ξ_V and $\mathcal{F}_W$ are (positively) transverse to Φ on their domains of definition,
(2) each of $\mathcal{F}_W$ and ξ_V is tangent to $\partial_h V$,
(3) $\xi_V = T\mathcal{F}_U$ on $\overline{U} \cap V$,
(4) ξ_V is a contact structure on $V \setminus U$, and
(5) $\chi_{\xi_V} < \chi_{T\mathcal{F}_W}$ on $(\partial_v V)\setminus\overline{U}$, when viewed from outside W.

A foliation $\mathcal{F}_W$ is *V–transitive* if every point in W can be connected by a path in a leaf of $\mathcal{F}$ to a point of V.

Let $V_{\gamma_i}(\tau_i)$ be the spanning collection of attracting neighborhoods constructed in Theorem 6.2. Then M can be decomposed by setting $V = V_{\gamma_1}(\tau_1) \cup \cdots \cup V_{\gamma_n}(\tau_n)$ and letting W be the closure of the complement of V. Since V is the union of a spanning collection of neighborhoods, $\mathcal{F}'_W$, the restriction of $\mathcal{F}'$ to W, is V–transitive.

The confoliation ξ_V constructed in Corollary 7.10 satisfies conditions (1)–(5), and we can therefore apply Theorem 6.10 of [24].

Theorem 9.2 [24, Theorem 6.10] *If M admits a positive $(\xi_V, \mathcal{F}_W, \Phi)$ decomposition such that $\mathcal{F}_W$ is V–transitive and ξ_V is ϵ C^0 close to $\mathcal{F}_W \cap \partial_v W$, then M admits a smooth positive contact structure ξ^+ which agrees with ξ_V on V and is ϵ C^0 close to $\mathcal{F}_W$ on W. The analogous result holds if M admits a negative $(\xi_{V'}, \mathcal{F}_{W'}, \Phi)$ decomposition, yielding a smooth negative contact structure ξ^-. If M admits both a positive $(\xi_V, \mathcal{F}_W, \Phi)$ decomposition and a negative $(\xi_{V'}, \mathcal{F}_{W'}, \Phi)$ decomposition, then these contact structures (M, ξ^+) and $(-M, \xi^-)$ are weakly symplectically fillable and universally tight.*

Our main theorem can now be proved.

Theorem 1.2 *Any taut transversely oriented $C^{1,0}$ foliation on a closed oriented 3–manifold $M \neq S^1 \times S^2$ can be C^0 approximated by both a positive ξ^+ and a negative ξ^- smooth contact structure. These contact structures (M, ξ^+) and $(-M, \xi^-)$ are weakly symplectically fillable and universally tight.*

Proof of Theorem 1.2 Given the earlier results in this paper on approximating foliations by smoother foliations, the methods developed for introducing holonomy in minimal sets, and the construction of approximating contact structures in attracting holonomy neighborhoods, Theorem 1.2 follows directly from Theorem 9.2. Since this is a long line of implications and constructions, we assemble and summarize the steps now.

The first step is to show that it suffices to restrict attention to the case that $\mathcal{F}$ is a transversely oriented $C^{\infty,0}$ foliation which is not a fibering and whose every minimal set contains a leaf which is not homeomorphic to $\mathbb{R}^2$.

Consider first the case that $\mathcal{F}$ is a transversely oriented $C^{1,0}$ foliation on M. By Theorem 2.10, $\mathcal{F}$ can be C^0 approximated by a transversely oriented $C^{\infty,0}$ foliation. Next, if $\mathcal{F}$ contains a minimal set all of whose leaves are planes, then by Corollary 8.16, it can be C^0 approximated by a smooth fibering. Finally, if $\mathcal{F}$ is a $C^{\infty,0}$ fibering, then by Theorem 5.2, it can be C^0 approximated by a transversely oriented $C^{\infty,0}$ foliation which is obtained by Denjoy blow up.

If $\mathcal{F}$ is minimal, set $\Lambda_1 = \mathcal{F}$. Otherwise, let $\Lambda_1, \dots, \Lambda_r$ denote the exceptional minimal sets of $\mathcal{F}$, and let $[L_1], \dots, [L_s]$ denote the isotopy classes of compact leaves of $\mathcal{F}$. Apply Corollary 4.12 to obtain a spanning collection of pairwise disjoint holonomy neighborhoods $V'_{\gamma_1}(\tau_1, A_1), \dots, V'_{\gamma_{r+s}}(\tau_{r+s}, A_{r+s})$. Let V' denote their union.

For each i with $1 \le i \le n$, let $R'_i = R_{\gamma_i}(\sigma_i, A_i)$, and set $R' = \bigcup_i R_i$. For each i with $1 \le i \le n$, fix a smooth open neighborhood $N_{R'_i}$ of R'_i in V'_i. Choose each $N_{R'_i}$ small enough that its closure, $\overline{N}_{R'_i}$, is a closed regular neighborhood of R'_i. Let $N_{R'}$ denote the union of the $N_{R'_i}$.

By Lemma 3.11, $\mathcal{F}$ can be C^0 approximated by a transversely oriented $C^{\infty,0}$ foliation which is strongly (V', P)–compatible for some choice of product neighborhood $(P, \mathcal{P})$ of $(V'; N_{R'})$. Hence, it suffices to restrict attention to the case that $\mathcal{F}$ is a transversely oriented, strongly (V', P)–compatible, $C^{\infty,0}$ foliation which is not a fibering and whose every minimal set contains a leaf which is not homeomorphic to $\mathbb{R}^2$. We now do so. In particular, $\mathcal{F} = \mathcal{P}$ on $N_{R'}$ and $\mathcal{F}$ is x–invariant in the (x, y, z) coordinates given by P.

Put the product metric on each component $P_i = [-1, 1] \times S^1 \times [-1, 1]$ of P, as described in Definition 3.12, and let g_0 denote the resulting metric on P. Fix a Riemannian metric $g = g(P)$ on M which restricts to g_0 on P. Fix $\epsilon > 0$.

Now apply Theorem 6.2 to obtain a transversely oriented $C^{\infty,0}$ foliation $\mathcal{G}$ that is ϵ C^0 close to $\mathcal{F}$, V'–compatible with $\mathcal{F}$, and strongly (V', P)–compatible, and a finite set of pairwise disjoint attracting neighborhoods

$$V_{\gamma_1}(\tau_1, A_1), \dots, V_{\gamma_m}(\tau_m, A_m),$$

with $m \ge n$, for $\mathcal{G}$ such that the V_{γ_i} are ϵ–flat, ϵ–horizontal, and if V denotes their union, $\mathcal{G}$ is V–transitive.

By Theorem 7.2, there are a regular neighborhood $N_v \subset V$ of the vertical edges of ∂Q in V and smooth approximating confoliations $\xi_V^{\pm}$ defined on V where the main properties are

(1) $\xi_V^{\pm} = T\mathcal{G}$ on $\overline{N_h \cup N_v}$ and is contact at all other points of V,

(2) ξ_V^+ dominates $\mathcal{G}$ along $\partial_v V$, with the domination strict outside $\overline{N_h \cup N_v}$, and

(3) ξ_V^- is dominated by $\mathcal{G}$ along $\partial_v V$, with the domination strict outside $\overline{N_h \cup N_v}$.

Let W denote the closure of the complement of V and let $\mathcal{G}_W$ denote the restriction of $\mathcal{G}$ to W. Applying Theorem 9.2 to the $(\xi_V^+, \mathcal{G}_W, \Phi)$ decomposition of M yields a positive contact structure ξ^+ on M which is $O(\epsilon)$ close to $\mathcal{G}$. By symmetry, there is a negative contact structure ξ^- on M which is $O(\epsilon)$ close to $\mathcal{G}$. When $\mathcal{F}$ is taut, these contact structures are weakly symplectically fillable and universally tight. Since $\mathcal{G}$ is ϵ close to $\mathcal{F}$, each of $\xi^\pm$ is $O(\epsilon)$ close to $\mathcal{F}$. □

References

[1] **J Bowden**, *Approximating C^0–foliations by contact structures*, Geom. Funct. Anal. 26 (2016) 1255–1296 MR

[2] **D Calegari**, *Leafwise smoothing laminations*, Algebr. Geom. Topol. 1 (2001) 579–585 MR

[3] **C Camacho**, **A Lins Neto**, *Geometric theory of foliations*, Birkhäuser, Boston (1985) MR

[4] **A Candel**, **L Conlon**, *Foliations, I*, Graduate Studies in Mathematics 23, Amer. Math. Soc., Providence, RI (2000) MR

[5] **O T Dasbach**, **T Li**, *Property P for knots admitting certain Gabai disks*, Topology Appl. 142 (2004) 113–129 MR

[6] **C Delman**, **R Roberts**, *Alternating knots satisfy Strong Property P*, Comment. Math. Helv. 74 (1999) 376–397 MR

[7] **P R Dippolito**, *Codimension one foliations of closed manifolds*, Ann. of Math. 107 (1978) 403–453 MR

[8] **Y M Eliashberg**, **W P Thurston**, *Confoliations*, University Lecture Series 13, Amer. Math. Soc., Providence, RI (1998) MR

[9] **W Floyd**, **U Oertel**, *Incompressible surfaces via branched surfaces*, Topology 23 (1984) 117–125 MR

[10] **D Gabai**, *Foliations and the topology of 3–manifolds*, J. Differential Geom. 18 (1983) 445–503 MR

[11] **D Gabai**, *Foliations and genera of links*, Topology 23 (1984) 381–394 MR

[12] **D Gabai**, *Detecting fibred links in S^3*, Comment. Math. Helv. 61 (1986) 519–555 MR

[13] **D Gabai**, *Genera of the alternating links*, Duke Math. J. 53 (1986) 677–681 MR

[14] **D Gabai**, *Foliations and the topology of 3–manifolds, II*, J. Differential Geom. 26 (1987) 461–478 MR

[15] **D Gabai**, *Foliations and the topology of 3–manifolds, III*, J. Differential Geom. 26 (1987) 479–536 MR

[16] **D Gabai**, *Foliations and 3–manifolds*, from "Proceedings of the International Congress of Mathematicians, I" (I Satake, editor), Math. Soc. Japan, Tokyo (1991) 609–619 MR

[17] **D Gabai**, *Taut foliations of 3–manifolds and suspensions of S^1*, Ann. Inst. Fourier (Grenoble) 42 (1992) 193–208 MR

[18] **D Gabai**, *Problems in foliations and laminations*, from "Geometric topology" (W H Kazez, editor), AMS/IP Stud. Adv. Math. 2, Amer. Math. Soc., Providence, RI (1997) 1–33 MR

[19] **D Gabai**, *Combinatorial volume preserving flows and taut foliations*, Comment. Math. Helv. 75 (2000) 109–124 MR

[20] **D Gabai**, **U Oertel**, *Essential laminations in 3–manifolds*, Ann. of Math. 130 (1989) 41–73 MR

[21] **G Hector**, **U Hirsch**, *Introduction to the geometry of foliations, B: Foliations of codimension one*, Aspects of Mathematics E3, Friedr. Vieweg & Sohn, Braunschweig (1983) MR

[22] **H Imanishi**, *On the theorem of Denjoy–Sacksteder for codimension one foliations without holonomy*, J. Math. Kyoto Univ. 14 (1974) 607–634 MR

[23] **T Kalelkar**, **R Roberts**, *Taut foliations in surface bundles with multiple boundary components*, Pacific J. Math. 273 (2015) 257–275 MR

[24] **W H Kazez**, **R Roberts**, *Approximating $C^{1,0}$–foliations*, from "Interactions between low-dimensional topology and mapping class groups" (R I Baykur, J Etnyre, U Hamenstädt, editors), Geom. Topol. Monogr. 19, Geom. Topol. Publ. (2015) 21–72 MR

[25] **W H Kazez**, **R Roberts**, *$C^{1,0}$ foliation theory*, preprint (2016) arXiv

[26] **W H Kazez**, **R Roberts**, *Taut foliations*, preprint (2016) arXiv To appear in Comm. Anal. Geom.

[27] **J M Lee**, *Introduction to smooth manifolds*, Graduate Texts in Mathematics 218, Springer (2003) MR

[28] **T Li**, *Commutator subgroups and foliations without holonomy*, Proc. Amer. Math. Soc. 130 (2002) 2471–2477 MR

[29] **T Li**, *Laminar branched surfaces in 3–manifolds*, Geom. Topol. 6 (2002) 153–194 MR

[30] **T Li**, *Boundary train tracks of laminar branched surfaces*, from "Topology and geometry of manifolds" (G Matić, C McCrory, editors), Proc. Sympos. Pure Math. 71, Amer. Math. Soc., Providence, RI (2003) 269–285 MR

[31] **T Li**, **R Roberts**, *Taut foliations in knot complements*, Pacific J. Math. 269 (2014) 149–168 MR

[32] **E E Moise**, *Affine structures in 3–manifolds, V: The triangulation theorem and Hauptvermutung*, Ann. of Math. 56 (1952) 96–114 MR

[33] **U Oertel**, *Incompressible branched surfaces*, Invent. Math. 76 (1984) 385–410 MR

[34] **U Oertel**, *Measured laminations in 3–manifolds*, Trans. Amer. Math. Soc. 305 (1988) 531–573 MR

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

[36] **R Roberts**, *Constructing taut foliations*, Comment. Math. Helv. 70 (1995) 516–545 MR

[37] **R Roberts**, *Taut foliations in punctured surface bundles, I*, Proc. London Math. Soc. 82 (2001) 747–768 MR

[38] **R Roberts**, *Taut foliations in punctured surface bundles, II*, Proc. London Math. Soc. 83 (2001) 443–471 MR

[39] **R Sacksteder**, *Foliations and pseudogroups*, Amer. J. Math. 87 (1965) 79–102 MR

[40] **R Sacksteder**, **J Schwartz**, *Limit sets of foliations*, Ann. Inst. Fourier (Grenoble) 15 (1965) 201–213 MR

[41] **L C Siebenmann**, *Deformation of homeomorphisms on stratified sets, I*, Comment. Math. Helv. 47 (1972) 123–136 MR

[42] **D Tischler**, *On fibering certain foliated manifolds over* S^1, Topology 9 (1970) 153–154 MR

[43] **T Vogel**, *Uniqueness of the contact structure approximating a foliation*, preprint (2013) arXiv

[44] **R F Williams**, *Expanding attractors*, Inst. Hautes Études Sci. Publ. Math. 43 (1974) 169–203 MR

Department of Mathematics, University of Georgia
Athens, GA, United States

Department of Mathematics, Washington University
St Louis, MO, United States

will@math.uga.edu, roberts@wustl.edu

Proposed: Cameron Gordon
Seconded: Yasha Eliashberg, András I Stipsicz

Received: 3 January 2016
Accepted: 30 January 2017

Geometry & Topology 21 (2017) 3659–3758

Boundaries and automorphisms of hierarchically hyperbolic spaces

MATTHEW GENTRY DURHAM
MARK F HAGEN
ALESSANDRO SISTO

Hierarchically hyperbolic spaces provide a common framework for studying mapping class groups of finite-type surfaces, Teichmüller space, right-angled Artin groups, and many other cubical groups. Given such a space $\mathcal{X}$, we build a bordification of $\mathcal{X}$ compatible with its hierarchically hyperbolic structure.

If $\mathcal{X}$ is proper, eg a hierarchically hyperbolic group such as the mapping class group, we get a compactification of $\mathcal{X}$; we also prove that our construction generalizes the Gromov boundary of a hyperbolic space.

In our first main set of applications, we introduce a notion of geometrical finiteness for hierarchically hyperbolic subgroups of hierarchically hyperbolic groups in terms of boundary embeddings.

As primary examples of geometrical finiteness, we prove that the natural inclusions of finitely generated Veech groups and the Leininger–Reid combination subgroups extend to continuous embeddings of their Gromov boundaries into the boundary of the mapping class group, both of which fail to happen with the Thurston compactification of Teichmüller space.

Our second main set of applications are dynamical and structural, built upon our classification of automorphisms of hierarchically hyperbolic spaces and analysis of how the various types of automorphisms act on the boundary.

We prove a generalization of the Handel–Mosher "omnibus subgroup theorem" for mapping class groups to all hierarchically hyperbolic groups, obtain a new proof of the Caprace–Sageev rank-rigidity theorem for many $\mathrm{CAT}(0)$ cube complexes, and identify the boundary of a hierarchically hyperbolic group as its Poisson boundary; these results rely on a theorem detecting *irreducible axial* elements of a group acting on a hierarchically hyperbolic space (which generalize pseudo-Anosov elements of the mapping class group and rank-one isometries of a cube complex not virtually stabilizing a hyperplane).

20F65; 20F67, 30F60

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3659

Introduction

The class of hierarchically hyperbolic spaces (HHSs) was introduced by Behrstock, Hagen and Sisto [5], and they gave a streamlined definition in [6], to provide a common framework for studying cubical groups and mapping class groups of surfaces. The definition was motivated by the observation that, under natural hypotheses, a CAT(0) cube complex is equipped with a collection of projections to hyperbolic spaces obeying rules reminiscent of the hierarchical structure of mapping class groups and projections to curve graphs introduced by Masur and Minsky [59; 60]. The class of HHSs includes the aforementioned spaces (mapping class groups and many CAT(0) cube complexes, including all universal covers of compact special cube complexes), along with Gromov-hyperbolic spaces, Teichmüller space with any of the usual metrics, and many others; see Behrstock, Hagen and Sisto [5; 6; 7] for an account of the current scope of the theory.

Much of the utility of HHSs comes from the fact that many features of Gromov-hyperbolic spaces have natural generalizations in the HHS world. Since one of the most useful objects associated to a hyperbolic space is its Gromov boundary, we provide here a generalization of the Gromov boundary to hierarchically hyperbolic spaces. The boundary of a hierarchically hyperbolic space is inspired by various boundaries associated to the salient examples of HHSs, eg the simplicial boundary of a CAT(0) cube complex and the Thurston compactification of Teichmüller space, projective measured lamination space $\mathbb{P}\mathcal{ML}(S)$.

Just as the Gromov boundary does for hyperbolic spaces and groups, the HHS boundary provides considerable information about the geometry of an HHS and the dynamics of its automorphisms; our aim in this paper is to explore some of these properties.

Introduction to HHSs

We first briefly and softly recall the HHS theory. A *hierarchically hyperbolic space* is a pair $(\mathcal{X}, \mathfrak{S})$ equipped with some additional data: $\mathcal{X}$ is a quasigeodesic metric space and $\mathfrak{S}$ is an index set equipped with a partial order $\sqsubseteq$, called *nesting*, with a unique maximal element S. There is also an *orthogonality* relation on $\mathfrak{S}$; when $\mathfrak{S}$ is the set of essential subsurfaces of a surface S, up to isotopy, orthogonality is just disjointness. We often call elements of $\mathfrak{S}$ *domains*.

Each $U \in \mathfrak{S}$ is equipped with a uniformly hyperbolic space $\mathcal{C}U$ and a coarse map $\pi_U \colon \mathcal{X} \to \mathcal{C}U$. There are also *relative projections* ρ^U_V, which are coarse maps $\mathcal{C}U \to \mathcal{C}V$ defined unless U and V are orthogonal. In the case where $\mathcal{X}$ is the marking complex of the surface S and $\mathfrak{S}$ is the set of subsurfaces of S, the associated hyperbolic spaces are the curve graphs of these subsurfaces and the projections are subsurface projections. We impose other rules reminiscent of the hierarchical structure of the mapping class group; see Definition 1.1.

The *distance formula* is crucial: for any $x, y \in \mathcal{X}$, the distance $\mathsf{d}_{\mathcal{X}}(x, y)$ differs, up to bounded multiplicative and additive error, from the sum of the distances

$$\mathsf{d}_{\mathcal{C}U}(\pi_U(x), \pi_U(y))$$

as $U \in \mathfrak{S}$ varies over those domains where that distance exceeds some predefined threshold; see Behrstock, Hagen and Sisto [6].

Just as quasiconvexity is vital to the study of hyperbolic spaces, *hierarchical quasiconvexity* is central in the study of HHSs. Roughly, $\mathcal{Y} \subseteq \mathcal{X}$ is hierarchically quasiconvex if $\pi_U(\mathcal{Y})$ is uniformly quasiconvex for each $U \in \mathfrak{S}$, and any point in $\mathcal{X}$ projecting under π_U close to $\pi_U(\mathcal{Y})$ for each U must lie close (in $\mathcal{X}$) to $\mathcal{Y}$. The fundamental example of a hierarchically quasiconvex subspaces is the *standard product region* P_U associated to each $U \in \mathfrak{S}$. Roughly, the subspace P_U consists of those points $x \in \mathcal{X}$ where $\pi_V(x)$ is close to ρ^U_V for any $V \in \mathfrak{S}$ that is not orthogonal to, or nested in, V. The factor of P_U obtained by fixing, in addition, the projections to domains orthogonal to U (and allowing movement in domains nested in U) is denoted by F_U, and the other factor is E_U. A familiar example here is the region of Teichmüller space with the Teichmüller metric where the boundary curves of some subsurface U are short: Minsky [61] proved that these so-called thin parts are quasiisometric to products of the Teichmüller spaces of the complementary subsurfaces, one of which is U.

What's needed from [5; 6] Behrstock, Hagen and Sisto [6] is the main foundational paper in the theory of HHSs. In the current paper, we use most of the background material developed in [6], with the notable exception of the combination theorems. In particular, we use the main definition of HHSs (which is equivalent to, but much simpler than, the original definition from [5]), the realization theorem, the distance formula, and the existence of hierarchy paths. The fact that mapping class groups are hierarchically hyperbolic groups, which is crucial for our applications to Veech and Leininger–Reid subgroups in Section 5, could be deduced from Behrstock [3], Behrstock, Kleiner, Minsky and Mosher [8] and Masur and Minsky [59; 60], but is also given a streamlined proof by Behrstock, Hagen and Sisto [6, Section 11].

From Behrstock, Hagen and Sisto [5], we need the acylindricity result (Theorem 14.3) and, for the purposes of Section 10, the HHS structure on CAT(0) cube complexes. We note that the acylindricity result from [5] is independent of the other HHS results in that paper.

Finally, the recent paper Behrstock, Hagen and Sisto [7] is completely independent of this one.[1]

The boundary

Consider an HHS $(\mathcal{X},\mathfrak{S})$. Since any two points of $\mathcal{X}$ are joined by a *hierarchy path* — a uniform quasigeodesic projecting to a uniform unparametrized quasigeodesic in $\mathcal{C}U$ for each $U\in\mathfrak{S}$ (see [6]) — a natural approach to constructing a boundary is to imitate the construction of the Gromov boundary, or the visual boundary of a CAT(0) space: boundary points would be asymptotic classes of "hierarchy rays" emanating from a fixed basepoint, and one might imagine topologizing this set by defining two boundary points to be close if the corresponding rays stay close "for a long time".

The boundary construction is motivated by this intuition. Given a hierarchy ray $\gamma\colon\mathbb{N}\to\mathcal{X}$, one first observes that the set of $U\in\mathfrak{S}$ for which $\pi_U\circ\gamma$ is unbounded is a pairwise-orthogonal collection — γ either spends a bounded amount of time in each standard product region, or γ wanders permanently into the (coarse) intersection of several standard product regions. Accordingly, the underlying set of the boundary $\partial(\mathcal{X},\mathfrak{S})$ is the set of formal linear combinations $p=\sum_{U\in\mathfrak{U}}a_U p_U$, where $\mathfrak{U}\subset\mathfrak{S}$ (the *support* of p) is a pairwise-orthogonal set, each p_U is a point in the Gromov boundary of $\mathcal{C}U$, each $a_U\in(0,1]$, and $\sum_U a_U=1$.

Regarding each $\partial\mathcal{C}U$ as a discrete set, the above construction yields a (highly disconnected, locally infinite) simplicial complex. The "rank-one hierarchy rays" — ie the

[1]The picture on Hagen's website shows the current state of the theory, indicating the main concepts and results and their interdependencies.

points of $\partial\mathcal{C}S$ — correspond to isolated 0–simplices, while the standard product regions contribute boundary subcomplexes isomorphic to simplicial joins. This complex is a kind of "Tits boundary" for $(\mathcal{X},\mathfrak{S})$. The actual boundary we define is related to this complex in much the same way that the visual boundary of a CAT(0) space is related to the Tits boundary; we define the boundary $\partial(\mathcal{X},\mathfrak{S})$ by imposing a coarser topology, described in Section 2. (When the context is clear, we denote $\partial(\mathcal{X},\mathfrak{S})$ by $\partial\mathcal{X}$, being mindful that this space depends, as far as we know, on the particular HHS structure $\mathfrak{S}$.)

The resulting space $\overline{\mathcal{X}} = \mathcal{X} \cup \partial\mathcal{X}$ is Hausdorff and separable; $\partial\mathcal{X}$ is a closed subset and $\mathcal{X}$ is dense (Proposition 2.17). Moreover, the Gromov boundary $\partial\mathcal{C}U$ embeds in $\partial(\mathcal{X},\mathfrak{S})$, in the obvious way, for each $U \in \mathfrak{S}$, by Theorem 4.3. Crucially:

Theorem 3.4 (compactness) *Let $(\mathcal{X},\mathfrak{S})$ be a hierarchically hyperbolic space with $\mathcal{X}$ proper. Then $\overline{\mathcal{X}}$ is compact.*

The definition of $\partial(\mathcal{X},\mathfrak{S})$ is given strictly in terms of $\mathfrak{S}$ and the accompanying hyperbolic spaces and projections; the standing assumption that $(\mathcal{X},\mathfrak{S})$ is *normalized* — each π_U is coarsely surjective — connects the boundary to the space $\mathcal{X}$ by ensuring that $\mathcal{X}$ is dense in $\overline{\mathcal{X}}$. Even so, it is not clear whether the homeomorphism type of $\partial(\mathcal{X},\mathfrak{S})$ depends on the particular choice of HHS structure:

Question 1 *Let $(\mathcal{X},\mathfrak{S})$ be a hierarchically hyperbolic space and let $(\mathcal{X},\mathfrak{S}')$ be a different hierarchically hyperbolic structure on the same space. Does the identity map $\mathcal{X} \to \mathcal{X}$ extend to a map $\mathcal{X} \cup \partial(\mathcal{X},\mathfrak{S}) \to \mathcal{X} \cup \partial(\mathcal{X},\mathfrak{S}')$ which restricts to a homeomorphism of boundaries?*

A positive answer to Question 1 would stand in contrast to the situation for CAT(0) spaces. For example, the right-angled Artin group A, presented by a path of length 3, famously has the property that the universal cover $\widetilde{X}$ of the Salvetti complex can be endowed with different CAT(0) metrics (obtained by perturbing angles in the 2–cells) with nonhomeomorphic visual boundaries; see Croke and Kleiner [22]. On the other hand, $\widetilde{X}$ admits a hierarchically hyperbolic structure $(\widetilde{X},\mathfrak{S})$ coming from the cubical structure of $\widetilde{X}$ (with no dependence on the CAT(0) metric). Perturbing the CAT(0) metric within its quasiisometry type does not change the HHS structure (and hence the HHS boundary), so the HHS boundary is in a sense more "canonical" than the visual boundary in this example (and indeed for all CAT(0) cube complexes with *factor systems*, which we discuss in more detail below).

Automorphisms and their actions on the boundary

An *automorphism* of $(\mathcal{X},\mathfrak{S})$ is a bijection $g\colon \mathfrak{S} \to \mathfrak{S}$ and an isometry $\mathcal{C}U \to \mathcal{C}g(U)$ for each $U \in \mathfrak{S}$ which satisfy certain compatibility conditions. The distance formula ensures that automorphisms induce uniform quasiisometries of $\mathcal{X}$, so the group $\mathrm{Aut}(\mathfrak{S})$ of automorphisms uniformly quasiacts by (uniform) quasiisometries on $\mathcal{X}$. The (quasi)action of $\mathrm{Aut}(\mathfrak{S})$ on $\mathcal{X}$ extends to an action on $\overline{\mathcal{X}}$ that restricts to an action by homeomorphisms on $\partial\mathcal{X}$ (Corollary 6.1).

In one of the main cases of interest, $\mathcal{X}$ is a Cayley graph of a finitely generated group G, and the action of G on itself by left multiplication corresponds to an action on $(G,\mathfrak{S})$ by HHS automorphisms. In this situation, if the action on $\mathfrak{S}$ is cofinite, then $(G,\mathfrak{S})$ is a *hierarchically hyperbolic group structure*; if a group G admits a hierarchically hyperbolic group structure, then G is a *hierarchically hyperbolic group* (HHG). The archetypal hierarchically hyperbolic group is the mapping class group of a connected, oriented surface of finite type [6, Section 11]. Other examples include many cubical groups [5], many graphs of hierarchically hyperbolic groups [6], and certain quotients of hierarchically hyperbolic groups [7]. If $(G,\mathfrak{S})$ is a hierarchically hyperbolic group, then the isometric action of G on itself by left multiplication extends to an action by homeomorphisms on $\overline{G}$ (Corollary 6.2). We describe in detail below our results regarding the dynamics and structure of groups of automorphisms.

Embeddings of subspace boundaries and geometrical finiteness

A desirable property of a boundary is that inclusions of subspaces that are "convex" in an appropriate sense induce embeddings of boundaries with closed images. In Section 5, we show that hierarchically quasiconvex subspaces of $\mathcal{X}$, which admit their own natural HHS structures [6], have this property: if $\mathcal{Y}\subset\mathcal{X}$ is hierarchically quasiconvex, then $\mathcal{Y}$ has a limit set in $\partial\mathcal{X}$ which is homeomorphic to $\partial\mathcal{Y}$ with the HHS structure inherited from $\mathcal{X}$. In fact, Theorem 5.6 provides more, by giving natural conditions on maps between HHSs ensuring that they extend continuously to the HHS boundary. This motivates the following definition:

Definition 2 (geometrical finiteness) We say a hierarchically hyperbolic subgroup H of a hierarchically hyperbolic group G is *geometrically finite* if the natural inclusion $\iota\colon H\hookrightarrow G$ extends continuously to an H–equivariant embedding $\partial\iota\colon \partial H\hookrightarrow\partial G$.

In what follows, we will be interested in developing this notion and establishing examples in the context of the mapping class group of a finite-type surface.

Comparison of the mapping class group boundary with $\mathbb{P}\mathcal{ML}(S)$

The archetypal hierarchically hyperbolic group is the mapping class group $\mathcal{MCG}(S)$ of a connected, oriented surface S of finite type. The hierarchically hyperbolic structure is provided by results of Aougab [1], Behrstock [3], Behrstock, Kleiner, Minsky and

Mosher [8], Bowditch [12], Clay, Rafi and Schleimer [21], Hensel, Przytycki and Webb [43], Mangahas [55], Masur and Minsky [59; 60], Przytycki and Sisto [66] and Webb [74] and is discussed in detail in Section 11 of Behrstock, Hagen and Sisto [6]. Roughly, $\mathfrak{S}$ is the set of essential subsurfaces of S, up to isotopy, $\mathcal{C}U$ is the curve graph of U for each $U \in \mathfrak{S}$, and projections are usual subsurface projections.

Traditionally, $\mathcal{MCG}(S)$ has been studied via its action on Teichmüller space $\mathcal{T}(S)$ with its Thurston compactification by $\mathbb{P}\mathcal{ML}(S)$. This approach has been fruitful especially when considering subgroups of $\mathcal{MCG}(S)$ defined via flat or hyperbolic geometry. Nonetheless, the $\mathcal{MCG}(S)$ action on $\mathcal{T}(S)$ is not cocompact and the orbits of many subgroups (in fact, any with Dehn twists) are distorted in $\mathcal{T}(S)$, which make $\mathcal{T}(S)$ imperfect for studying the coarse geometry of $\mathcal{MCG}(S)$ and its subgroups.

The situation is further complicated when one attempts to extend the $\mathcal{MCG}(S)$ action on $\mathcal{T}(S)$ to its various boundaries. Teichmüller geodesics are unique and thus geodesic rays based at a point form a natural visual compactification of $\mathcal{T}(S)$, but Kerckhoff [49] proved that it is basepoint dependent and thus the $\mathcal{MCG}(S)$ action fails to extend continuously. While Thurston [72] defined a compactification via $\mathbb{P}\mathcal{ML}(S)$ to which the $\mathcal{MCG}(S)$ action does extend continuously, Thurston's compactification is defined via hyperbolic geometry and the Teichmüller metric is defined via flat geometry, which leads to an incoherence between the internal geometry and its asymptotics in $\mathbb{P}\mathcal{ML}(S)$; see Brock, Leininger, Modami and Rafi [16], Chaika, Masur and Wolf [19], Leininger, Lenzhen and Rafi [51], Lenzhen [53] and Masur [58].

The boundary $\partial(\mathcal{MCG}(S), \mathfrak{S})$ provides the first compactification of $\mathcal{MCG}(S)$, so the action of $\mathcal{MCG}(S)$ on itself by left multiplication extends to a continuous action on the boundary with the dynamical properties we discuss below (see also Section 6). While many of these dynamical properties were originally proven via the $\mathcal{MCG}(S)$–action on $\mathcal{T}(S)$ with its Thurston compactification, many of the pathologies described above vanish in our construction, as we discuss presently.

On geometrically finite subgroups of $\mathcal{MCG}(S)$

Problem 5 of Hamenstädt [41] and Section 6 of Mosher [62] regard the development of a notion of geometrical finiteness for subgroups of $\mathcal{MCG}(S)$. Mosher suggests a definition that requires an external proper hyperbolic space X on which the candidate subgroup acts with a collection of cusp subgroups in some appropriate sense; geometric finiteness would then require that X and ∂X embed quasiisometrically in $\mathcal{T}(S)$ and continuously in $\mathbb{P}\mathcal{ML}(S)$, respectively. Masur's theorem makes it unreasonable to expect a simultaneous continuous embedding $X \cup \partial X \to \mathcal{T}(S) \cup \mathbb{P}\mathcal{ML}(S)$.

We will argue that replacing $\mathcal{T}(S) \cup \mathbb{P}\mathcal{ML}(S)$ with $\mathcal{MCG}(S) \cup \partial\mathcal{MCG}(S)$ as in Definition 2 generates a robust theory of geometrical finiteness. In particular, we prove:

Theorem 3 *Suppose that $H < \mathcal{MCG}(S)$ is one of the following:*

(1) *The standard embedding of $\mathcal{MCG}(Y)$ for some proper subsurface $Y \subset S$.*

(2) *Convex cocompact in the sense of Farb and Mosher [31].*

(3) *A finitely generated Veech group.*

(4) *A Leininger–Reid combination subgroup [52].*

Then H is a geometrically finite subgroup of $\mathcal{MCG}(S)$.

Hence geometrical finiteness generalizes convex cocompactness for subgroups of $\mathcal{MCG}(S)$ to a broader class of groups. Theorem 3(a) is proven in Theorem 5.11 and Theorem 3(b) is Theorem 5.12. We discuss presently the Veech and Leininger–Reid examples in more detail.

Veech and Leininger–Reid combinations subgroups For Mosher (see Problem 6.1 of [62]), the main test cases for a definition of geometrical finiteness for subgroups of mapping class groups are finitely generated Veech groups and the Leininger–Reid subgroups. It is worth noting that while the former are explicitly defined via flat geometry and the latter somewhat less so, the aforementioned coherence pathologies between the Teichmüller geometry and the Thurston compactification give an obstruction to considering embeddings of natural boundaries associated to them into $\mathbb{P}\mathcal{ML}(S)$. We prove that this obstruction disappears with $\partial\mathcal{MCG}(S)$. We now briefly give some background.

Given a holomorphic quadratic differential q on S, there is an associated copy of $\mathbb{H}^2$ called a Teichmüller disk, $\mathrm{TD}(q)$, which is a convex subset of $\mathcal{T}(S)$. The stabilizer of $\mathrm{TD}(q)$ in $\mathcal{MCG}(S)$ is $\mathrm{Aff}(q)$, those elements with a representative which act by affine homeomorphisms with respect to the flat metric determined by q. A Veech group V is a subgroup of $\mathrm{Aff}(q)$ which acts properly on $\mathrm{TD}(q)$; we consider only finitely generated Veech groups. The visual boundary of $\mathrm{TD}(q)$ is naturally identified by $\mathbb{P}\mathcal{ML}(q)$, which admits a natural embedding in $\mathbb{P}\mathcal{ML}(S)$ that parametrizes the limit set of V in $\mathbb{P}\mathcal{ML}(S)$ — see Kent and Leininger [47] — but a theorem of Masur [58] implies that this embedding does not give an everywhere continuous extension $\mathrm{TD}(q) \cup \mathbb{P}\mathcal{ML}(q) \hookrightarrow \mathcal{T}(S) \cup \mathbb{P}\mathcal{ML}(S)$.

Leininger and Reid [52] construct subgroups of $\mathcal{MCG}(S)$ which are combinations of Veech groups; some are surface groups in which all but one conjugacy class is pseudo-Anosov. The *boundary* of such a surface subgroup is its limit set in $\partial\mathbb{H}^2$. Problem 3.3 of Reid [68] asks if there is a continuous, equivariant embedding of this boundary into $\mathbb{P}\mathcal{ML}(S)$.

While we do not answer this question directly, we do prove something strictly stronger for $\partial\mathcal{MCG}(S)$:

Theorem 5.20 *Let $H < \mathcal{MCG}(S)$ be either a finitely generated Veech or Leininger–Reid subgroup as above. Then the inclusion $H \hookrightarrow \mathcal{MCG}(S)$ extends to a continuous H–equivariant embedding $\partial H \hookrightarrow \partial\mathcal{MCG}(S)$ with closed image. In particular, H is a geometrically finite subgroup of $\mathcal{MCG}(S)$.*

Other candidates for geometrical finiteness Perhaps the next best candidates for geometrically finite subgroups of $\mathcal{MCG}(S)$ are the various right-angled Artin groups constructed by Clay, Leininger and Mangahas [20] and Koberda [50]. These subgroups are HHGs and the former are even known to be quasiisometrically embedded in $\mathcal{MCG}(S)$.

Question 4 *Are the Clay–Leininger–Mangahas and Koberda right-angled Artin subgroups of $\mathcal{MCG}(S)$ geometrically finite? Hierarchically quasiconvex?*[2]

The HHS boundary of Teichmüller space and $\mathbb{P}\mathcal{ML}(S)$ Slight modifications of the above hierarchical structures endow the Teichmüller space, $\mathcal{T}(S)$, with either the Teichmüller or Weil–Petersson metrics, with an HHS structure, as explained in [5; 6] using results of Brock [15], Durham [26] and Eskin, Masur and Rafi [29]; see also Bowditch [14; 13] for closely related results.

Question 5 *How is the HHS boundary $\partial\mathcal{T}(S)$ of $\mathcal{T}(S)$, with the Teichmüller metric and the above HHS structure, related to the projective measured lamination space $\mathbb{P}\mathcal{ML}(S)$?*

In fact, there is a natural map $\mathbb{P}\mathcal{ML}(S) \to \partial\mathcal{T}(S)$ which collapses certain simplices of measures on given laminations to points, while being injective on the set of uniquely ergodic laminations, whose image in $\partial\mathcal{T}(S)$ can be identified with a subset of $\partial\mathcal{C}S \subset \partial\mathcal{T}(S)$. A promising strategy is to attempt to use this map, along with a result of Edwards — see Daverman [24] and Edwards [28] — to prove that $\partial\mathcal{T}(S)$ is homeomorphic to $\mathbb{P}\mathcal{ML}(S)$, ie to $\mathbb{S}^{2\xi(S)-1}$. The missing ingredient is a positive answer to:

Question 6 *Does $\partial\mathcal{T}(S)$ have the* ***disjoint discs property****?*

A metric space M has the disjoint disks property if any two maps $D^2 \to M$ admit arbitrarily small perturbations with disjoint image; the above question makes sense since it is not hard to see, using Proposition 2.17, that $\partial\mathcal{T}(S)$ is metrizable. The difficulty

[2] Since we initially posted this paper, Mousley [63] answered this question negatively.

here involves nonuniquely ergodic laminations, which cause a similar problem to the extensions discussed above related to the Leininger–Reid subgroups.

Another question, subject to much recent study, is about the limit sets of Teichmüller geodesics in Thurston's compactification. The analogous question in our setting is:

Question 7 *What are the limit sets of Teichmüller geodesics in $\partial\mathcal{T}(S)$?*

There are now several constructions of geodesics with limits sets that are bigger than a point — see Brock, Leininger, Modami and Rafi [16], Chaika, Masur and Wolf [19], Leininger, Lenzhen and Rafi [51] and Lenzhen [53] — but these constructions fundamentally depend on the fact that filling minimal laminations can admit simplices of measures, which collapse in $\partial\mathcal{T}(S)$. The geodesics constructed in [16; 19; 51] will have unique limits $\partial\mathcal{T}(S)$ as their asymptotics with respect to $\partial\mathcal{T}(S)$ are determined by their asymptotics in the curve graph $\mathcal{C}S$. On the other hand, the situation becomes more opaque for Teichmüller geodesics with vertical laminations with multiple components. Using work of Rafi [67], one can determine that the coefficients a_Y of the components $Y \subset S$ supporting the potential limits in $\partial\mathcal{T}(S)$ are determined by limits of ratios of the rates of divergence in the various subsurface curve graphs $\mathcal{C}Y$. However, it seems unlikely that these limits of ratios always exist, suggesting that such geodesics need not have unique limits in $\partial\mathcal{T}(S)$.

Dynamical and structural results

Our second main collection of applications of the boundary are about the dynamics of the action on the boundary and the structure of subgroups. In Section 6, we study automorphisms of hierarchically hyperbolic spaces:

Classification of automorphisms Given $f \in \mathrm{Aut}(\mathfrak{S})$, the set $\mathrm{Big}(f)$ of $U \in \mathfrak{S}$ for which $\langle f\rangle \cdot x$ (for some basepoint $x \in \mathcal{X}$) projects to an unbounded set in $\mathcal{C}U$ is a possibly empty finite set of pairwise-orthogonal domains preserved by the action of $\langle f\rangle$ on $\mathfrak{S}$. We classify f according to the nature of $\mathrm{Big}(f)$. First, if $\mathrm{Big}(f) = \varnothing$, then f has bounded orbits in each $\mathcal{C}U$ and hence has bounded orbits in $\mathcal{X}$, by Proposition 6.4; in this case, f is *elliptic*. Second, if $\langle f\rangle \cdot x$ projects to a quasiline in $\mathcal{C}U$ for some $U \in \mathrm{Big}(f)$, then $\langle f\rangle \cdot x$ is a quasiline in $\mathcal{X}$, by Proposition 6.12, and f is *axial*. Otherwise, f is *distorted*.

If $\mathrm{Big}(f) = \{S\}$, then f is *irreducible*, and f is *reducible* otherwise. Perhaps the most important class of HHS automorphisms are irreducible axial automorphisms. In the mapping class group, these are the pseudo-Anosov elements; in a hierarchically hyperbolic cube complex, these are the rank-one elements that do not virtually preserve hyperplanes; see [5] and Hagen [36]. In the case where $(G, \mathfrak{S})$ is a hierarchically hyperbolic group, each irreducible axial element is Morse — this follows from Theorem 6.15 — but the converse does not hold. The question of when irreducible axial elements exist is of major interest later.

Dynamics and fixed points In Section 6.2, we study the dynamics of $f \in \mathrm{Aut}(\mathfrak{S})$ on $\partial\mathcal{X}$. First, we show that irreducible axial automorphisms act as expected:

Proposition 6.18 (north–south dynamics) *If $g \in \mathrm{Aut}(\mathfrak{S})$ is irreducible axial, then g has exactly two fixed points $\lambda_+, \lambda_- \in \partial\mathcal{X}$. Moreover, for any boundary neighborhoods $\lambda_+ \in U_+$ and $\lambda_- \in U_-$, there exists an $N > 0$ such that $g^N(\partial\mathcal{X} - U_-) \subset U_+$.*

In Propositions 6.19 and 6.20, we show that if f is irreducible distorted, then f fixes a unique point $p \in \partial\mathcal{X}$, which is an "attracting fixed point". We also prove analogues of these results for reducible automorphisms (Propositions 6.22 and 6.25).

We then study hierarchically hyperbolic groups. First, we rule out distortion:

Theorem 7.1 (coarse semisimplicity) *If $(G, \mathfrak{S})$ is a hierarchically hyperbolic group, then each $g \in G$ is either elliptic or axial; in fact g is undistorted in each element of* $\mathrm{Big}(g)$.

In the event that G contains irreducible axial elements, we have:

Theorem 6.29 (topological transitivity) *Let $(G, \mathfrak{S})$ be hierarchically hyperbolic with an irreducible axial element and let G be nonelementary. Then any G–orbit in ∂G is dense.*

Below, we will describe when $(G, \mathfrak{S})$ has an irreducible axial element.

Uses of the boundary

We use the boundary, and actions thereon, in numerous ways.

Finding and exploiting irreducible axials In Section 9, we study irreducible axial elements of groups of automorphisms of hierarchically hyperbolic spaces. The setting is an HHS $(\mathcal{X}, \mathfrak{S})$ with $\mathcal{X}$ proper and $\mathfrak{S}$ countable, and we consider a countable subgroup $G \le \mathrm{Aut}(\mathfrak{S})$. This holds, for example, when $\mathcal{X} = G$ is an HHG. The main technical statement is:

Propositions 9.4 and 9.2 (finding irreducible axials) *Suppose that either G acts properly and coboundedly on $\mathcal{X}$ and cofinitely on $\mathfrak{S}$, or G acts with unbounded orbits in $\mathcal{X}$ and no fixed point in $\partial\mathcal{C}S$. Then either G contains an irreducible axial element, or there exists $U \in \mathfrak{S} - \{U\}$ which is fixed by a finite-index subgroup of G.*

These two propositions are proved in tandem. The strategy is to consider probability measures on G and corresponding G–stationary measures on $\partial\mathcal{X}$; the main lemma, Lemma 9.8, shows that, unless G has a finite orbit in $\partial\mathcal{C}S$ or $\mathfrak{S}-\{S\}$, such a measure must be supported on $\partial\mathcal{C}S \subset \partial\mathcal{X}$. In particular, if $\mathcal{C}S$ is bounded, then there must be a finite orbit in $\mathfrak{S}-\{S\}$. We emphasize that, for the above proposition and all of its applications, compactness of the HHS boundary (ie Theorem 3.4) is absolutely vital.

Using the above propositions, we prove:

Theorem 9.15 (HHG Tits alternative) *Let $(G,\mathfrak{S})$ be an HHG and let $H \le G$. Then H either contains a nonabelian free group or is virtually abelian.*

By analyzing supports of global fixed points in the boundary of an HHS, we then prove:

Theorem 9.20 (omnibus subgroup theorem) *Let $(G,\mathfrak{S})$ be a hierarchically hyperbolic group and let $H \le G$. Then there exists an element $g \in H$ with $\mathfrak{A}(H) = \mathrm{Big}(g)$. Moreover, for any $g' \in H$ and each $U \in \mathrm{Big}(g')$, there exists $V \in \mathrm{Big}(g)$ with $U \sqsubseteq V$.*

Here, $\mathfrak{A}(H)$ is the set of domains U on which H has unbounded projection. The theorem we actually prove is more general than the above, but the version stated here is sufficient to imply the omnibus subgroup theorem for mapping class groups, due to Handel and Mosher [42], which they proved as an umbrella theorem for several subgroup structure theorems, including the Tits alternative; see also Mangahas [56] for further discussion.

We also obtain a coarse/HHS version of the rank-rigidity conjecture for CAT(0) spaces:

Theorems 9.13 and 9.14 (coarse rank-rigidity) *Let $(\mathcal{X},\mathfrak{S})$ be an HHS with $\mathcal{X}$ unbounded and proper and $\mathfrak{S}$ countable. Let $G \le \mathrm{Aut}(\mathfrak{S})$ be a countable subgroup and suppose that one of the following holds:*

(1) *G acts essentially on $\mathcal{X}$ with no fixed point in $\partial\mathcal{X}$.*

(2) *G acts properly and coboundedly on $\mathcal{X}$ and cofinitely on $\mathfrak{S}$.*

Then either $(\mathcal{X},\mathfrak{S})$ is a ***product HHS with unbounded factors*** *or there exists an axial element $g \in G$ such that $\mathrm{Big}(g)$ consists of a single domain W, for which $\mathcal{C}U$ is bounded if $U \perp W$.*

Such an element g is a *rank-one automorphism*; all of its quasigeodesic axes of any fixed quality lie in some neighborhood of one another (of radius depending on the quality). The HHS is a *product with unbounded factors* if there exists $U \in \mathfrak{S}$ such that $\mathcal{X}$ coarsely coincides with the standard product region P_U, and each of E_U and F_U is unbounded.

In particular, if $\mathcal{X}$ is any of the cube complexes shown in [5] to be hierarchically hyperbolic (ie those admitting "factor systems"), then our methods allow us to recover the Caprace–Sageev rank-rigidity theorem [18] for $\mathcal{X}$:

Corollary 9.24 (rank-rigidity for many cube complexes) *Let $\mathcal{X}$ be a* CAT(0) *cube complex with a factor system. Let G act on $\mathcal{X}$ and suppose that one of the following holds:*

(1) *$\mathcal{X}$ is unbounded and G acts on $\mathcal{X}$ properly and cocompactly.*

(2) *G acts on $\mathcal{X}$ with no fixed point in $\mathcal{X} \cup \partial_{\triangle}\mathcal{X}$.*

Then $\mathcal{X}$ contains a G–invariant convex subcomplex $\mathcal{Y}$ such that either G contains a rank-one isometry of $\mathcal{Y}$ or $\mathcal{Y} = \mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are unbounded convex subcomplexes.

It is difficult to construct cube complexes without factor systems that satisfy the remaining hypotheses of this theorem. At least in the cocompact case, we believe that our proof works without explicitly hypothesizing the existence of a factor system — see Question A of [6], which asks whether the presence of a geometric group action on a cube complex guarantees that a factor system exists (see Remark 9.25).[3]

Other applications, examples, and questions

The HHS boundary in the cubical case If $\mathcal{X}$ is a CAT(0) cube complex with a factor system $\mathfrak{F}$ (here $\mathfrak{F}$ more properly denotes the set of parallelism classes of elements of the factor system), then the resulting hierarchically hyperbolic structure (which is fundamentally derived from the hyperplanes of $\mathcal{X}$ and how they interact) has a boundary which is, perhaps unsurprisingly, closely related to the *simplicial boundary* $\partial_{\triangle}\mathcal{X}$ introduced by Hagen [36] (which is derived from how certain infinite families of hyperplanes interact). Specifically:

Theorem 10.1 (simplicial and HHS boundaries) *Let $\mathcal{X}$ be a* CAT(0) *cube complex with a factor system $\mathfrak{F}$, and let $(\mathcal{X}, \mathfrak{F})$ be the associated hierarchically hyperbolic structure. There is a topology $\mathcal{T}$ on the simplicial boundary $\partial_{\triangle}\mathcal{X}$ such that:*

(1) *There is a homeomorphism $b\colon (\partial_{\triangle}\mathcal{X}, \mathcal{T}) \to \partial(\mathcal{X}, \mathfrak{F})$.*

(2) *For each component C of the simplicial complex $\partial_{\triangle}\mathcal{X}$, the inclusion $C \hookrightarrow (\partial_{\triangle}\mathcal{X}, \mathcal{T})$ is an embedding.*

In particular, if $\mathfrak{F}$ and $\mathfrak{F}'$ are factor systems on $\mathcal{X}$, then $\partial(\mathcal{X}, \mathfrak{F})$ is homeomorphic to $\partial(\mathcal{X}, \mathfrak{F}')$.

[3] After we initially posted this paper, Hagen and Susse [38] showed that every CAT(0) cube complex with a geometric group action admits a factor system and is thus hierarchically hyperbolic.

This theorem highlights the relationship between the question of when factor systems exist, and when $\mathcal{X}$ is *visible* in the sense that every simplex of the simplicial boundary corresponds to a geodesic ray in $\mathcal{X}$; this is discussed in Remark 10.9.

Detecting splittings and cubulations from the boundary It is not difficult to show, from the definitions and Stallings' theorem [71] on ends of groups, that if $(G, \mathfrak{S})$ is a hierarchically hyperbolic group, then $\partial(G, \mathfrak{S})$ is disconnected if and only if G splits over a finite subgroup.

Question 8 *Can the JSJ splitting of G over slender subgroups* (see Dunwoody and Sageev [25], Fujiwara and Papasoglu [33] and Rips and Sela [69]) *be detected by examining separating spheres in $\partial(G, \mathfrak{S})$, as is the case for hyperbolic groups and splittings over two-ended subgroups* (see Bowditch [10])?

One can also consider producing actions of hierarchically hyperbolic groups on CAT(0) cube complexes other than trees. As usual, this divides into two separate issues, namely detecting a profusion of codimension-1 subgroups and then choosing a finite collection sufficient to produce an action on a cube complex with good finiteness properties. It appears as though $\partial(G, \mathfrak{S})$ can be used to produce a proper action on a cube complex from a sufficiently rich collection of hierarchically quasiconvex codimension-1 subgroups by a method exactly analogous to that used to cubulate various hyperbolic groups by Bergeron and Wise [9]. The main difference is that G does not act as a uniform convergence group on $\partial(G, \mathfrak{S})$; one must replace the space of triples of distinct boundary points by the space of triples $(p, q, r) \in \partial G$ such that any two of p, q and r are *antipodal*, ie joined by a biinfinite hierarchy path.

Question 9 *Let $(G, \mathfrak{S})$ be a hierarchically hyperbolic group. Give conditions on G ensuring that for any antipodal p, $q \in \partial G$, there exists a hierarchically quasiconvex codimension-1 subgroup H such that p and q are in distinct components of ∂gH for some $g \in G$.*

We have not included a detailed discussion of the above "boundary cubulation for HHG" technique in the present paper since there are not yet any applications; these could be provided by an answer to Question 9.

Poisson boundaries and C^*–simplicity In Section 9.8.1, we show that the boundary of an HHG is a topological model for the Poisson boundary:

Theorem 9.26 (Poisson boundary) *Let $(G, \mathfrak{S})$ be an HHG with* $\operatorname{diam} \mathcal{C}S = \infty$, *$\mu$ be a nonelementary probability measure on G with finite entropy and finite first logarithmic moment, and ν the resulting μ–stationary measure on ∂G. Then $(\partial G, \nu)$ is the Poisson boundary for (G, μ).*

In fact, $\partial\mathcal{C}S$ is a model for the Poisson boundary [5], but $\partial(G,\mathfrak{S})$ has the advantage of being compact, while in general $\partial\mathcal{C}S$ is not compact. The space ∂G is a G*–boundary*, ie a compactum on which G acts minimally and proximally. Moreover:

Proposition 10 *The action of G on ∂G is topologically free, ie for each $g \in G-\{1\}$, the set of $p \in \partial\mathcal{X}$ with $gp \neq p$ is dense in $\partial\mathcal{X}$.*

Proof Let $g \in G-\{1\}$, let $q \in \partial G$, and let U be a neighborhood of q. Suppose for a contradiction that g fixes U pointwise. By Proposition 9.4, G contains an irreducible axial element, so by Proposition 6.28, $\partial\mathcal{C}S$ is dense in ∂G, whence, since G is nonelementary, g fixes infinitely many distinct points of $\partial\mathcal{C}S$. If g is reducible axial, then Lemma 6.24 yields a contradiction, since g cannot fix any point in $\partial\mathcal{C}S$ by the lemma. If g is irreducible axial, then g fixes exactly two points in $\partial\mathcal{C}S$, again a contradiction. Otherwise, g is elliptic and hence has finite order and we are done by hypothesis. □

By a result of Kalantar and Kennedy [46, Theorem 1.5], the above proposition gives a new proof that a nonelementary HHG G with $\partial\mathcal{C}S$ unbounded is C^*–simple (ie the reduced C^*–algebra of G is simple) provided finite-order elements have finite fixed-point sets in $\partial\mathcal{C}S$. However, G is known to be C^*–simple under these circumstances, since G is acylindrically hyperbolic [5] and has no finite normal subgroup; see Dahmani, Guirardel and Osin [23].

In light of the HHG structure on cubulated groups discussed above, Theorem 9.26 should be compared to the results of [64], in which Nevo and Sageev construct the Poisson boundary for a cubical group using the Roller boundary of the cube complex.

Outline of this paper

In Section 1, we review hierarchically hyperbolic spaces. In Section 2, we define the HHS boundary. Section 3 is devoted to the proof that proper HHSs have compact boundaries, and in Section 4, we show that the HHS boundary of a hyperbolic HHS is homeomorphic to the Gromov boundary. In Section 5, we discuss continuous extensions of maps between HHSs to the boundary, and consider this phenomenon in the context of Veech and Leininger–Reid subgroups of the mapping class group. Automorphisms of hierarchically hyperbolic structures induce homeomorphisms of the boundary; in Section 6, we classify their automorphisms and study fixed sets and dynamics of the actions of automorphisms on the boundary. In particular, in Section 7, we show that cyclic subgroups of hierarchically hyperbolic groups are undistorted. Section 8 is a brief technical discussion of essential HHSs and actions, supporting Section 9, in which we prove the coarse rank-rigidity theorem and some of its consequences. In Section 10, we consider CAT(0) cube complexes with HHS structures coming from [5], relating the HHS boundary to the *simplicial boundary* from Hagen [36].

Acknowledgments Hagen is grateful to Dan Guralnik and Alessandra Iozzi for a discussion clarifying the situation now described in Remark 10.8 and to Dave Futer and Brian Rushton for a discussion about some alternative topologies on the boundary of a cube complex related to the one discussed in Section 10. Durham is grateful to Henry Wilton and Cambridge University for their hospitality during a visit in the fall of 2015. We are grateful to Jason Behrstock for several helpful discussions and for assistance with the HHS infographic mentioned above. We are grateful to Bob Daverman for a useful MathOverflow comment on the shrinkable decompositions of manifolds, which has implications for the structure of the HHS boundary of Teichmüller space. We thank Robert Tang for a useful suggestion related to Section 5. We also thank the organizers of the *Young Geometric Group Theory IV* meeting, where initial discussions about this project took place. We are also very grateful to Sarah Mousley, Jacob Russell, and the anonymous referee for numerous helpful comments and corrections. Durham was supported by the National Science Foundation under grant number NSF 1045119. Hagen was supported by NSF grant number 1045119 and by the EPSRC grant of Henry Wilton. Sisto was supported by the Swiss National Science Foundation project 144373.

1 Background

1.1 Hierarchically hyperbolic spaces

We begin by recalling the definition of a hierarchically hyperbolic space, introduced in [5] and axiomatized in a more efficient fashion in [6], as follows. We begin by defining a hierarchically hyperbolic space. We will work in the context of a *quasigeodesic space*, $\mathcal{X}$, ie a metric space where any two points can be connected by a uniform-quality quasigeodesic.

Definition 1.1 (hierarchically hyperbolic space) The q–quasigeodesic space $(\mathcal{X}, \mathsf{d}_{\mathcal{X}})$ is a *hierarchically hyperbolic space* if there exists $\delta \geq 0$, an index set $\mathfrak{S}$, whose elements we call *domains*, and a set $\{\mathcal{C}W : W \in \mathfrak{S}\}$ of δ–hyperbolic spaces $(\mathcal{C}U, \mathsf{d}_U)$, such that the following conditions are satisfied:

(1) **Projections** There is a set $\{\pi_W\colon \mathcal{X} \to 2^{\mathcal{C}W} \mid W \in \mathfrak{S}\}$ of *projections* sending points in $\mathcal{X}$ to sets of diameter bounded by some $\xi \geq 0$ in the various $\mathcal{C}W \in \mathfrak{S}$. Moreover, there exists K such that each π_W is (K, K)–coarsely Lipschitz.

(2) **Nesting** $\mathfrak{S}$ is equipped with a partial order $\sqsubseteq$, and either $\mathfrak{S} = \varnothing$ or $\mathfrak{S}$ contains a unique $\sqsubseteq$–maximal element; when $V \sqsubseteq W$, we say V is *nested* in W. We require that $W \sqsubseteq W$ for all $W \in \mathfrak{S}$. For each $W \in \mathfrak{S}$, we denote by $\mathfrak{S}_W$ the set of $V \in \mathfrak{S}$ such that $V \sqsubseteq W$. Moreover, for all $V, W \in \mathfrak{S}$ with $V \sqsubsetneq W$ there is a specified subset $\rho^V_W \subset \mathcal{C}W$ with $\mathrm{diam}_{\mathcal{C}W}(\rho^V_W) \le \xi$. There is also a *projection* $\rho^W_V \colon \mathcal{C}W \to 2^{\mathcal{C}V}$. (The notation is justified by viewing ρ^V_W as a coarsely constant map $\mathcal{C}V \to 2^{\mathcal{C}W}$.)

(3) **Orthogonality** $\mathfrak{S}$ has a symmetric and antireflexive relation called *orthogonality*: we write $V \perp W$ when V and W are orthogonal. Also, whenever $V \sqsubseteq W$ and $W \perp U$, we require that $V \perp U$. We require that for each $T \in \mathfrak{S}$ and each $U \in \mathfrak{S}_T$ for which $\{V \in \mathfrak{S}_T : V \perp U\} \neq \varnothing$, there exists $W \in \mathfrak{S}_T - \{T\}$ such that whenever $V \perp U$ and $V \sqsubseteq T$, we have $V \sqsubseteq W$. Finally, if $V \perp W$, then V and W are not $\sqsubseteq$–comparable.

(4) **Transversality and consistency** If $V, W \in \mathfrak{S}$ are not orthogonal and neither is nested in the other, then we say V and W are *transverse*, denoted by $V \pitchfork W$. There exists $\kappa_0 \ge 0$ such that if $V \pitchfork W$, then there are sets $\rho^V_W \subseteq \mathcal{C}W$ and $\rho^W_V \subseteq \mathcal{C}V$ each of diameter at most ξ and satisfying:

$$\min\{\mathsf{d}_W(\pi_W(x), \rho^V_W), \mathsf{d}_V(\pi_V(x), \rho^W_V)\} \le \kappa_0$$

for all $x \in \mathcal{X}$.

For $V, W \in \mathfrak{S}$ satisfying $V \sqsubseteq W$ and for all $x \in \mathcal{X}$,

$$\min\{\mathsf{d}_W(\pi_W(x), \rho^V_W),\ \mathrm{diam}_{\mathcal{C}V}\bigl(\pi_V(x) \cup \rho^W_V(\pi_W(x))\bigr)\} \le \kappa_0.$$

The preceding two inequalities are the *consistency inequalities* for points in $\mathcal{X}$. Finally, if $U \sqsubseteq V$, then $\mathsf{d}_W(\rho^U_W, \rho^V_W) \le \kappa_0$ whenever $W \in \mathfrak{S}$ satisfies $W \not\sqsubseteq U$ and either $V \sqsubsetneq W$ or $V \pitchfork W$.

(5) **Finite complexity** There exists $n \ge 0$, the *complexity* of $\mathcal{X}$ (with respect to $\mathfrak{S}$), such that any set of pairwise-$\sqsubseteq$–comparable elements has cardinality at most n.

(6) **Large links** There exist $\lambda \ge 1$ and $E \ge \max\{\xi, \kappa_0\}$ such that the following holds: Let $W \in \mathfrak{S}$ and let $x, x' \in \mathcal{X}$. Let $N = \lambda \mathsf{d}_W(\pi_W(x), \pi_W(x')) + \lambda$. Then there exists $\{T_i\}_{i=1,\dots,\lfloor N \rfloor} \subseteq \mathfrak{S}_W - \{W\}$ such that for all $T \in \mathfrak{S}_W - \{W\}$, either $T \in \mathfrak{S}_{T_i}$ for some i or $\mathsf{d}_T(\pi_T(x), \pi_T(x')) < E$. Also, $\mathsf{d}_W(\pi_W(x), \rho^{T_i}_W) \le N$ for each i.

(7) **Bounded geodesic image** For all $W \in \mathfrak{S}$, all $V \in \mathfrak{S}_W - \{W\}$, and all geodesics γ of $\mathcal{C}W$, either $\mathrm{diam}_{\mathcal{C}V}(\rho^W_V(\gamma)) \le E$ or $\gamma \cap \mathcal{N}_E(\rho^V_W) \neq \varnothing$.

(8) **Partial realization** There exists a constant α with the following property: Let $\{V_j\}$ be a family of pairwise-orthogonal elements of $\mathfrak{S}$ and let $p_j \in \pi_{V_j}(\mathcal{X}) \subseteq \mathcal{C}V_j$. Then there exists $x \in \mathcal{X}$ such that

- $\mathsf{d}_{V_j}(x, p_j) \le \alpha$ for all j,

- for each j and each $V \in \mathfrak{S}$ with $V_j \sqsubseteq V$, we have $\mathsf{d}_V(x, \rho_V^{V_j}) \leq \alpha$, and
- if $W \pitchfork V_j$ for some j, then $\mathsf{d}_W(x, \rho_W^{V_j}) \leq \alpha$.

(9) **Uniqueness** For each $\kappa \geq 0$, there exists $\theta_u = \theta_u(\kappa)$ such that if $x, y \in \mathcal{X}$ and $\mathsf{d}(x, y) \geq \theta_u$, then there exists $V \in \mathfrak{S}$ such that $\mathsf{d}_V(x, y) \geq \kappa$.

We often refer to $\mathfrak{S}$, together with the nesting and orthogonality relations, the projections, and the hierarchy paths, as a *hierarchically hyperbolic structure* for the space $\mathcal{X}$.

Notation 1.2 Given $U \in \mathfrak{S}$, we often suppress the projection map π_U when writing distances in $\mathcal{C}U$: given $x, y \in \mathcal{X}$ and $p \in \mathcal{C}U$ we write $\mathsf{d}_U(x, y)$ for $\mathsf{d}_U(\pi_U(x), \pi_U(y))$ and $\mathsf{d}_U(x, p)$ for $\mathsf{d}_U(\pi_U(x), p)$. To measure distance between a pair of sets, we take the infimal distance between the two sets. Given $A \subset \mathcal{X}$ and $U \in \mathfrak{S}$ we let $\pi_U(A)$ denote $\bigcup_{a \in A} \pi_U(a)$.

Remark 1.3 (summary of constants) Each hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ is associated with a collection of constants often, as above, denoted by δ, ξ, n, κ_0, E, θ_u and K, where

(1) $\mathcal{C}U$ is δ–hyperbolic for each $U \in \mathfrak{S}$,

(2) each π_U has image of diameter at most ξ and is (K, K)–coarsely Lipschitz, and each ρ_V^U has (image of) diameter at most ξ,

(3) for each $x \in \mathcal{X}$, the tuple $(\pi_U(x))_{U \in \mathfrak{S}}$ is κ_0–consistent,

(4) E is the constant from the bounded geodesic image axiom.

Whenever working in a fixed hierarchically hyperbolic space, we use the above notation freely. We can, and shall, assume that $E \geq q$, $E \geq \delta$, $E \geq \xi$, $E \geq \kappa_0$, $E \geq K$ and $E \geq \alpha$.

Lemma 1.4 ("finite dimension") *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space of complexity n and let $U_1, \ldots, U_k \in \mathfrak{S}$ be pairwise orthogonal. Then $k \leq n$.*

Proof Definition 1.1(3) provides $W_1 \in \mathfrak{S}$, not $\sqsubseteq$–maximal, such that $U_2, \ldots, U_k \sqsubseteq W_1$. Using Definition 1.1 inductively yields a sequence $W_{k-1} \subsetneq W_{k-2} \subsetneq \cdots \subsetneq W_1 \sqsubseteq S$, with S $\sqsubseteq$–maximal, such that $U_{i-1}, \ldots, U_k \sqsubseteq W_i$ for $1 \leq i \leq k-1$. Hence $k \leq n$ by Definition 1.1(5). □

The next lemma is a simple consequence of the axioms and also appears in [7]:

Lemma 1.5 *Let $U, V, W \in \mathfrak{S}$ satisfy $U \perp V$, and $U, V \not\perp W$, and $W \not\sqsubseteq U, V$. Then $\mathsf{d}_W(\rho_W^U, \rho_W^V) \leq 2E$.*

Proof Our assumptions imply that $U \sqsubsetneq W$ or $U \pitchfork W$, and the same is true for V. Applying partial realization yields a point $x \in \mathcal{X}$ such that $\mathsf{d}_T(x, \rho_T^U)$, $\mathsf{d}_T(x, \rho_T^V) \leq E$ whenever $T \not\sqsubseteq U, V$ and $T \not\perp U, V$. The claim follows from the triangle inequality. □

Definition 1.6 For $D \geq 1$, a path γ in $\mathcal{X}$ is a D*–hierarchy path* if

(1) γ is a (D, D)–quasigeodesic,

(2) $\pi_W \circ \gamma$ is an unparametrized (D, D)–quasigeodesic for each $W \in \mathfrak{S}$.

An unbounded hierarchy path $[0, \infty) \to \mathcal{X}$ is a *hierarchy ray*.

The following theorems are proved in [6]:

Theorem 1.7 (realization theorem) *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic. Then for each κ there exist θ_e and θ_u such that the following holds. Let $\vec{\mathrm{b}} \in \prod_{W \in \mathfrak{S}} 2^{\mathcal{C}W}$ have each coordinate correspond to a subset of $\mathcal{C}W$ of diameter at most κ; for each W, let b_W denote the $\mathcal{C}W$–coordinate of $\vec{\mathrm{b}}$. Suppose that whenever $V \pitchfork W$ we have*

$$\min\{\mathsf{d}_W(b_W, \rho_W^V), \mathsf{d}_V(b_V, \rho_V^W)\} \leq \kappa$$

and whenever $V \sqsubseteq W$ we have

$$\min\{\mathsf{d}_W(b_W, \rho_W^V), \operatorname{diam}_{\mathcal{C}V}(b_V \cup \rho_V^W(b_W))\} \leq \kappa.$$

Then the set of all $x \in \mathcal{X}$ such that $\mathsf{d}_W(b_W, \pi_W(x)) \leq \theta_e$ for all $\mathcal{C}W \in \mathfrak{S}$ is nonempty and has diameter at most θ_u.

Theorem 1.8 (existence of hierarchy paths) *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic. Then there exists D_0 such that any $x, y \in \mathcal{X}$ are joined by a D_0–hierarchy path.*

Theorem 1.9 (distance formula) *Let $(X, \mathfrak{S})$ be hierarchically hyperbolic. Then there exists $s_0 \geq \xi$ such that for all $s \geq s_0$ there exist constants K and C such that, for all $x, y \in \mathcal{X}$,*

$$\mathsf{d}_{\mathcal{X}}(x, y) \asymp_{(K,C)} \sum_{W \in \mathfrak{S}} \{\{\mathsf{d}_W(\pi_W(x), \pi_W(y))\}\}_s.$$

The notation $\{\{A\}\}_B$ denotes the quantity which is A if $A \geq B$ and 0 otherwise.

1.2 Hieromorphisms, automorphisms and hierarchically hyperbolic groups

Morphisms in the category of hierarchically hyperbolic spaces were defined in [6], along with the related notion of a hierarchically hyperbolic group; we recall these definitions here.

Definition 1.10 (hieromorphism) Let $(\mathcal{X}, \mathfrak{S})$ and $(\mathcal{X}', \mathfrak{S}')$ be hierarchically hyperbolic structures on the spaces $\mathcal{X}$ and $\mathcal{X}'$, respectively. A *hieromorphism*

$$\big(f, \pi(f), \{\rho(f,U)\colon U \to \pi(f)(U) \mid U \in \mathfrak{S}\}\big)\colon (\mathcal{X}, \mathfrak{S}) \longrightarrow (\mathcal{X}', \mathfrak{S}')$$

consists of a map $f\colon \mathcal{X} \to \mathcal{X}'$, a map $\pi(f)\colon \mathfrak{S} \to \mathfrak{S}'$ preserving nesting, transversality and orthogonality, and a set $\{\rho(f,U)\colon U \to \pi(f)(U) \mid U \in \mathfrak{S}\}$ of quasiisometric embeddings with uniform constants such that the following two diagrams coarsely commute for all nonorthogonal $U, V \in \mathfrak{S}$:

$$\begin{tikzcd} \mathcal{X} \arrow[r, "f"] \arrow[d, "\pi_U"] & \mathcal{X}' \arrow[d, "\pi_{\pi(f)(U)}"] \\ \mathcal{C}U \arrow[r, "\rho(f,U)"] & \mathcal{C}\pi(f)(U) \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} \mathcal{C}U \arrow[r, "\rho(f,U)"] \arrow[d, "\rho^U_V"] & \mathcal{C}\pi(f)(U) \arrow[d, "\rho^{\pi(f)(U)}_{\pi(f)(V)}"] \\ \mathcal{C}V \arrow[r, "\rho(f,V)"] & \mathcal{C}\pi(f)(V) \end{tikzcd}$$

where $\rho^U_V\colon \mathcal{C}U \to \mathcal{C}V$ is the map from Definition 1.1.

Definition 1.11 (automorphism of an HHS, automorphism group) A hieromorphism $f\colon (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}, \mathfrak{S})$ is an *automorphism* if $\pi(f)\colon \mathfrak{S} \to \mathfrak{S}$ is a bijection and $\rho(f,U)\colon \mathcal{C}U \to \mathcal{C}\pi(f)(U)$ is an isometry for each $U \in \mathfrak{S}$. When the context is clear, we will continue to use f to denote f, $\pi(f)$ and $\rho(f,U)$.

Observe that if f and f' are automorphisms of $(\mathcal{X}, \mathfrak{S})$, then $f \circ f'\colon \mathcal{X} \to \mathcal{X}$ is also an automorphism: compose the maps $\mathfrak{S} \to \mathfrak{S}$, and compose isometries of the hyperbolic spaces in the obvious way. Declare automorphisms f and f' *equivalent* if $\pi(f) = \pi(f')$ and $\rho(f,U) = \rho(f',U)$ for all $U \in \mathfrak{S}$. Note that $f, f'\colon \mathcal{X} \to \mathcal{X}$ uniformly coarsely coincide in this case.

Denote by $\mathrm{Aut}(\mathfrak{S})$ the set of equivalence classes of automorphisms, so $\mathrm{Aut}(\mathfrak{S})$ is a group with the obvious multiplication. If $[f] \in \mathrm{Aut}(\mathfrak{S})$, then $[f]^{-1}$ is represented by the quasiinverse of f associated to $\pi(f)^{-1}$ and $\{\rho(f,U)^{-1} \mid U \in \mathfrak{S}\}$.

Observe that $\mathrm{Aut}(\mathfrak{S})$ quasiacts on $\mathcal{X}$ by uniform quasiisometries. We will sometimes abuse language and refer to individual automorphisms as elements of $\mathrm{Aut}(\mathfrak{S})$, and refer to the "action" of $\mathrm{Aut}(\mathfrak{S})$ on $\mathcal{X}$. By an *action* of a group G on $(\mathcal{X}, \mathfrak{S})$, we mean a homomorphism $G \to \mathrm{Aut}(\mathfrak{S})$. "Coarse" properties of an action, like properness and coboundedness, make sense in this context.

Definition 1.12 (equivariant) Let $f\colon (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}', \mathfrak{S}')$ be a hieromorphism, $G \le \mathrm{Aut}(\mathfrak{S})$ and $G' \le \mathrm{Aut}(\mathfrak{S}')$, and $\phi\colon G \to G'$ a homomorphism. Then f is *ϕ–equivariant* if

$$\begin{array}{ccc} \mathfrak{S} & \xrightarrow{f} & \mathfrak{S}' \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle \phi(g)} \\ \mathfrak{S} & \xrightarrow{f} & \mathfrak{S}' \end{array} \quad \text{and} \quad \begin{array}{ccc} \mathcal{C}U & \xrightarrow{f} & \mathcal{C}f(U) \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle \phi(g)} \\ \mathcal{C}gU & \xrightarrow{f} & \mathcal{C}\phi(g)f(U) \end{array}$$

(coarsely) commute for all $g \in G$ and $U \in \mathfrak{S}$. This implies that $\phi(g)f(x) \asymp f(gx)$ for all $x \in \mathcal{X}$ and $g \in G$. If ϕ is an isomorphism and f is ϕ–equivariant, then f is *G–equivariant*.

Definition 1.13 (hierarchically hyperbolic group) A finitely generated group G is *hierarchically hyperbolic* if there exists a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$ such that $G \leq \mathrm{Aut}(\mathfrak{S})$, the action on $\mathcal{X}$ is proper and cobounded, and G acts on $\mathfrak{S}$ with finitely many orbits. In this case we can assume $\mathcal{X} = G$ (with any fixed word-metric) and that the action $G \to \mathrm{Aut}(\mathfrak{S})$ sends each $g \in G$ to an automorphism whose underlying map $G \to G$ is left multiplication by g. In this case, we say that $(G, \mathfrak{S})$ is *hierarchically hyperbolic*.

1.3 Standard product regions

The notion of a standard product region in a hierarchically hyperbolic space, introduced in [6], plays an important role in several places, so we recall the definition here. Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space and let $U \in \mathfrak{S}$. Let $\mathfrak{S}_U$ be the set of $V \in \mathfrak{S}$ with $V \sqsubseteq U$ (in particular, $U \in \mathfrak{S}_U$ is the unique $\sqsubseteq$–maximal element). Let $\mathfrak{S}_U^\perp$ be the set of $V \in \mathfrak{S}$ such that $V \perp U$, together with some $\sqsubseteq$–minimal $A \in \mathfrak{S}$ such that $V \sqsubseteq A$ for all such V.

Fix $\kappa \geq \kappa_0$ and let F_U be the space of κ–consistent tuples in $\prod_{V \in \mathfrak{S}_U} 2^{\mathcal{C}V}$ whose coordinates are sets of diameter $\leq \xi$. Similarly, let E_U be the set of κ–consistent tuples in $\prod_{V \in \mathfrak{S}_U^\perp - \{A\}} 2^{\mathcal{C}V}$ whose coordinates are sets of diameter $\leq \xi$. In fact, $(F_U, \mathfrak{S}_U)$ and $(E_U, \mathfrak{S}_U^\perp)$ are HHSs (the hyperbolic space associated to A is $\mathrm{im}_A(E_U)$), and there are hieromorphisms (see [6] or Definition 1.10), inducing quasiisometric embeddings, $F_U \to \mathcal{X}$ and $E_U \to \mathcal{X}$, extending to a coarsely defined map $F_U \times E_U \to \mathcal{X}$ whose image is hierarchically quasiconvex in the sense of [6] (or see below). Specifically, each tuple $\vec{b} \in F_U$ is sent to the tuple that coincides with $\vec{b}$ on $\mathfrak{S}_U$ and has coordinate ρ_V^U for all $V \in \mathfrak{S} - \{U\}$ such that $V \pitchfork U$ or $U \sqsubseteq V$, and is fixed at some base element of E_U on $\mathfrak{S}_U^\perp - \{A\}$. The map $E_U \to \mathcal{X}$ is defined analogously. The spaces F_U and E_U are the *standard nesting factor* and the *standard orthogonality factor*, respectively, associated to U. The maps are the *standard hieromorphisms* associated to U, and the image P_U of $F_U \times E_U$ is a *standard product region*. Where it will not cause confusion, we sometimes denote by E_U and F_U the images of the corresponding standard hieromorphisms.

Remark 1.14 (automorphisms of product regions) Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space and let $U \in \mathfrak{S}$. Recall that $(F_U, \mathfrak{S}_U)$ is a hierarchically hyperbolic space, where the hyperbolic spaces and projections implicit in the hierarchically hyperbolic structure are exactly those inherited from $\mathfrak{S}$. Recall that $(E_U, \mathfrak{S}_U^{\perp})$ is a hierarchically hyperbolic space, where $\mathcal{C}V$ is as in $(\mathcal{X}, \mathfrak{S})$ except when $V = A$ is the $\sqsubseteq$–maximal element. The hieromorphism $(E_U, \mathfrak{S}_U^{\perp}) \to (\mathcal{X}, \mathfrak{S})$ is determined by the choice of $A \in \mathfrak{S}$ that is $\sqsubseteq$–minimal among all those containing each V with $V \perp U$, which we take as the $\sqsubseteq$–maximal element of $\mathfrak{S}_U^{\perp}$.

Let $\mathcal{A}_U$ be the group of automorphisms g of $\mathfrak{S}$ such that $g \cdot U = U$. Then there are *restriction homomorphisms* $\theta_U, \theta_U^{\perp}\colon \mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U), \mathrm{Aut}(\mathfrak{S}_U^{\perp})$, defined as follows. Given $g \in \mathcal{A}_U$, let $\theta_U(g)$ act like g on $\mathfrak{S}_U$ and like g on each $\mathcal{C}V$ with $V \sqsubseteq U$.

Define $\theta^{\perp}$ analogously, to give an automorphism of $\mathfrak{S}_U^{\perp} - \{A\}$ restricting the action of g on $\mathfrak{S}$, and fixing A. When defining $g\colon \mathrm{im}_A(E_U) \to \mathrm{im}_A(E_U)$, we draw attention to two cases, which it will be important to distinguish in Section 9:

- There exist infinitely many $A_i \in \mathfrak{S}$ that are $\sqsubseteq$–minimal with the property that $V \sqsubseteq A_i$ whenever $V \perp U$. The minimality assumption implies that these A_i are pairwise nonnested, so, using Lemma 1.4 and the consistency axiom, we see that $\pi_{A_i}(E_U)$ has diameter bounded independently of A_i (in fact, just in terms of E); thus, when building the HHS $(E_U, \mathfrak{S}_U^{\perp})$, we can take the hyperbolic space $\mathrm{im}_A(E_U)$ associated to the maximal element A to be a single point, and define $g\colon \mathrm{im}_A(E_U) \to \mathrm{im}_A(E_U)$ in the obvious way. This conclusion holds, more generally, if there are two transverse $\sqsubseteq$–minimal "containers" A_i and A_j for the domains orthogonal to U.

- The set $\{A_i\}$ of domains that are $\sqsubseteq$–minimal with the property that $V \sqsubseteq A_i$ whenever $V \perp U$ is a pairwise-orthogonal set. In this case, there are at most n such A_i, where n is the complexity, by Lemma 1.4. Again, we choose $A \in \{A_i\}$ arbitrarily and define the HHS structure on $(E_U, \mathfrak{S}_U^{\perp})$ using A as the $\sqsubseteq$–maximal element, with associated hyperbolic space $\mathrm{im}_A(E_U)$. Now, if there exists $h \in \mathrm{Aut}(\mathfrak{S})$ such that $hA = A_i$ for some i, then $\mathrm{im}_{A_i}(E_U)$ is uniformly quasiisometric to $\mathrm{im}_A(E_U)$. In particular, $g\colon \mathrm{im}_A(E_U) \to \mathrm{im}_A(E_U)$ can be defined so that the restriction homomorphism $\theta_U^{\perp}$ makes sense.

Note that, if $f \in \mathcal{A}_U$ and $x \in P_U \subset \mathcal{X}$, then $\mathsf{d}_{F_U \times E_U}\big(\theta_U(f)(r_U(x)), r_U(f(x))\big)$ is uniformly bounded, where $r_U\colon P_U \cong_{\mathrm{qi}} F_U \times E_U \to F_U$ is coarse projection to the first factor, and a similar statement holds for $\theta_U^{\perp}$ and projection to E_U.

Finally, recall that the standard product region P_U is defined to be the image of $F_U \times E_U$ under the product of the hieromorphisms $(F_U, \mathfrak{S}_U), (E_U, \mathfrak{S}_U^{\perp}) \to (\mathcal{X}, \mathfrak{S})$. This map is coarsely defined, but it is convenient to fix maps $F_U \times E_U \to \mathcal{X}$ (realizing those hieromorphisms) such that $P_{gU} = gP_U$ for all $U \in \mathfrak{S}$ and $g \in \mathrm{Aut}(\mathfrak{S})$. Similarly, the image of F_{gU} coincides with gF_U, etc. The set $\{P_U : U \in \mathfrak{S}\}$ is $\mathrm{Aut}(\mathfrak{S})$–invariant.

1.4 Normalized hierarchically hyperbolic spaces and hierarchical quasiconvexity

Hierarchically hyperbolic spaces, in the sense of Definition 1.1, need not coarsely surject to the associated hyperbolic spaces, but in almost all cases of interest, they do. Accordingly:

Definition 1.15 (normalized HHS) The HHS $(\mathcal{X}, \mathfrak{S})$ is *normalized* if there exists C such that for all $U \in \mathfrak{S}$, we have $\mathcal{C}U = \mathcal{N}_{\mathcal{C}U}(\pi_U(\mathcal{X}))$.

Proposition 1.16 *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space. Then $\mathcal{X}$ admits a normalized hierarchically hyperbolic structure $(\mathcal{X}, \mathfrak{S}')$ with a hieromorphism $f\colon (\mathcal{X}, \mathfrak{S}') \to (\mathcal{X}, \mathfrak{S})$, where $f\colon \mathcal{X} \to \mathcal{X}$ is the identity and $f\colon \mathfrak{S}' \to \mathfrak{S}$ is a bijection. Moreover, if $G \leq \mathrm{Aut}(\mathfrak{S})$, then there is a monomorphism $G \to \mathrm{Aut}(\mathfrak{S}')$ making f equivariant.*

Proof Let $\mathfrak{S}' = \mathfrak{S}$, and retain the same nesting, orthogonality, and transversality relations. For each $U \in \mathfrak{S}'$, the associated hyperbolic space $\mathcal{C}_{\mathrm{norm}}U$ is chosen to be uniformly quasiisometric to the uniformly quasiconvex subset $\pi_U(\mathcal{X})$ of $\mathcal{C}U$. The projection $\pi_U\colon \mathcal{X} \to \mathcal{C}_{\mathrm{norm}}U$ is, up to composition with a uniform quasiisometry, unchanged (and therefore continues to be coarsely Lipschitz). Let $p_U\colon \mathcal{C}U \to \mathcal{C}_{\mathrm{norm}}U$ be the composition of the coarse closest-point projection $\mathcal{C}U \to \pi_U(\mathcal{X})$, composed with the uniform quasiisometry $\pi_U(\mathcal{X}) \to \mathcal{C}_{\mathrm{norm}}U$. Then, for all U and V with $U \pitchfork V$ or $U \sqsubseteq V$, define the relative projection $\mathcal{C}_{\mathrm{norm}}U \to \mathcal{C}_{\mathrm{norm}}V$ to be the composition of $p_U \circ \rho_V^U\colon \pi_U(\mathcal{X}) \to \mathcal{C}_{\mathrm{norm}}V$ with the quasiisometry $\mathcal{C}_{\mathrm{norm}}U \to \pi_U(\mathcal{X})$. The remaining assertions are a matter of checking definitions. □

Recall from [6] that the subspace $\mathcal{Y}$ of $(\mathcal{X}, \mathfrak{S})$ is *hierarchically quasiconvex* if there exists $k_0 \geq 0$ such that $\pi_U(\mathcal{Y})$ is k_0–quasiconvex in $\mathcal{C}U$ for all $U \in \mathfrak{S}$ and if, for all $\kappa \geq \kappa_0$, each κ–consistent tuple $\vec{b} \in \prod_{U \in \mathfrak{S}} \mathcal{C}U$ with U–coordinate in $\pi_U(\mathcal{Y})$ for all U has the property that any associated realization point $x \in \mathcal{X}$ lies at distance from $\mathcal{Y}$, depending only on κ.

In the interest of staying in the class of normalized hierarchically hyperbolic spaces, we will always work with a normalized hierarchically hyperbolic structure on $\mathcal{Y}$, namely the one provided by Proposition 1.16. Moreover, we will (abusively) eschew the notation $\mathcal{C}_{\mathrm{norm}}U$ and use the same notation for $\pi_U(\mathcal{Y})$ and its thickening; in other words, we will regard $\pi_U(\mathcal{Y})$ as a genuine (uniformly) hyperbolic geodesic space.

Finally, we recall the following notion from [6, Definition 5.3, Lemma 5.4]. Let $\mathcal{Y} \subset \mathcal{X}$ be a hierarchically quasiconvex subspace. Then there is a coarsely Lipschitz map $\mathfrak{g}_{\mathcal{Y}}\colon \mathcal{X} \to \mathcal{Y}$ (the coarse Lipschitz constants depend only on the constants from Definition 1.1 and the constants implicit in the definition of hierarchical quasiconvexity) with the following property: for each $U \in \mathfrak{S}$ and $x \in \mathcal{X}$, the projection $\pi_U(\mathfrak{g}_{\mathcal{Y}}(x))$ uniformly coarsely coincides with the coarse closest-point projection of $\pi_U(x)$ to the quasiconvex subspace $\pi_U(\mathcal{Y})$. The map $\mathfrak{g}_{\mathcal{Y}}$ is the *gate map* associated to $\mathcal{Y}$.

2 Definition of the boundary

Fix a hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$. For each $S \in \mathfrak{S}$, denote by $\partial \mathcal{C}S$ the Gromov boundary, ie the space of equivalence classes of sequences $(x_n \in \mathcal{C}S)$, where (x_n) and (y_n) are equivalent if for some (hence any) fixed basepoint $x \in \mathcal{C}S$, we have $(x_n, y_n)_x \to \infty$. In particular, $\partial \mathcal{C}S$ need not be compact if $\mathcal{C}S$ is not proper. The topology is as usual.

Remark 2.1 (extending the Gromov product) For $U \in \mathfrak{S}$, any $p, q \in \mathcal{C}U \cup \partial \mathcal{C}U$ are joined to $u \in \mathcal{C}U$ by $(1, 20\delta)$–quasigeodesics, enabling extension of the Gromov product to $\partial \mathcal{C}U$.

2.1 Supports and boundary points

We first define $\partial \mathcal{X} = \partial(\mathcal{X}, \mathfrak{S})$ as a set.

Definition 2.2 (support set, boundary point) A *support set* $\overline{S} \subset \mathfrak{S}$ is a set with $S_i \perp S_j$ for all $S_i, S_j \in \overline{S}$. Given a support set $\overline{S}$, a *boundary point* with *support* $\overline{S}$ is a formal sum $p = \sum_{S \in \overline{S}} a_S^p p_S$, where each $p_S \in \partial \mathcal{C}S$, and $a_S^p > 0$, and $\sum_{S \in \overline{S}} a_S^p = 1$. Such sums are necessarily finite, by Lemma 1.4. We denote the support $\overline{S}$ of p by $\mathrm{Supp}(p)$.

Definition 2.3 (boundary) The *boundary* $\partial(\mathcal{X}, \mathfrak{S})$ of $(\mathcal{X}, \mathfrak{S})$ is the set of boundary points.

Notation 2.4 When the specific HHS structure is clear, we write $\partial \mathcal{X}$ to mean $\partial(\mathcal{X}, \mathfrak{S})$.

2.2 Topologizing $\partial \mathcal{X}$

We topologize $\partial \mathcal{X}$ using the visual topologies on the Gromov boundaries of elements of $\{\mathcal{C}S : S \in \mathfrak{S}\}$. The main challenge is to incorporate these topologies into a coherent topology on the whole boundary, allowing boundary points supported on nonorthogonal domains to interact. This requires some preliminary definitions.

Definition 2.5 (remote point) Let $\overline{S} \subset \mathfrak{S}$ be a support set. A point $p \in \partial\mathcal{X}$ is *remote (with respect to $\overline{S}$, or with respect to some $q \in \partial\mathcal{X}$ with support $\overline{S}$)* if

(1) $\mathrm{Supp}(p) \cap \overline{S} = \varnothing$, and

(2) for all $S \in \overline{S}$, there exists $T \in \mathrm{Supp}(p)$ such that S and T are *not* orthogonal.

Denote by $\partial^{\mathrm{rem}}_{\overline{S}}\mathcal{X}$ the set of all remote points with respect to $\overline{S}$.

For each $S \in \mathfrak{S}$, let $\mathcal{B}(\mathcal{C}S)$ be the set of all bounded sets in $\mathcal{C}S$. If $\overline{S} \subset \mathfrak{S}$ is a support set, we denote by $\overline{S}^{\perp}$ the set of all $U \in \mathfrak{S}$ such that $U \perp S$ for all $S \in \overline{S}$.

Definition 2.6 (boundary projection) Let $\overline{S} \subset \mathfrak{S}$ be a support set. For each $q \in \partial^{\mathrm{rem}}_{\overline{S}}\mathcal{X}$, let $\overline{S}_q$ be the union of $\overline{S}$ and the set of domains $T \in \overline{S}^{\perp}$ such that T is not orthogonal to W_T for some $W_T \in \mathrm{Supp}(q)$. Define a *boundary projection* $\partial\pi_{\overline{S}}(q) \in \prod_{S \in \overline{S}_q} \mathcal{C}S$ as follows. Let $q = \sum_{T \in \overline{T}} a^p_T q_T$ be a remote point with respect to $\overline{S}$. For each $S \in \overline{S}_q$, let $T_S \in \mathrm{Supp}(q)$ be chosen so that S and T_S are not orthogonal. Define the S–coordinate $(\partial\pi_{\overline{S}}(q))_S$ of $\partial\pi_{\overline{S}}(q)$ as follows:

(1) If $T_S \sqsubseteq S$ or $T_S \pitchfork S$, then $(\partial\pi_{\overline{S}}(q))_S = \rho^{T_S}_S$.

(2) Otherwise, $S \sqsubseteq T_S$. Choose a $(1, 20\delta)$–quasigeodesic ray γ in $\mathcal{C}T_S$ joining $\rho^S_{T_S}$ to q_{T_S}. By the bounded geodesic image axiom, there exists $x \in \gamma$ such that $\rho^{T_S}_S$ is coarsely constant on the subray of γ beginning at x. Let $(\partial\pi_{\overline{S}}(q))_S = \rho^{T_S}_S(x)$.

Lemma 2.7 *The map $\partial\pi_{\overline{S}}$ is coarsely independent of the choice of $\{T_S\}_{S \in \overline{S}}$.*

Proof Suppose that $T_S, T'_S \in \overline{T}$ are chosen so that T_S and T'_S are not orthogonal to S and suppose that $S \not\sqsubseteq T_S, T'_S$. In other words, either $T_S \sqsubseteq S$ or $T_S \pitchfork S$ and the same is true for T'_S. By partial realization (Definition 1.1(8)), there therefore exists $y \in \mathcal{X}$ such that $\mathrm{d}_S(\rho^{T_S}_S, y), \mathrm{d}_S(\rho^{T'_S}_S, y) \leq E$, whence $\rho^{T_S}_S$ and $\rho^{T'_S}_S$ coarsely coincide. If $S \sqsubseteq T_S$, then $S \perp T'_S$ since $T_S \perp T'_S$; this contradicts the defining property of T'_S. Hence, in all allowable situations, $\rho^{T_S}_S$ coarsely coincides with $\rho^{T'_S}_S$; the claim follows. □

Fix a basepoint $x_0 \in \mathcal{X}$. We are now ready to define a neighborhood basis for each $p = \sum_{S \in \overline{S}} a^p_S p_S$, where $p_S \in \mathcal{C}S$ for all $S \in \mathrm{Supp}(p) = \overline{S}$. For each $S \in \mathfrak{S}$, choose a cone-topology neighborhood U_S of p_S in $\mathcal{C}S \cup \partial\mathcal{C}S$, and choose $\epsilon > 0$. For convenience, given $q \in \partial\mathcal{X}$, we let $a^q_T = 0$ when $T \in \mathfrak{S} - \mathrm{Supp}(q)$.

We define the basic set $\mathcal{N}_{\{U_S\},\epsilon}(p)$ as the union of a *remote part*, a *nonremote part*, and an *interior part*, as follows:

Definition 2.8 (remote part) The *remote part* is

$$\mathcal{N}^{\mathrm{rem}}_{\{U_S\},\epsilon}(p)=\left\{q\in\partial^{\mathrm{rem}}_{\overline{S}}\mathcal{X}\;\middle|\;\forall S\in\overline{S},\ (\partial\pi_{\overline{S}}(q))_S\in U_S \text{ and } \sum_{T\in\overline{S}^{\perp}} a^q_T<\epsilon \right.$$
$$\left.\text{and } \forall S\in\overline{S}_q,\ S'\in\overline{S},\ \left|\frac{\mathsf{d}_S(x_0,(\partial\pi_{\overline{S}}(q))_S)}{\mathsf{d}_{S'}(x_0,(\partial\pi_{\overline{S}}(q))_{S'})}-\frac{a^p_S}{a^p_{S'}}\right|<\epsilon\right\}.$$

Definition 2.9 (nonremote part) Given $p,q\in\partial\mathcal{X}$, let $A=\operatorname{Supp}(p)\cap\operatorname{Supp}(q)$. The *nonremote part* is

$$\mathcal{N}^{\mathrm{non}}_{\{U_S\},\epsilon}(p)=\left\{q=\sum_T a^q_T q_T\in\partial\mathcal{X}-\partial^{\mathrm{rem}}_{\overline{S}}\mathcal{X}\;\middle|\;\forall T\in A,\ |a^q_T-a^p_T|<\epsilon \text{ and } q_T\in U_T,\right.$$
$$\left.\text{and } \sum_{V\in\operatorname{Supp}(q)-A} a^q_V<\epsilon\right\}.$$

Definition 2.10 (interior part) The *interior part* is

$$\mathcal{N}^{\mathrm{int}}_{\{U_S\},\epsilon}(p)=\left\{x\in\mathcal{X}\;\middle|\;\forall S,S'\in\overline{S},\ \forall T\in\overline{S}^{\perp},\ \pi_S(x)\in U_S \text{ and } \left|\frac{a_S}{a_{S'}}-\frac{\mathsf{d}_S(x_0,x)}{\mathsf{d}_{S'}(x_0,x)}\right|<\epsilon\right.$$
$$\left.\text{and } \frac{\mathsf{d}_T(x_0,x)}{\mathsf{d}_S(x_0,x)}<\epsilon\right\}.$$

Definition 2.11 (topology on $\mathcal{X}\cup\partial\mathcal{X}$) For each $p\in\partial\mathcal{X}$ with $\operatorname{Supp}(p)=\overline{S}$, and $\{U_S : S\in\overline{S}\}$ and $\epsilon>0$ as above, let

$$\mathcal{N}_{\{U_S\},\epsilon}(p)=\mathcal{N}^{\mathrm{rem}}_{\{U_S\},\epsilon}(p)\cup\mathcal{N}^{\mathrm{non}}_{\{U_S\},\epsilon}(p)\cup\mathcal{N}^{\mathrm{int}}_{\{U_S\},\epsilon}(p).$$

We declare the set of all such $\mathcal{N}_{\{U_S\},\epsilon}(p)$ to form a neighborhood basis at p. Also, we include in the topology on $\mathcal{X}\cup\partial\mathcal{X}$ the open sets in $\mathcal{X}$. This topology does not depend on x_0.

Remark 2.12 The $\mathcal{N}_{\{U_S\},\epsilon}(p)$ need not be open; a priori, they may have empty interior!

The following is an obvious consequence of the definitions:

Proposition 2.13 *For all $U\in\mathfrak{S}$, the inclusion $\partial\mathcal{C}U\hookrightarrow\partial\mathcal{X}$ is an embedding.*

Proposition 2.17 gives basic properties of $\partial\mathcal{X}$; first we need a definition and some lemmas.

Definition 2.14 (basically Hausdorff) Let $\mathcal{H}$ be a topological space and let $\mathcal{B}$ be a neighborhood basis. Then $(\mathcal{H},\mathcal{B})$ is *basically Hausdorff* if for all distinct $h,h'\in\mathcal{H}$, there exist disjoint $B,B'\in\mathcal{B}$ with $h\in B$ and $h'\in B'$.

Lemma 2.15 *Let $(\mathcal{X},\mathfrak{S})$ be hierarchically hyperbolic and let $\overline{\mathcal{X}}=\mathcal{X}\cup\partial(\mathcal{X},\mathfrak{S})$. Then, equipped with the neighborhood basis declared above, $\overline{\mathcal{X}}$ is basically Hausdorff.*

Proof Let $p,q\in\overline{\mathcal{X}}$ be distinct. The statement is obvious when p or q is in $\mathcal{X}$, so assume that $p,q\in\partial\mathcal{X}$. Fix a basepoint $x_0\in\mathcal{X}$.

For each $U\in\mathrm{Supp}(p)$, choose a neighborhood Y_U^p of p in $\mathcal{C}U\cup\partial\mathcal{C}U$ that does not contain $(\partial\pi_{\mathrm{Supp}(p)}(q))_U$, provided it is defined. For each $T\in\mathrm{Supp}(q)$, choose a neighborhood Y_T^q of q in $\mathcal{C}T\cup\partial\mathcal{C}T$ that does not intersect $\mathcal{N}_{1000E+\omega}(\{\pi_T(x_0)\})$ and, when it is defined, $\mathcal{N}_{1000E+\omega}((\partial\pi_{\mathrm{Supp}(q)}(p))_T)$, where $\omega\geq 0$ is to be determined; also choose Y_T^q so that $Y_T^p\cap Y_T^q=\varnothing$ when $T\in\mathrm{Supp}(p)\cap\mathrm{Supp}(q)$, unless $p_T=q_T$, in which case we choose $Y_T^p=Y_T^q$. Fix $\epsilon>0$, to be determined. Let $\mathcal{N}(p)=\mathcal{N}_{\{Y_U^p\},\epsilon}(p)$ and $\mathcal{N}(q)=\mathcal{N}_{\{Y_V^q\},\epsilon}(q)$.

Finally, for any $w,v\in\partial\mathcal{X}$, let $\mathrm{Supp}(w)_v=\mathrm{Supp}(w)\cup(\mathrm{Supp}(w)^\perp-\mathrm{Supp}(v)^\perp)$.

We need an auxiliary claim:

Claim 1 Let $x,p,q\in\partial\mathcal{X}$. Suppose there exist $W_p,W_q\in\mathrm{Supp}(x)$, and $U\in\mathrm{Supp}(p)_x$ and $V\in\mathrm{Supp}(q)_x$, such that $W_p\not\perp U$ and $W_p\neq U$, and $W_q\not\perp V$ and $W_q\neq V$. Then there exists $y\in P_{W_p}\cap P_{W_q}\subset\mathcal{X}$ such that $(\partial\pi_{\mathrm{Supp}(p)}(x))_U$ $100E$–coarsely coincides with $\pi_U(y)$, and $(\partial\pi_{\mathrm{Supp}(q)}(x))_V$ $100E$–coarsely coincides with $\pi_V(y)$.

(P_{W_p} is the standard product region associated to W_p, defined in Section 1.3.)

Proof of Claim 1 If $W_p\pitchfork U$ or $W_p\sqsubseteq U$, and $W_q\pitchfork V$ or $W_q\sqsubseteq V$, then any $y\in P_{W_p}\cap P_{W_q}$ suffices. If $U\sqsubsetneq W_p$, use partial realization to see that, given a $(1,20\delta)$–quasigeodesic ray γ in $\mathcal{C}W_p$ with endpoint x_{W_p}, we can choose a sequence (y_n) in $P_{W_p}\cap P_{W_q}$ projecting uniformly close to an unbounded sequence in γ. This provides the desired y. $\triangleleft$

Suppose that $x\in\mathcal{N}(p)\cap\mathcal{N}(q)$. We consider the following cases:

(1) ***x* ∈ ∂𝒳 is *p*–remote and *q*–remote** First of all, notice that by definition of remote, for any $U\in\mathrm{Supp}(p)$ there exists W_p as in Claim 1, and similarly for $V\in\mathrm{Supp}(q)$. We now consider the following subcases:

- (a) There exists $U\in\mathrm{Supp}(p)\cap\mathrm{Supp}(q)$ with $p_U\neq q_U$.
- (b) There exists $U\in\mathrm{Supp}(p)\cap\mathrm{Supp}(q)$ with $p_U=q_U$ but $a_U^p\neq a_U^q$.
- (c) Up to swapping p and q, there exists $V\in\mathrm{Supp}(q)-\mathrm{Supp}(p)$, and there exists $U\in\mathrm{Supp}(p)$ not orthogonal to V.
- (d) The previous case does not apply and, up to swapping p and q, there exists $V\in(\mathrm{Supp}(q)-\mathrm{Supp}(p))\cap\mathrm{Supp}(p)^\perp$.

(a) Then we would have that $(\partial\pi_{\mathrm{Supp}(p)}(x))_U$ is contained in both Y_U^p and Y_U^q, which are disjoint, a contradiction.

(b) Let $\mathcal{U} = \mathrm{Supp}(p) \cap \mathrm{Supp}(q)$. For each $V \in \mathcal{U}$ we have that the ratio

$$\frac{d_V\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(x))_V\big)}{d_U\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(x))_U\big)}$$

is ϵ–close to both a_V^p/a_U^p and a_V^q/a_U^q. Hence, if there exists $V \in \mathcal{U}$ such that $a_V^p/a_U^p \neq a_V^q/a_U^q$, we can choose ϵ small enough to give a contradiction. Otherwise, since the coefficients sum to 1, the supports of p and q do not coincide, and we deal with this in the next subcases.

(c) If $U \pitchfork V$, then by our choice of $\mathcal{N}(p)$ and $\mathcal{N}(q)$, we have $\mathsf{d}_U(y, \rho_U^V) > E$ and $\mathsf{d}_V(y, \rho_V^U) > E$ for y as in Claim 1, contradicting consistency. If $U \subsetneq V$ or $V \subsetneq U$, then we reach a similar contradiction of consistency.

(d) Suppose also that $\mathrm{Supp}(p) \subseteq \mathrm{Supp}(q) \cup \mathrm{Supp}(q)^\perp$ but $\mathrm{Supp}(p) \cap \mathrm{Supp}(q)^\perp \neq \varnothing$, since otherwise either (a) or (b) holds. Let $U \in \mathrm{Supp}(p) - \mathrm{Supp}(q)$. By remoteness of x, we have $U \in \mathrm{Supp}(q)^\perp - \mathrm{Supp}(x)^\perp$, so $U \in \mathrm{Supp}(q)_x$. Hence the definition of q–remoteness gives

$$\left|\frac{\mathsf{d}_U\big(x_0, (\partial\pi_{\mathrm{Supp}(q)}(x))_U\big)}{\mathsf{d}_V\big(x_0, (\partial\pi_{\mathrm{Supp}(q)}(x))_V\big)} - \frac{a_U^q}{a_V^q}\right| < \epsilon.$$

Similarly, we have $V \in \mathrm{Supp}(p)_x$, so the definition of p–remoteness gives

$$\left|\frac{d_V\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(x))_V\big)}{d_U\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(x))_U\big)} - \frac{a_V^p}{a_U^p}\right| < \epsilon.$$

Now, since $V \notin \mathrm{Supp}(p), U \notin \mathrm{Supp}(q)$, we have $a_V^p = a_U^q = 0$, so, we may take y to be the point in $\mathcal{X}$ provided by Claim 1, and hence we have $\mathsf{d}_V(y, x_0)/\mathsf{d}_U(y, x_0) < 2\epsilon$ and $\mathsf{d}_U(y, x_0)/\mathsf{d}_V(y, x_0) < 2\epsilon$ provided ω in Claim 1 was chosen sufficiently large in terms of ϵ and E. This is a contradiction.

(2) $\boldsymbol{x \in \mathcal{X}}$ In this case, x can play the role of y in the above arguments.

(3) **$\boldsymbol{x \in \partial\mathcal{X}}$ is $\boldsymbol{p}$–nonremote and $\boldsymbol{q}$–nonremote** In this case, first choose $\epsilon \in \big(0, \frac{1}{2}\big)$ smaller than $\frac{1}{10}|a_W^p - a_W^q|$ for each $W \in \mathrm{Supp}(p) \cap \mathrm{Supp}(q)$. The definition of the nonremote part now ensures that x cannot exist.

(4) **$\boldsymbol{x \in \partial\mathcal{X}}$ is $\boldsymbol{p}$–remote and $\boldsymbol{q}$–nonremote** In this case, there exists $U \in \mathrm{Supp}(p)$ and $V \in \mathrm{Supp}(q)$, and $W_p, W_q \in \mathrm{Supp}(x)$, such that W_p is distinct from and not orthogonal to U while $W_p = V$ or $W_p \perp V$. If for each such W_q we have $W_q \in \mathrm{Supp}(q)^\perp$,

then by choosing $\epsilon < 1$, we have that $\sum_{T \in \mathrm{Supp}(x)} a_T^x < 1$, a contradiction. Thus we may take $W_q = V \in \mathrm{Supp}(q)$.

Now, choose $y \in P_{W_p}$ so that $(\partial\pi_{\mathrm{Supp}(p)}(x))_U$ $100E$–coarsely coincides with $\pi_U(y)$. If $U = W_q$, then our choice of $\mathcal{N}(p)$ and $\mathcal{N}(q)$ ensures that x cannot lie in both. Suppose that $U \pitchfork W_q$. Then $\pi_U(y)$, $\rho_U^{W_q}$ and $\rho_U^{W_p}$ all $10E$–coarsely coincide and lie at distance $50E$ from the required neighborhood of p_U, so $x \notin \mathcal{N}(p)$. When $U \subsetneq W_q$ or $W_q \subsetneq U$, a similar argument shows that $x \notin \mathcal{N}(p) \cap \mathcal{N}(q)$.

Hence it remains to consider the case where $W_q \perp U$. By definition, $|a_{W_q}^x - a_{W_q}^q| < \epsilon$. On the other hand, we can assume $W_q \in \mathrm{Supp}(p)^\perp$, for otherwise we could rechoose U and W_q to be in one of the above cases. Thus, by definition, $a_{W_q}^x < \epsilon$. This yields a contradiction provided we choose, say, $\epsilon \in \bigcap_{T \in \mathrm{Supp}(q)} \big(0, \frac{1}{10} a_T^q\big)$.

Hence our choice of $\mathcal{N}(p)$ and $\mathcal{N}(q)$ ensures $\mathcal{N}(p) \cap \mathcal{N}(q) = \varnothing$, as required. □

Lemma 2.16 *$\overline{\mathcal{X}}$ is Hausdorff.*

Proof In light of Lemma 2.15, it suffices to show that for all $p \in \partial\mathcal{X}$, with $p = \sum_{T \in \mathrm{Supp}(p)} a_T p_T$, all $\epsilon > 0$, and all collections $\{U_T : T \in \mathrm{Supp}(p)\}$ with each U_T a neighborhood of p_T in $\mathcal{C}T \cup \partial\mathcal{C}T$, the corresponding basic set $\mathcal{N}_{\{U_T\},\epsilon}(p)$ has nonempty interior.

The topology of basic convergence Given a sequence $\{p_n\}$ with each $p_n \in \overline{\mathcal{X}}$, we say that p_n *basically converges* to $p \in \partial\mathcal{X}$ if for all $\epsilon > 0$ and all choices of $\{U_T\}$ as above, we have $p_n \in \mathcal{N}_{\{U_T\},\epsilon}(p)$ for all but finitely many $n \in \mathbb{N}$. Similarly, $\{p_n\}$ *basically converges* to $p \in \mathcal{X}$ if, for all $\epsilon > 0$, we have $p_n \in \mathcal{N}_\epsilon(p)$ for all sufficiently large n.

Define a topology on $\overline{\mathcal{X}}$ as follows: the set $A \subset \overline{\mathcal{X}}$ is declared to be closed if $a \in A$ whenever there is a sequence $\{a_n\}$ such that $a_n \in A$ for all n and a_n basically converges to a. Denote by $\mathfrak{M}$ the space $\overline{\mathcal{X}}$ endowed with this topology.

Nonempty interior of basic sets Let $\mathcal{N} = \mathcal{N}_{\{U_T\},\epsilon}(p)$ be a basic set as above. We claim that $p \in \mathrm{Int}(\mathcal{N})$. Otherwise, there exists a sequence $\{p_n\}$ in $\overline{\mathcal{X}} - \mathcal{N}$ that basically converges to p. This is a contradiction since basic convergence to p needs $\{p_n\}$ to enter $\mathcal{N}$.

Equivalence of the topologies To complete the proof that basic sets in $\overline{\mathcal{X}}$ have nonempty interior (with respect to the original topology), and thereby complete the proof of the lemma, it suffices to show that $\overline{\mathcal{X}}$ is homeomorphic to $\mathfrak{M}$.

Now, a set $A \subseteq \overline{\mathcal{X}}$ is closed in $\overline{\mathcal{X}}$ (ie has open complement) if and only if, for each $p \in \overline{\mathcal{X}} - A$, we can choose $\epsilon > 0$ and neighborhoods $\{U_T : T \in \mathrm{Supp}(p)\}$ so that

$\mathcal{N}_{\{U_T\},\epsilon}(p)$ is disjoint from A. But this is equivalent to the following: for all basically convergent $\{a_n\}$ with each $a_n \in A$, the (basic) limit a lies in A. This is in turn equivalent to the assertion that A is closed in $\mathfrak{M}$. □

Proposition 2.17 *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic, and let $\overline{\mathcal{X}} = \mathcal{X} \cup \partial(\mathcal{X}, \mathfrak{S})$.*

(1) *$\overline{\mathcal{X}}$ is Hausdorff and, if $\mathcal{X}$ is separable (eg if it is proper), then $\overline{\mathcal{X}}$ is separable.*
(2) *$\partial\mathcal{X}$ is closed in $\overline{\mathcal{X}}$.*
(3) *$\mathcal{X}$ is dense in $\overline{\mathcal{X}}$.*

Proof The "Hausdorff" part of assertion (1) follows from Lemma 2.16. Separability of $\overline{\mathcal{X}}$ follows from density of the metric space $\mathcal{X}$ in $\overline{\mathcal{X}}$, ie part (3). Assertion (2) is obvious: no bounded neighborhood of an interior point contains a boundary point, so no sequence of boundary points converges to an interior point.

It remains to prove (3). Pick a neighborhood $\mathcal{N}_{\{U_S\},\epsilon}(p)$ of $p = \sum_{S \in \mathrm{Supp}(p)} a_S^p p_S \in \partial\mathcal{X}$ with $p_S \in \partial\mathcal{C}S$ for $S \in \mathrm{Supp}(p)$. For each $S_i \in \mathrm{Supp}(p) = \{S_1, \dots, S_d\}$, fix a uniform quasigeodesic ray γ_i in $\mathcal{C}S$ from $\pi_S(x_0)$ to p_S.

First, suppose that $d = 1$. Then for each t, there exists x_1^t such that $\pi_{S_1}(x_1^t)$ coarsely coincides with $\gamma_1(a_{S_1}^p \cdot t)$ and, in view of the quasiisometric embedding $F_{S_1} \times E_{S_1} \to \mathcal{X}$ described in Section 1.3, the point x_1^t can be chosen so that $\pi_T(x_1^t)$ coarsely equals $\pi_T(x_0)$ for each $T \perp S_1$. (Here we have used that $(\mathcal{X}, \mathfrak{S})$ is normalized.)

Now suppose $d \geq 2$. By induction, for all t, there exists $x_{d-1}^t \in E_{S_d}$ such that for all $i \leq d-1$, the projection $\pi_{S_i}(x_{d-1}^t)$ coarsely coincides with $\gamma_i(a_{S_i}^p \cdot t)$, and also $\pi_T(x_{d-1}^t)$ coarsely coincides with $\pi_T(x_0)$ for each T orthogonal to each S_i. In view of the quasiisometric embedding $F_{S_d} \times E_{S_d} \to \mathcal{X}$, there exists a point x_d^t such that $\mathfrak{g}_{E_{S_d}}(x_d^t)$ coarsely coincides with x_{d-1}^t and $\pi_{S_d}(x_d^t)$ coarsely coincides with $\gamma_d(a_{S_d}^p \cdot t)$. (Here, $\mathfrak{g}_{E_{S_d}}$ is the gate map defined at the end of Section 1.) For each sufficiently large t, the point x_d^t lies in $\mathcal{N}_{\{U_{S_i}\},\epsilon}(p)$, as required. □

Remark 2.18 By regarding each $\partial\mathcal{C}U$, with $U \in \mathfrak{S}$, as a discrete set, we can endow $\partial(\mathcal{X}, \mathfrak{S})$ with an alternate topology as a simplicial complex, as follows. For each $U \in \mathfrak{S}$ and each $p \in \partial\mathcal{C}U$, we have a 0–simplex, and the 0–simplices $p_i \in \partial\mathcal{C}U_i$ for $i = 0, \dots, k$ span a k–simplex if $U_i \perp U_j$ for $0 \leq i < j \leq k$. There is an obvious bijection from the resulting simplicial complex to $\partial(\mathcal{X}, \mathfrak{S})$, which is an embedding on each simplex.

3 Compactness for proper HHS

In this section, we will prove that proper HHSs have compact HHS boundaries.

3.1 Preliminary lemmas

Definition 3.1 Let $(X,\mathfrak{S})$ be hierarchically hyperbolic. The *level* ℓ_U of $U\in\mathfrak{S}$ is defined inductively as follows. If U is $\sqsubseteq$–minimal, then $\ell_U=1$. We inductively define $\ell_U=k+1$ if k is the maximal integer such that there exists $V\sqsubseteq U$ with $\ell_V=k$ and $V\neq U$.

The following is a slightly modified version of Lemma 2.5 in [6].

Lemma 3.2 *Let $(X,\mathfrak{S})$ be hierarchically hyperbolic. Then there exists N with the following property: Let $x,y\in\mathcal{X}$ and let $\{S_i\}_{i=1,\dots,N}\subseteq\mathfrak{S}$ be such that $\mathsf{d}_{\mathcal{C}S_i}(x,y)\geq 50E$ for each $i=1,\dots,N$. Then there exist $S\in\mathfrak{S}$ and i such that $S_i\sqsubsetneq S$ and $\mathsf{d}_{\mathcal{C}S}(x,y)\geq 100E$. Moreover, for each $T\in\mathfrak{S}$ such that each $S_i\sqsubseteq T$, we can choose $S\sqsubseteq T$.*

Proof The proof is by induction on the level k of a $\sqsubseteq$–minimal $S\in\mathfrak{S}$ into which each S_i is nested. The base case $k=1$ is empty.

Suppose that the statement holds for a given $N=N(k)$ when the level of S as above is at most k. Suppose instead that $|\{S_i\}|\geq N(k+1)$ (where $N(k+1)$ is a constant much larger than $N(k)$ that will be determined shortly) and there exists a $\sqsubseteq$–minimal $S\in\mathfrak{S}$ of level $k+1$ into which each S_i is nested. There are two cases.

If $\mathsf{d}_{\mathcal{C}S}(x,y)\geq 100E$, then we are done. If not, then the large link axiom (Definition 1.1(6)) implies that there exists $K=K(100E)$ and $T_1,\dots,T_K$, each properly nested into S (and hence of level less than $k+1$), so that any S_i is nested into some T_j. In particular, if $N(k+1)\geq KN(k)$, there exists j such that at least $N(k)$ elements of $\{S_i\}$ are nested into T_j. By the induction hypothesis, we are done.

Note that the proof still works replacing $\mathfrak{S}$ with $\mathfrak{S}_T$ when each $S_i\sqsubseteq T$. In this case, we can take $S\sqsubseteq T$ and the T_i produced by the large link axiom will also have $T_i\sqsubseteq S\sqsubseteq T$ for each i, as required for the second statement. □

Lemma 3.3 *Let $(X,\mathfrak{S})$ be hierarchically hyperbolic. Then for every hierarchy ray γ there exists $S\in\mathfrak{S}$ such that $\pi_S(\gamma)$ is unbounded. Moreover, if $T\in\mathfrak{S}$ has the property that $\{\mathrm{diam}_{\mathcal{C}T'}(\gamma):T'\sqsubseteq T\}$ is unbounded, then there exists $S\sqsubseteq T$ such that $\pi_S(\gamma)$ is unbounded.*

Proof The proof of the "moreover" part is a minor variation; we prove the first assertion and indicate parenthetically how to adapt the proof.

By the distance formula (Theorem 1.9) and the fact that γ is a quasigeodesic, there exists an increasing sequence $\{n_i\}$ of natural numbers such that for each positive integer i, there exists $S_i' \in \mathfrak{S}$ such that $\mathrm{d}_{\mathcal{C}S_i'}(\gamma(n_i), \gamma(n_{i+1})) \geq 100E$. (For the purposes of the "moreover" part, we choose S_i' nested into T.) Since γ is a hierarchy path, it makes coarsely monotonic progress in $\mathcal{C}U$ for each $U \in \mathfrak{S}$, and thus for each $t \geq 0$ we have

$$\mathrm{d}_{\mathcal{C}U}(\gamma(0), \gamma(t)) \geq 50E \cdot |\{i : n_i \leq t,\ S_i' = U\}|.$$

Let $\mathcal{S} \subset \mathfrak{S}$ be the collection of domains in which γ makes significant progress; that is, $\mathcal{S}$ is the set of all $S \in \mathfrak{S}$ for which there exists $t_S \geq 0$ such that for any $t \geq t_S$ we have $\mathrm{d}_{\mathcal{C}S}(\gamma(0), \gamma(t)) \geq 50E$. (In the proof of the "moreover" part, we further require that S is nested into T.) If $|\mathcal{S}| < \infty$, then we are done by the above inequality, so assume $|\mathcal{S}| = \infty$.

Let $S \in \mathfrak{S}$ be $\sqsubseteq$–minimal with the property that there are infinitely many $S' \in \mathcal{S}$ nested into S. (In the proof of the "moreover" part, S is nested into T.) Suppose for a contradiction that $\mathrm{diam}_S(\pi_S(\gamma)) = D < \infty$.

Denote by $\mathcal{S}^j$ the set of all level-j elements of $\mathcal{S}$ nested into S, and let k be maximal with the property that $\mathcal{S}^k$ is infinite. Note that this assumption and finite complexity imply that $\bigcup_{k'>k} \mathcal{S}^{k'}$ is finite. To derive a contradiction, we will use the large link axiom and Lemma 3.2 to construct an infinite sequence of distinct $S_i \in \bigcup_{k'>k} \mathcal{S}^{k'}$.

By the large link axiom (Definition 1.1(6)), there exists $K = K(D)$ such that, for any t, there exist $T_1^t, \ldots, T_K^t$ properly nested into S such that if $X \in \mathcal{S}$ has $X \sqsubseteq S$ and $t_X \leq t$, then $X \sqsubseteq T_j^t$ for some j. If we take t_0 large enough, we can apply Lemma 3.2 to a sufficiently large subset of $\mathcal{S}^k$, all of whose elements are nested into some $T_j^{t_0}$, and we get some S_0 of level $k_0 > k$ such that $\mathrm{d}_{\mathcal{C}S_0}(\gamma(0), \gamma(t)) \geq 100E$ for $t \geq t_0$. Note that Lemma 3.2 allows us to take $S_0 \sqsubseteq T_j^{t_0}$, so that $S_0 \sqsubseteq S$ and thus $S_0 \in \mathcal{S}^{k_0}$. By minimality of S, there are finitely many elements of $\mathcal{S}^k$ nested into S_0. We can now choose $t_1 > t_0$ and apply Lemma 3.2 to a sufficiently large subset of $\mathcal{S}^k$ all of whose elements are nested into some $T_j^{t_1}$ but not nested into S_0, and get another element $S_1 \in \mathcal{S}^{k_1}$ for some $k_1 > k$ which is properly nested into S. We can then proceed inductively and construct infinitely many distinct elements $S_i \sqsubseteq S$ of level greater than k, giving us our contradiction. □

3.2 Compactness

We are ready to prove:

Theorem 3.4 *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic, and let $\overline{\mathcal{X}} = \mathcal{X} \cup \partial(\mathcal{X}, \mathfrak{S})$. If $\mathcal{X}$ is proper, then $\overline{\mathcal{X}}$ is compact.*

Proof It suffices to show that $\overline{\mathcal{X}}$ is sequentially compact since it is separable by Proposition 2.17. We will first show that any internal sequence $\{x_n\} \subset \mathcal{X}$ subconverges to some point in $\overline{\mathcal{X}}$. Then we will show this suffices for the theorem.

Internal sequences subconverge Let $\{x_n\} \subset \mathcal{X}$ be a sequence of interior points. For each n, let γ_n be a uniformly Lipschitz hierarchy path between x_0 and x_n, whose existence is guaranteed by Theorem 1.8. Since $\mathcal{X}$ is proper, either the sequence x_n subconverges to an interior point and we are done, or we can assume that the sequence of hierarchy paths γ_n converges to a hierarchy ray, γ_∞.

Lemma 3.3 implies there exists $T \in \mathfrak{S}$ such that $\pi_T \circ \gamma_\infty$ is unbounded. The collection $\{T_i\}_{i=1}^k$ for which this is true must be a collection of pairwise-orthogonal elements by the consistency inequalities (Definition 1.1(4)). For each T_i, the quasigeodesic ray $\pi_{T_i} \circ \gamma_\infty \subset \mathcal{C}T_i$ represents a point $p_{T_i} \in \partial \mathcal{C}T_i$. Set $\overline{T} = \{T_i\}_{i=1}^k$.

We now consider two cases, depending on the behavior of the sequence $\{x_n\}$ in $\overline{T}^\perp$. First, suppose $\liminf_n \sup\{\mathrm{d}_{\mathcal{C}T}(x_0, x_n) : T \in \overline{T}^\perp\} < \infty$. Up to passing to a further subsequence of $\{x_n\}$, we have well-defined limits for $1 \le i, j \le k$,

$$r_{i,j} = \lim_n \frac{\mathrm{d}_{\mathcal{C}T_i}(x_0, x_n)}{\mathrm{d}_{\mathcal{C}T_j}(x_0, x_n)} \in [0, \infty],$$

which determine coefficients $\{a_i^p \in [0,1]\}$ such that $a_i^p / a_j^p = r_{i,j}$ and $\sum a_i^p = 1$. It is straightforward to check that $\{x_n\}$ eventually lies in the interior part of any $\mathcal{N}_{\{U_{T_i}\},\epsilon}(p)$, implying that $\{x_n\}$ subconverges to $p = \sum_{T \in \overline{T}} a_T^p p_T$.

Now suppose that, up to passing to a subsequence,

$$\liminf_n \sup\{\mathrm{d}_{\mathcal{C}T}(x_0, x_n) : T \in \overline{T}^\perp\} = \infty.$$

Consider the sequence $\{y_n\} = \{\mathfrak{g}_{E_{\overline{T}}}(x_n)\}$ of gates in the orthogonal complement of $\overline{T}$.

Since $(E_{\overline{T}}, \mathfrak{S}_{\overline{T}^\perp})$ is an HHS with complexity strictly less than that of $(\mathcal{X}, \mathfrak{S})$, by induction on the complexity of $(\mathcal{X}, \mathfrak{S})$, the sequence $\{y_n\}$ subconverges to $q \in \partial \mathcal{X}$, where $\mathrm{Supp}(q) = \{T_i\}_{i=k+1}^{k'}$ and $T_i \perp T_j$ whenever $i \le k < j$. Since $(E_{\overline{T}}, \mathfrak{S}_{\overline{T}^\perp}) \subset (\mathcal{X}, \mathfrak{S})$ is hierarchically quasiconvex, we can take $q \in \partial E_{\overline{T}}$. For each $j > k$, let $q_{T_j} \in \partial \mathcal{C}T_j$, so that q is a linear combination of the q_{T_j}. As before, up to passing to a further subsequence, for any $1 \le i, j \le k'$, we can define

$$r_{i,j} = \lim_n \frac{\mathrm{d}_{\mathcal{C}T_i}(x_0, x_n)}{\mathrm{d}_{\mathcal{C}T_j}(x_0, x_n)} \in [0, \infty],$$

which determine coefficients $\{a_{T_i}^p\}_{i=1}^k \cup \{a_{T_j}^q\}_{j=k+1}^{k'}$ such that

- $a_{T_i}^r / a_{T_j}^{r'} = r_{i,j}$ when $r, r' \in \{p, q\}$ and $a_{T_i}^r$ and $a_{T_j}^{r'}$ are defined, and
- $\sum_{i=1}^k a_{T_i}^p + \sum_{j=k+1}^{k'} a_{T_j}^q = 1$.

If some $a^r_{T_i}=0$ for $r\in\{p,q\}$, we disregard T_i. We now claim that $\{x_n\}$ (sub)converges to

$$p=\sum_{i=1}^{k} a^p_{T_i} p_{T_i} + \sum_{i=k+1}^{k'} a^q_{T_i} q_{T_i}.$$

Pick a neighborhood $\mathcal{N}_{\{U_{T_i}\},\epsilon}(p)$ of p. For large enough n, we have $x_n\in\mathcal{N}_{\{U_{T_i}\},\epsilon}(p)$ because

- $\pi_{T_i}(x_n)\in U_{T_i}$ for $i\le k$ since $(\pi_{T_i}(x_n)\,|\,p_{T_i})_{\pi_{T_i}(x_0)}\to\infty$,
- $\pi_{T_i}(x_n)\in U_{T_i}$ for $i>k$ since $\pi_{T_i}(x_n)$ coarsely equals $\pi_{T_i}(y_n)$ and $y_n\to q$,
- $|a^r_{T_j}/a^{r'}_{T_i}-\mathsf{d}_{T_j}(x_0,x_n)/\mathsf{d}_{T_i}(x_0,x_n)|<\epsilon$ by definition, when $r,r'\in\{p,q\}$ and $a^r_{T_i}$ and $a^{r'}_{T_j}$ are defined, and
- $\mathsf{d}_T(x_0,x_n)/\mathsf{d}_{T_i}(x_0,x_n)<\epsilon$ for $T\in(\{T_i\}_{i=1}^{k'})^\perp$ and any $1\le i\le k'$, as we now show.

Let $T\in(\{T_i\}_{i=1}^{k'})^\perp$ and choose i so that $a^r_{T_i}\neq 0$ for $r\in\{p,q\}$. Observe that

$$\frac{\mathsf{d}_T(x_0,x_n)}{\mathsf{d}_{T_i}(x_0,x_n)}=\frac{\mathsf{d}_T(x_0,x_n)}{\mathsf{d}_{T_{k+1}}(x_0,x_n)}\cdot\frac{\mathsf{d}_{T_{k+1}}(x_0,x_n)}{\mathsf{d}_{T_i}(x_0,x_n)}.$$

The first term on the right-hand side can be made arbitrarily small by increasing n since $\mathsf{d}_T(x_0,x_n)$ and $\mathsf{d}_{T_{k+1}}(x_0,x_n)$ coarsely coincide with $\mathsf{d}_T(x_0,y_n)$ and $\mathsf{d}_{T_{k+1}}(x_0,y_n)$, respectively, and $\{y_n\}$ converges to q. Since the second term converges to $r_{k+1,i}<\infty$, this proves the claim and completes the internal sequence case.

Reduction to the internal sequence case Recall the definition of the boundary projection, Definition 2.8. By passing to a subsequence if necessary, it suffices to consider any boundary sequence $\{z_n\}\subset\partial\mathcal{X}$, where $z_n=\sum_{S\in\mathrm{Supp}(z_n)} a^{z_n}_S p^n_S$ for each n.

We first find $\{x_n\}\subset\mathcal{X}$ with the properties (1)–(7) below, and then verify that $\{z_n\}$ subconverges to the limit of $\{x_n\}$:

(1) $\mathsf{d}_{\mathcal{X}}(x_0,x_n)\ge n$.

(2) $(\pi_S(x_n)\,|\,p^n_S)_{\pi_S(x_0)}\ge n$ for each $S\in\mathrm{Supp}(z_n)$ (we remind the reader that the notation $(\bullet\,|\,\bullet)_\bullet$ denotes the Gromov product with respect to the subscripted basepoint).

(3) $|a^n_S/a^n_{S'}-\mathsf{d}_S(x_0,x_n)/\mathsf{d}_{S'}(x_0,x_n)|<1/n$ for any distinct $S,S'\in\mathrm{Supp}(z_n)$.

(4) $\mathsf{d}_T(x_0,x_n)/\mathsf{d}_S(x_0,x_n)<1/n$ for any $T\in(\mathrm{Supp}(z_n))^\perp$ and $S\in\mathrm{Supp}(z_n)$.

(5) For all n and $S^n\in\mathrm{Supp}(z_n)$, if $T\pitchfork S^n$ or $S^n\sqsubseteq T$, then $\mathsf{d}_T(\rho^{S^n}_T,x_n)<K$ for some uniform $K>0$. Moreover, $\mathsf{d}_T(x_0,x_n)\le\mathsf{d}_{S^n}(x_0,x_n)$ for all such T.

(6) $\{x_n\}$ converges to $p = \sum_{T \in \mathrm{Supp}(p)} a_T^p p_T \in \partial\mathcal{X}$ with the following property: if there are infinitely many n for which $z_n \in \partial^{\mathrm{rem}}\mathcal{X}$ (with respect to $\mathrm{Supp}(p)$), then there are infinitely many remote z_n such that the following holds for some fixed $T \in \mathrm{Supp}(p)$: there exists $S_T^n \in \mathrm{Supp}(z_n)$ such that $S_T^n \pitchfork T$ or $S_T^n \subsetneq T$, or $T \subsetneq S_T^n$ but $\mathsf{d}_{S_T^n}(\rho^T_{S_T^n}, x_0) \le 100K'E$ for some constant $K' \ge 1$ depending on $\{z_n\}$ and p but not on n. Moreover, $\mathsf{d}_{\mathcal{C}T}(x_0, x_n) \le \mathsf{d}_{\mathcal{C}S^n}(x_0, x_n)$ for all such T.

(7) $\{x_n\}$ converges to $p = \sum_{T \in \mathrm{Supp}(p)} a_T^p p_T \in \partial\mathcal{X}$ with the following property: if there are infinitely many n for which $z_n \in \partial^{\mathrm{rem}}\mathcal{X}$ (with respect to $\mathrm{Supp}(p)$), then there are infinitely many remote z_n such that $\mathsf{d}_T((\partial\pi_{\mathrm{Supp}(p)}(z_n))_T, x_n) \le K''$ for some K'' independent of n and all $T \in \mathrm{Supp}(p)_{z_n}$. Moreover, $\mathsf{d}_{\mathcal{C}T}(x_0, x_n) \le \mathsf{d}_{\mathcal{C}S^n}(x_0, x_n)$ for all such T.

To see that such an internal sequence exists, choose a sequence $\{x_n\}$ so that $x_n \in P$ for all n, where

$$P = \mathrm{im}\biggl(\prod_{S \in \mathrm{Supp}(z_n)} F_S \to \mathcal{X}\biggr);$$

the sequence $\{x_n\}$ satisfies (1)–(4) (which can be done since they are componentwise conditions); and

$$\min_{S \in \mathrm{Supp}(z_n)} \frac{\mathsf{d}_{\mathcal{X}}(\mathfrak{g}_{F_S}(x_n), x_0)}{\mathsf{d}_{\mathcal{X}}(\mathfrak{g}_{F_S}(x_0), x_0)} \to \infty$$

as $n \to \infty$. Here we fix, for each n, a basepoint $(p_S)_{S \in \mathrm{Supp}(z_n)}$ and let $F_S = F_S \times \{(P_{S'})_{S' \neq S}\}$.

(Recall from [6, Remark 5.12] that, whenever $U_1, \dots, U_k \in \mathfrak{S}$ are pairwise orthogonal, we have a standard quasiisometric embedding $\prod_{i=1}^k F_{U_i} \to \mathcal{X}$ whose image is hierarchically quasiconvex and which is, for each $i \le k$, the restriction of the usual map $F_{U_i} \times E_{U_i} \to \mathcal{X}$.)

We can verify condition (5) by examining the product regions $\prod_{S \in \mathrm{Supp}(z_n)} F_S \to \mathcal{X}$. Let $T \pitchfork S^n$ or $S^n \subsetneq T$ for $S^n \in \mathrm{Supp}(z_n)$. Since x_n coarsely lies in $\prod_{S \in \mathrm{Supp}(z_n)} F_S$, it follows that $\mathrm{diam}_T(\rho_T^{S_n} \cup \pi_T(F_{S^n})) \asymp 1$ and $\mathsf{d}_T(\pi_T(F_{S^n}), x_n) \asymp 1$. We thus have, for some uniform C,

$$\mathsf{d}_T(x_0, x_n) \le C\,\mathsf{d}_{\mathcal{X}}\biggl(x_0, \prod_{S \in \mathrm{Supp}(z_n)} F_S\biggr) + C.$$

For sufficiently large n, our choice of $\{x_n\}$ ensures that

$$\mathsf{d}_{S_n}(x_0, x_n) \ge C\,\mathsf{d}\biggl(x_0, \prod_{S \in \mathrm{Supp}(z_n)} F_S\biggr) + C,$$

verifying the "moreover" part of assertion (5).

Let $\{x_n\}$ satisfy (1)–(5). We now prove there is a subsequence of $\{x_n\}$ satisfying (6).

By replacing $\{x_n\}$ with a subsequence (and replacing $\{z_n\}$ with the corresponding subsequence of $\{z_n\}$), we can apply the proof that internal sequences subsequentially converge to conclude $\{x_n\}$ converges to $p = \sum_{T \in \mathrm{Supp}(p)} a_T^p p_T \in \partial \mathcal{X}$.

Consider the set $\mathbb{G}$ of $n \in \mathbb{N}$ such that z_n is remote with respect to p. If $\mathbb{G}$ is finite, then (6) holds vacuously. Otherwise, by replacing $\mathbb{G}$ with an infinite subset, we find $T \in \mathrm{Supp}(p)$ such that for all $n \in \mathbb{G}$, there exists $S^n \in \mathrm{Supp}(z_n)$ with either $T \pitchfork S^n$ or $S^n \subsetneq T$ or $T \subsetneq S^n$.

First consider the case where $\{S^n : n \in \mathbb{G}\}$ is infinite. By passing to a subsequence if necessary, and then applying finite complexity, Lemma 1.4, and Ramsey's theorem, we can assume that $S^n \pitchfork S^m$ when $n \neq m$. Let $\mathbb{G}_T \subseteq \mathbb{N}$ be the set of $n \in \mathbb{G}$ such that $T \subsetneq S^n$. Then for all $m, n \in \mathbb{G}_T$, we have $\mathsf{d}_{S^m}(\rho^T_{S^m}, \rho^{S^n}_{S^m}) \le E$ by the consistency inequalities. Hence, again by the consistency inequalities and the triangle inequality, we have $\mathsf{d}_{S^n}(\rho^T_{S^n}, x_0) \le 2E$ for all but at most one element of $\mathbb{G}_T$. Indeed, if $\mathsf{d}_{S^n}(\rho^T_{S^n}, x_0) > 2E$, then $\mathsf{d}_{S^n}(\rho^{S^m}_{S^n}, x_0) > E$ for any $m \in \mathbb{G}_T - \{n\}$, so by consistency $\mathsf{d}_{S^m}(\rho^{S^n}_{S^m}, x_0) \le E$; the claim follows from the triangle inequality since $\mathsf{d}_{S^m}(\rho^T_{S^m}, \rho^{S^n}_{S^m}) \le E$. Hence, by replacing $\{z_n\}$ with a subsequence, for all $T \in \mathrm{Supp}(p)$ with $T \subsetneq S^n$, we have $\mathsf{d}_{S^n}(\rho^T_{S^n}, x_0) \le 100K'E$. Letting $S^n_T = S^n$ for $n \in \mathbb{G}$, this establishes assertion (6) when $\{S^n : n \in \mathbb{G}\}$ is infinite.

When $\{S^n : n \in \mathbb{G}\}$ is finite, we can assume that $S^n = S^m$ for all m, n by passing to a subsequence. Hence, there exists $S \in \mathfrak{S}$ such that for all $n \in \mathbb{G}$, and all $U \in \mathrm{Supp}(z_n)$, either $U = S$ or $U \perp T$. Fix T and S as above, and replace (z_n) with a subsequence so that for each $n \in \mathbb{G}$, we have $S \in \mathrm{Supp}(z_n)$. Then, for each $n \in \mathbb{G}$, set $S^n_T = S$ and observe that either $S \sqsubseteq T$, $S \pitchfork T$ or $T \sqsubseteq S$. In the latter case, take $K' = \mathsf{d}_S(\rho^T_S, x_0)$, which depends on p and $\{z_n\}$ but not on n. This completes the proof of (6).

We now deduce condition (7) from (1)–(6). Assume $\mathbb{G}$ is infinite, so that, by (6), there exists $T' \in \mathrm{Supp}(p)$ such that, after replacing $\mathbb{G}$ with an infinite subset if necessary, we have, for each $n \in \mathbb{G}$, some $S^n_{T'} \in \mathrm{Supp}(z_n)$ such that $\mathsf{d}_{S^n_{T'}}(\rho^T_{S^n_{T'}}, x_0) \le 100K'E$. Let $T \in \mathrm{Supp}(p)_{z_n}$. First suppose that $T \subsetneq S^n_{T'}$. Then, since $T \perp T'$ or $T = T'$, Lemma 1.5 implies that $\mathsf{d}_{S^n_{T'}}(\rho^T_{S^n_{T'}}, x_0) \le 200K'E$. It follows from (2) that

$$(\pi_{S^n_{T'}}(x_n) \,|\, p^n_{S^n_{T'}})_{\rho^T_{S^n_{T'}}} \to \infty$$

as $n \to \infty$, so that, by discarding finitely many n and applying the bounded geodesic image axiom, we have $\mathsf{d}_T\big((\partial\pi_{\mathrm{Supp}(p)}(z_n))_T, x_n\big) \le E$ for all $n \in \mathbb{G}$. In the remaining cases, where $T \pitchfork S^n_{T'}$ or $S^n_{T'} \subsetneq T$, we reach the same conclusion, using (5) instead of (6). This completes the proof of condition (7).

Subconvergence of $\{z_n\}$ Fix a neighborhood $\mathcal{N} = \mathcal{N}_{\{U_S\},\epsilon}(p)$ of p; we must check that for infinitely many values of n, we have $z_n \in \mathcal{N}$. For each n, either $z_n \in \partial^{\text{rem}}\mathcal{X}$ (recall that this means that $\operatorname{Supp}(z_n) \cap \operatorname{Supp}(p) = \varnothing$ and for all $T \in \operatorname{Supp}(p)$, there exists $S \in \operatorname{Supp}(z_n)$ with $T \not\perp S$) or $z_n \in \partial\mathcal{X} - \partial^{\text{rem}}\mathcal{X}$ (so that either $\operatorname{Supp}(z_n) \cap \operatorname{Supp}(p) \neq \varnothing$ or there exists $T \in \operatorname{Supp}(p)$ with $T \perp S$ for all $S \in \operatorname{Supp}(z_n)$).

The nonremote case We will consider the nonremote case first. Recall that $z_n = \sum_{S \in \operatorname{Supp}(z_n)} a_S^{z_n} p_S^n$. We must check the following conditions:

(a) For each $S \in \operatorname{Supp}(p) \cap \operatorname{Supp}(z_n)$, and infinitely many n, we have $p_S^n \in U_S$.

(b) For each $S \in \operatorname{Supp}(p) \cap \operatorname{Supp}(z_n)$ and infinitely many n, we have $a_S^n \to a_S^p$.

(c) $\sum_{T \in \operatorname{Supp}(p) - \operatorname{Supp}(z_n)} a_T^p < K'\epsilon$ for infinitely many n and some uniform K'.

Up to passing to a subsequence, (a) follows from (2) and the fact that $x_n \to p$.

For (b), we have three cases. If $\operatorname{Supp}(p) \cap \operatorname{Supp}(z_n) = \varnothing$, then this holds vacuously. If $\operatorname{Supp}(p) \cap \operatorname{Supp}(z_n)$ has multiple elements, then this follows from (3) and the fact that $x_n \to p$. If $\operatorname{Supp}(p) \cap \operatorname{Supp}(z_n) = \{S\}$, then this follows from (3) and (c), proved momentarily.

To see (c), first observe that $\operatorname{Supp}(p) - \operatorname{Supp}(z_n) \subset (\operatorname{Supp}(z_n))^{\perp}$ by nonremoteness. Let $T \in \operatorname{Supp}(p) - \operatorname{Supp}(z_n)$ and $S \in \operatorname{Supp}(p) \cap \operatorname{Supp}(z_n)$; note that such an $S \in \operatorname{Supp}(p) \cap \operatorname{Supp}(z_n)$ exists, otherwise one of $x_n \to p$ or (4) is contradicted. By definition of $x_n \to p$,

$$\left| \frac{a_T^p}{a_S^p} - \frac{\mathrm{d}_T(x_0, x_n)}{\mathrm{d}_S(x_0, x_n)} \right| < \epsilon.$$

It follows from (4) that $\mathrm{d}_T(x_0, x_n)/\mathrm{d}_S(x_0, x_n) < 1/n$. Since each $a_S^p \le 1$, it follows that

$$\sum_{T \in \operatorname{Supp}(p) - \operatorname{Supp}(z_n)} a_T^p < \xi(\mathcal{X})\Big(\epsilon + \frac{1}{n}\Big) \le 2\xi(\mathcal{X})\epsilon,$$

completing the proof of (c) and thus the nonremote case.

The remote case We must check the following conditions:

(i) For any $T \in \operatorname{Supp}(p)$, and infinitely many n, we have $(\partial\pi_{\operatorname{Supp}(p)}(z_n))_T \in U_T$.

(ii) For infinitely many n and any $T \in \operatorname{Supp}(p)_{z_n}$, $T' \in \operatorname{Supp}(p)$, we have

$$\left| \frac{\mathrm{d}_T(x_0, (\partial\pi_{\operatorname{Supp}(p)}(z_n))_T)}{\mathrm{d}_{T'}(x_0, (\partial\pi_{\operatorname{Supp}(p)}(z_n))_{T'})} - \frac{a_T^p}{a_{T'}^p} \right| < \epsilon.$$

(iii) We have $\sum_{T \in \operatorname{Supp}(p)^{\perp} \cap \operatorname{Supp}(z_n)} a_T^{z_n} < K\epsilon$ for some uniform K.

For any $T \in \mathrm{Supp}(p)$ and each n, choose $S^n_T \in \mathrm{Supp}(z_n)$ so that T and S^n_T are not orthogonal. If $\mathbb{G}$ is infinite, then we may pass to a subsequence such that S^n_T and T are always nonorthogonal: that is, $T \subsetneq S^n_T$, or $T \pitchfork S^n_T$, or $S^n_T \subsetneq T$.

We now show that assertion (i) holds for infinitely many n; the proof divides into three cases according to the above possibilities, which influence the definition of $(\partial\pi_{\mathrm{Supp}(p)}(z_n))_T$.

First, if $S^n_T \pitchfork T$, then $(\partial\pi_{\mathrm{Supp}(p)}(z_n))_T = \rho_T^{S^n_T}$. In this case, (i) follows immediately from conditions (2) and (5) in the definition of $\{x_n\}$. The same is true if $S^n_T \subsetneq T$. If $T \sqsubseteq S^n_T$, then (i) follows from (2), (7) and the triangle inequality.

Assertion (ii), in the case when $T, T' \in \mathrm{Supp}(p)$, follows from (7). In fact, since $\{x_n\}$ converges to p, we have

$$(*) \qquad \left| \frac{\mathsf{d}_T(x_0, x_n)}{\mathsf{d}_{T'}(x_0, x_n)} - \frac{a^p_T}{a^p_{T'}} \right| \to 0,$$

and $\mathsf{d}_T(x_0, x_n) \to \infty$, $\mathsf{d}_{T'}(x_0, x_n) \to \infty$. By (7), we have that $\mathsf{d}_T(x_0, x_n)$ coarsely coincides with $\mathsf{d}_T(x_0, (\partial\pi_{\mathrm{Supp}(p)}(z_n))_T)$, and similarly for T'. Hence, $(*)$ implies that the ratio in assertion (ii) satisfies the required inequality. If $T \in \mathrm{Supp}(p)_{z_n} - \mathrm{Supp}(p)$, then we have to verify $\left|\mathsf{d}_T\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(z_n))_T\big)/\mathsf{d}_{T'}\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(z_n))_{T'}\big)\right| \to 0$. We still know $(*)$ (with $a^p_T/a^p_{T'}$ replaced by 0) and $\mathsf{d}_{T'}(x_0, x_n) \to \infty$. If $\mathsf{d}_T(x_0, x_n)$ does not diverge, we are done. If it does, we can approximate $\mathsf{d}_T\big(x_0, (\partial\pi_{\mathrm{Supp}(p)}(z_n))_T\big)$ by $\mathsf{d}_T(x_0, x_n)$ and we can conclude as above.

It remains to verify assertion (iii). For each n, let $T^n \in (\mathrm{Supp}(p))^\perp \cap \mathrm{Supp}(z_n)$ and choose $S^n \in \mathrm{Supp}(z_n) - (\mathrm{Supp}(p))^\perp$. Fix $P \in \mathrm{Supp}(p)$ so that, after passing to a subsequence, P is not orthogonal to any of the S^n. By either (5) or (7), we have $\mathsf{d}_{\mathcal{C}S^n}(x_0, x_n)/\mathsf{d}_{\mathcal{C}P}(x_0, x_n) \le 1$, while $\mathsf{d}_{\mathcal{C}P}(x_0, x_n)/\mathsf{d}_{\mathcal{C}T^n}(x_0, x_n) < \epsilon$ since $x_n \to p$. Hence $a^{z_n}_{T^n}/a^{z_n}_{S^n} \le \epsilon + 1/n$, by (3), and the desired inequality follows since the number of terms in the sum is bounded by $\xi(\mathcal{X})$, as in the nonremote case. This completes the proof that $\{z_n\}$ subconverges to p, and thus completes the proof that $\partial\mathcal{X}$ is compact. $\square$

4 The HHS boundary of a Gromov-hyperbolic space

In this section, we prove that the HHS boundary of a hyperbolic space is its Gromov boundary, regardless of the chosen HHS structure.

Lemma 4.1 *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic. If $\mathcal{X}$ is hyperbolic, then there exists $C > 0$ such that if $U, V \in \mathfrak{S}$ and $U \perp V$, then either* $\mathrm{diam}\,\mathcal{C}U < C$ *or* $\mathrm{diam}\,\mathcal{C}V < C$.

Proof Recall from [6] that if $U \perp V$, then there exists a quasiisometric embedding $F_U \times F_V \hookrightarrow \mathcal{X}$. Hyperbolicity uniformly bounds the diameter of one of the factors. $\square$

Lemma 4.2 *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic and let $\mathcal{X}$ be hyperbolic. If $\gamma\colon [0,\infty) \to \mathcal{X}$ is a hierarchy ray with $\gamma(0) = x_0$, then there exists a unique $U \in \mathfrak{S}$ with $\pi_U \circ \gamma\colon [0,\infty) \to \mathcal{C}U$ a parametrized quasigeodesic ray. In particular,* $\operatorname{diam}_{\mathcal{C}V}(\gamma) < \infty$ *for all $V \in \mathfrak{S}$ with $V \neq U$.*

Proof By Lemma 3.3, there exists $U \in \mathcal{C}S$ such that $\operatorname{diam}_{\mathcal{C}U}(\gamma)$ is unbounded. Let $V \in \mathfrak{S}$ be such that $V \neq U$; by Lemma 4.1, there are three cases: $V \sqsubseteq U$, $U \sqsubseteq V$ and $V \pitchfork U$.

Let $t_M \in [0,\infty)$ be such that $\mathsf{d}_{\mathcal{C}U}(\gamma(0), \gamma(t)) > E^2$ for $t \geq t_M$. If $U \sqsubseteq V$, then by the consistency inequality, $\mathsf{d}_V\big(\gamma(t), \rho^V_U(\gamma(0))\big) < E$ for all $t > t_M$. If $V \sqsubseteq U$, then $\mathsf{d}_{\mathcal{C}V}(\gamma(t), \rho^U_V) < E$ for all $t > t_M$. Similarly, if $U \pitchfork V$, then $\mathsf{d}_{\mathcal{C}V}(\gamma(t), \rho^V_U) < E$ for all $t > t_M$ by the transverse case of the consistency inequality. Thus, in each case, $\operatorname{diam}_{\mathcal{C}V}(\gamma) < \infty$. $\square$

Theorem 4.3 *Let $(\mathcal{X}, \mathfrak{S})$ be hierarchically hyperbolic and suppose that $\mathcal{X}$ is hyperbolic. Let $\overline{\mathcal{X}}^{\mathrm{Gr}} = \mathcal{X} \cup \partial^{\mathrm{Gr}}\mathcal{X}$, where $\partial^{\mathrm{Gr}}\mathcal{X}$ is the Gromov boundary of $\mathcal{X}$, and let $\overline{\mathcal{X}} = \mathcal{X} \cup \partial\mathcal{X}$. Then the identity map $\mathcal{X} \to \mathcal{X}$ extends uniquely to a homeomorphism $\overline{\mathcal{X}}^{\mathrm{Gr}} \to \overline{\mathcal{X}}$.*

Proof Lemma 4.1 gives $\partial\mathcal{X} = \bigsqcup_{U \in \mathfrak{S}} \partial\mathcal{C}U$ and Lemma 4.2 gives $|\operatorname{Supp}(p)| = 1$ for all $p \in \partial\mathcal{X}$.

Fix $x_0 \in \mathcal{X}$ and let $p \in \partial^{\mathrm{Gr}}\mathcal{X}$. Let $\gamma_p\colon [0,1) \to \overline{\mathcal{X}}^{\mathrm{Gr}}$ be a geodesic from x_0 to p. For any $n \in \mathbb{N}$, let $\gamma_n\colon [0,n) \to \mathcal{X}$ be a hierarchy path between x_0 and $\gamma_p(n)$. Since $\mathcal{X}$ is hyperbolic, each γ_n uniformly fellow-travels γ_p and thus $\gamma = \lim_n \gamma_n$ is a hierarchy ray from x_0 to p. The ray γ is independent of the choice of (γ_n) and is thus uniquely determined by p. By Lemma 4.2, there exists a unique $U \in \mathfrak{S}$ such that $\operatorname{diam}_{\mathcal{C}U}(\gamma)$ is an unbounded quasigeodesic ray. By hyperbolicity of $\mathcal{C}U$, there exists $q \in \partial\mathcal{C}U$ such that $\pi_{\mathcal{C}U}(\gamma)$ limits to q.

The above discussion yields a well-defined map $\phi^{\mathrm{Gr}}\colon \partial^{\mathrm{Gr}}\mathcal{X} \to \partial\mathcal{X}$ given by $\phi^{\mathrm{Gr}}(p) = q$. Define $\phi\colon \overline{\mathcal{X}}^{\mathrm{Gr}} \to \overline{\mathcal{X}}$ by $\phi|_{\mathcal{X}} = \mathrm{id}_{\mathcal{X}}$ and $\phi|_{\overline{\mathcal{X}}^{\mathrm{Gr}}} = \phi^{\mathrm{Gr}}$. We claim that ϕ is a homeomorphism.

Bijectivity The map ϕ is clearly bijective on $\mathcal{X}$. Let $p, q \in \partial^{\mathrm{Gr}}\mathcal{X}$ and suppose that $p \neq q$. Then there exist geodesic rays $\gamma_p, \gamma_q\colon [0,\infty] \to \mathcal{X}$ with $[\gamma_p] = p$, $[\gamma_q] = q$ and $\gamma_p(0) = \gamma_q(0) = x_0$. Since $p \neq q$, hyperbolicity of $\mathcal{X}$ implies that $\mathsf{d}_{\mathcal{X}}(\gamma_p(t), \gamma_q(t)) \to \infty$.

By Lemma 4.2, γ_p and γ_q have unique domains U_p and U_q, respectively, to which they have unbounded projections. If $U_p \neq U_q$, we are done. Otherwise, $U_p = U_q = U$, and Lemma 4.2, the distance formula, and the triangle inequality imply that $d_U(\gamma_p(t), \gamma_q(t)) \to \infty$, whence $\phi(p) \neq \phi(q)$, by definition. Thus ϕ is injective; surjectivity of ϕ follows from Theorem 1.7.

Basic sets in $\overline{\mathcal{X}}$ For convenience, we describe basic sets $\mathcal{N}(p)$ for $p \in \partial(\mathcal{X}, \mathfrak{S})$, in our current simple situation. Observe that $\mathrm{Supp}(p)$ consists of a single $S \in \mathfrak{S}$, while $\partial^{\mathrm{rem}}_{\mathrm{Supp}(p)}\mathcal{X}$ consists of those $q \in \partial(\mathcal{X}, \mathfrak{S})$ with $\mathrm{Supp}(q) = \{T\}$ with $T \neq S$. It is automatic that T is not orthogonal to S if $T \perp S$, then Lemma 4.1 implies only one of $\mathcal{C}S$ or $\mathcal{C}T$ can be unbounded and thus have nonempty Gromov boundary. It follows that $\mathrm{Supp}(q) \cap (\mathrm{Supp}(p))^{\perp} = \varnothing$.

Choosing $\epsilon > 0$ and $p \in \mathcal{U}_S \subset \mathcal{C}S \cup \partial\mathcal{C}S$, a remote neighborhood of p in $\overline{\mathcal{X}}$ is

$$\mathcal{N}^{\mathrm{rem}}_{\mathcal{U}_S,\epsilon}(p) = \left\{ q \in \bigsqcup_{S \neq T} \partial\mathcal{C}T \;\middle|\; \rho^T_S \in \mathcal{U}_S \right\}.$$

Meanwhile, the nonremote part of the boundary is just $\partial\mathcal{C}S$, so

$$\mathcal{N}^{\mathrm{non}}_{\mathcal{U}_S,\epsilon}(p) = \mathcal{U}_S.$$

Finally, the interior part is

$$\mathcal{N}^{\mathrm{int}}_{\mathcal{U}_S,\epsilon}(p) = \left\{ x \in \mathcal{X} \;\middle|\; \pi_S(x) \in U_S,\ \frac{\mathrm{d}_T(x_0, x)}{\mathrm{d}_S(x_0, x)} < \epsilon \text{ for all } T \perp S \right\}.$$

The above descriptions will be useful in proving that ϕ is a homeomorphism.

Continuity of ϕ and ϕ^{-1} Choose $p \in \partial(\mathcal{X}, \mathfrak{S})$, supported on $S \in \mathfrak{S}$, a neighborhood $\mathcal{U}_S$ of $p \in \partial\mathcal{C}S$, and $\epsilon > 0$. We may assume that

$$\mathcal{U}_S = \left\{ y \in \mathcal{C}S \cup \partial\mathcal{C}S \mid p_n \to p \text{ and } \liminf_n (y \,|\, \pi_S(p_n))_{\pi_S(x_0)} > r \text{ for some } (p_n) \right\}$$

for some $r \geq 0$. Choose $q \in \partial^{\mathrm{Gr}}\mathcal{X}$ so that $\phi(q) = p$. For each $r' \geq 0$, let

$$\boldsymbol{U}(q, r') = \{ y \in \mathcal{X} \cup \partial^{\mathrm{Gr}}\mathcal{X} \mid (y \,|\, q)_{x_0} \geq r' \}.$$

Recall that sets of this type yield a neighborhood basis in $\overline{\mathcal{X}}^{\mathrm{Gr}}$.

We exhibit $r' \geq 0$, depending on p, r, ϵ and the distance formula constants, such that

$$\phi(\boldsymbol{U}(q, r')) \subseteq \mathcal{N}_{\mathcal{U}_S,\epsilon}(p).$$

Indeed, if $y \in \boldsymbol{U}(q, r') \cap \partial^{\mathrm{Gr}}\mathcal{X}$, and r' is sufficiently large, then any geodesic ray or segment representing $[\pi_S \circ \gamma_y]$ has an initial segment of length at least r lying 2δ–close to the corresponding segment for p. This implies that $\phi(y) \in \mathcal{U}_S$, which is exactly the nonremote part of $\mathcal{N}_{\mathcal{U}_S,\epsilon}(p)$ (regardless of the choice of ϵ). If $y \in \boldsymbol{U}(q, r')$ is an interior point, and r' is sufficiently large, then similarly $\pi_S(x) \in \mathcal{U}_S$.

If $T \perp S$, then, by Lemma 4.1, there exists a uniform $C > 0$ such that $\mathrm{d}_T(x_0, y) \leq C$. Moreover, choosing r' sufficiently large compared to r, C and the constants in the distance formula, we have $\mathrm{d}_S(x_0, y) \geq C/\epsilon$. Hence either y is interior or $y \in \partial \mathcal{C}S$, and so

$$\phi(\boldsymbol{U}(q, r')) \subseteq \mathcal{N}^{\mathrm{non}}_{\mathcal{U}_S,\epsilon}(p) \cup \mathcal{N}^{\mathrm{int}}_{\mathcal{U}_S,\epsilon}(p).$$

Continuity follows easily: Given an open set $\mathcal{O} \subseteq \overline{\mathcal{X}}$, let $q \in \phi^{-1}(\mathcal{O})$. Then, since $\mathcal{O}$ is open, it contains a neighborhood $\mathcal{N}$ of $\phi(q)$. The preceding discussion shows that q lies in some neighborhood $\boldsymbol{U}$ which in turn lies in $\phi^{-1}(\mathcal{N}) \subset \phi^{-1}(\mathcal{O})$, so $\phi^{-1}(\mathcal{O})$ is open. Continuity of ϕ^{-1} is proved similarly. □

5 Extending hieromorphisms to the boundary

Hieromorphisms need not extend continuously to the boundary, but under additional hypotheses on the quasiisometries implicit in the hieromorphism, such extensions do exist. However, the class of hieromorphisms that extend continuously to the boundary is contained in a larger class of maps with this property, and, given the examples we study later in this section, it is in our interest to focus on this larger class of maps.

Definition 5.1 (slanted hieromorphism) Let $(\mathcal{X}, \mathfrak{S})$ and $(\mathcal{X}', \mathfrak{S}')$ be hierarchically hyperbolic spaces. A *slanted hieromorphism* $f\colon (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}', \mathfrak{S}')$ consists of

(1) a map $f\colon \mathcal{X} \to \mathcal{X}'$;

(2) a map $\pi(f)\colon \mathfrak{S} \to 2^{\mathfrak{S}'}$ such that $\pi(f)(U)$ is a collection of pairwise-orthogonal elements of $\mathfrak{S}'$ for each $U \in \mathfrak{S}$;

(3) for each $U \in \mathfrak{S}$, a map $\rho(f, U)\colon \mathcal{C}U \to \prod_{V \in \pi(f)(U)} \mathcal{C}V$

such that:

(I) If $U, V \in \mathfrak{S}$ satisfy $U \sqsubsetneq V$, then for each $W' \in \pi(f)(V)$, there exists $W \in \pi(f)(U)$ with $W \sqsubsetneq W'$, and for every $W \in \pi(f)(U)$ there exists (a unique) $W' \in \pi(f)(V)$ with $W \sqsubsetneq W'$.

(II) If $U, V \in \mathfrak{S}$ satisfy $U \perp V$, then $W \perp W'$ for all distinct $W \in \pi(f)(U)$ and $W' \in \pi(f)(V)$.

(III) If $U, V \in \mathfrak{S}$ satisfy $U \pitchfork V$, then for all $W \in \pi(f)(U)$ there exists $W' \in \pi(f)(V)$ with $W \pitchfork W'$ and vice versa.

(IV) Each $\rho(f, U)$ is a (uniform) quasiisometric embedding.

(V) For all $U \in \mathfrak{S}$, the following diagram (uniformly) coarsely commutes:

$$\begin{tikzcd} \mathcal{X} \arrow[r, "f"] \arrow[d, "\pi_U"] & \mathcal{X}' \arrow[d, "\prod_{W\in\pi(f)(U)} \pi_W"] \\ \mathcal{C}U \arrow[r, "\rho(f,U)"] & \prod_{W\in\pi(f)(U)} \mathcal{C}W \end{tikzcd}$$

(VI) If $U, V \in \mathfrak{S}$ satisfy $U \subsetneq V$ or $U \pitchfork V$, then

$$\begin{tikzcd} \mathcal{C}U \arrow[r, "\rho(f,U)"] \arrow[d, "\rho^U_V"] & \prod_{W\in\pi(f)(U)} \mathcal{C}W \arrow[d, "g"] \\ \mathcal{C}V \arrow[r, "\rho(f,V)"] & \prod_{W'\in\pi(f)(V)} \mathcal{C}W' \end{tikzcd}$$

uniformly coarsely commutes, where g is a coarsely constant map such that: if $U \subsetneq V$, then for each $W' \in \pi(f)(V)$, the W'–coordinate of g is $\rho^W_{W'}$ for some (hence any, by Lemma 1.5) $W \in \pi(f)(U)$ with $W \subsetneq W'$, and if $U \pitchfork V$, then for each $W' \in \pi(f)(V)$, the W'–coordinate of g is $\rho^W_{W'}$ for some (hence any) $W \in \pi(f)(U)$ with $W \pitchfork W'$.

(VII) If $V \subsetneq U$, then

$$\begin{tikzcd} \mathcal{C}U \arrow[r, "\rho(f,U)"] \arrow[d, "\rho^U_V"] & \prod_{W\in\pi(f)(U)} \mathcal{C}W \arrow[d, "h"] \\ \mathcal{C}V \arrow[r, "\rho(f,V)"] & \prod_{W'\in\pi(f)(V)} \mathcal{C}W' \end{tikzcd}$$

uniformly coarsely commutes, where the map h is defined as follows: given $(x_{W'})_{W'\in\pi(f)(U)}$, for each $W \in \pi(f)(V)$, the W–coordinate of $h((x_{W'}))$ is $\rho^{W''}_W(x_{W''})$, where W'' is the unique element of $\pi(f)(U)$ with $W \subsetneq W''$.

Remark 5.2 (hieromorphisms are slanted hieromorphisms) Any hieromorphism f is a slanted hieromorphism in which $|\pi(f)(U)| = 1$ for all $U \in \mathfrak{S}$.

Remark 5.3 There is presumably a still more general version of Definition 5.1 encompassing morphisms $f\colon (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}', \mathfrak{S}')$ where $f\colon \mathcal{X} \to \mathcal{X}'$ is a map, $f\colon 2^{\mathfrak{S}} \to 2^{\mathfrak{S}'}$

sends pairwise-orthogonal sets to pairwise-orthogonal sets, and f sends appropriate products of hyperbolic spaces to products of hyperbolic spaces. Simple examples like rotation in $\mathbb{E}^2$ require such a definition in order to be regarded as maps of hierarchically hyperbolic spaces.

Definition 5.4 (coarse similarity) Let M and M' be metric spaces. Then $f\colon M \to M'$ is a (λ,ϵ)*–coarse similarity* if there exist $\lambda > 0$ and $\epsilon \geq 0$ such that for all $p,q \in M$,

$$\lambda \mathsf{d}_M(p,q) - \epsilon \leq \mathsf{d}_{M'}(f(p), f(q)) \leq \lambda \mathsf{d}_M(p,q) + \epsilon.$$

Definition 5.5 (extensible slanted hieromorphism) Let $f\colon (\mathcal{X},\mathfrak{S}) \to (\mathcal{X}',\mathfrak{S}')$ be a slanted hieromorphism. Then f is *extensible* if there exist $0 < \lambda_1 \leq \lambda_2$ and $K < \infty$ such that:

(1) $\pi(f)\colon \mathfrak{S} \to 2^{\mathfrak{S}'}$ is injective.

(2) For all $V \in \mathfrak{S}'$, either there is $U \in \mathfrak{S}$ with $V \in \pi(f)(U)$ or
$$\operatorname{diam}_{\mathcal{C}V}\big(\pi_V(f(\mathcal{X}))\big) \leq K.$$

(3) For all $U \in \mathfrak{S}$ and $W \in \pi(f)(U)$, the composition
$$\mathcal{C}U \xrightarrow{\rho(f,U)} \prod_{V\in\pi(f)(U)} \mathcal{C}V \to \mathcal{C}W$$
is a (λ,λ')–coarse similarity, where the second map is the canonical projection and $\lambda \in [\lambda_1,\lambda_2]$ (λ can depend on U and V) and $\lambda' \geq 0$.

Theorem 5.6 (extending slanted hieromorphisms to the boundary) *Let $(\mathcal{X},\mathfrak{S})$ and $(\mathcal{X}',\mathfrak{S}')$ be hierarchically hyperbolic structures on the spaces $\mathcal{X}$ and $\mathcal{X}'$, respectively. Suppose that $f\colon (\mathcal{X},\mathfrak{S}) \to (\mathcal{X}',\mathfrak{S}')$ is an extensible slanted hieromorphism. Then there is a map $\overline{f}\colon \overline{\mathcal{X}} \to \overline{\mathcal{X}}'$ such that*

(1) *$\overline{f}|_{\mathcal{X}} = f$;*

(2) *$\overline{f}|_{\partial\mathcal{X}}$ is injective;*

(3) *for all $f(p) \in \partial\mathcal{X}'$ and basic neighborhoods $f(p) \in \mathcal{N}$ of $\overline{\mathcal{X}}'$, the set $\overline{f}^{-1}(\mathcal{N})$ contains a basic neighborhood of $p \in \overline{\mathcal{X}}$, ie $\overline{f}$ is continuous at each point in $\partial\mathcal{X}$.*

In particular, if $\mathcal{X}$ is proper, then $\overline{f}|_{\partial\mathcal{X}}$ is an embedding with closed image and, if f is an embedding, then $\overline{f}\colon \overline{\mathcal{X}} \to \overline{\mathcal{X}}'$ is an embedding whose image is closed.

Proof For convenience, when the domains of the various maps are understood, we shall denote each map $f\colon \mathcal{X} \to \mathcal{X}'$, $\pi(f)\colon \mathfrak{S} \to 2^{\mathfrak{S}'}$, and $\rho(f,U)\colon \mathcal{C}U \to \prod_{W\in\pi(f)(U)} \mathcal{C}W$ by f.

Boundary maps on hyperbolic domains Let $U \in \mathfrak{S}$. To each sequence (x_n) in $\mathcal{C}U$, associate the sequence $(f(x_n))_n$ in $\prod_{W \in \mathcal{C}\pi(f)(U)} \mathcal{C}W$. For each $W \in \pi(f)(U)$, let $w_n(W) \in \mathcal{C}W$ be the W–coordinate of $f(x_n)$. Fix a basepoint $x \in \mathcal{C}U$ and $p_W = \pi_W(\rho(f,U)(x)) \in \mathcal{C}W$ for each $W \in \pi(f)(U)$.

Suppose that $(x_n)_n$ represents a point in $\partial\mathcal{C}U$, ie $(x_i \mid x_j)_x \to \infty$ as $i,j \to \infty$. Since $\rho(f,U)$ is a uniform quasiisometric embedding, we have for each $W \in \pi(f)(U)$ that $(w_i(W) \mid w_j(W))_{p_W} \to \infty$ as $i,j \to \infty$. Hence $w_i(W)$ converges to a point $p(W) \in \partial\mathcal{C}W$.

For each $W \in \pi(f)(U)$, choose $\alpha_W \in (0,1]$ so that

$$\frac{\alpha_W}{\alpha_{W'}} = \lim_n \frac{\mathsf{d}_W(p_W, w_n(W))}{\mathsf{d}_{W'}(p_{W'}, w_n(W'))}$$

for all $W, W' \in \pi(f)(W)$, which exists because of the coarse similarity assumption. Then define $p \in \star_{W \in \mathcal{C}\pi(f)(U)} \partial\mathcal{C}W$ to be the linear combination $\sum_{W \in \pi(f)(U)} \alpha_W p_W$. The assignment $\bar{f}_U((x_n)) = p$ thus provides a map

$$\bar{f}_U \colon \mathcal{C}U \cup \partial\mathcal{C}U \to \prod_{W \in \mathcal{C}\pi(f)(U)} \mathcal{C}W \cup \star_{W \in \mathcal{C}\pi(f)(U)} \mathcal{C}W$$

extending the map $\rho(f,U)$.

For any $U \in \mathfrak{S}$, the map $\bar{f}_U$ defined above is injective since the composition of f with any of the canonical projections $\prod_{W \in \pi(f)(U)} \mathcal{C}W \to \mathcal{C}W$ is a uniform quasiisometric embedding, and quasiisometric embeddings coarsely preserve Gromov products.

Definition of $\bar{f}$ Let $p \in \partial\mathcal{X}$, so that $p = \sum_{U \in \mathrm{Supp}(p)} \beta_U p_U$, where $p_U \in \partial\mathcal{C}U$ for each U, each $\beta_U \in (0,1]$, and $\sum_U \beta_U = 1$. For each $U \in \mathrm{Supp}(p)$, we defined $\bar{f}_U(p_U) = \sum_{W \in \pi(f)(U)} \alpha_W^U q_W$ above, where $q_W \in \partial\mathcal{C}W$ and $\sum_W \alpha_W^U = 1$. Let

$$\bar{f}(p) = \sum_{U \in \mathrm{Supp}(p)} \sum_{W \in \pi(f)(U)} \beta_U \alpha_W^U \cdot q_W,$$

which is a point in $\partial\mathcal{X}'$ since $\sum_U \sum_W \beta_U \alpha_W^U = 1$ and since $\bigcup_{U \in \mathrm{Supp}(p)} \pi(f)(U)$ is a pairwise-orthogonal set by Definition 5.1 since f is a slanted hieromorphism.

Injectivity of $\bar{f}|_{\partial\mathcal{X}}$ Injectivity of $\bar{f}|_{\partial\mathcal{X}}$ follows from injectivity of $\bar{f}_U$ on each $\partial\mathcal{C}U$ for $U \in \mathfrak{S}$ together with injectivity of $\pi(f)$ and the fact that each $\bar{f}_U \colon \mathcal{C}U \to \prod_{W \in \pi(f)(U)} \mathcal{C}W$ is "fully supported" in the sense that each $\alpha_W^U > 0$.

Continuity at boundary points First consider $p \in \partial\mathcal{X}$. By Proposition 2.17, there exists (x_n) in $\mathcal{X}$ such that $x_n \to p$ as $n \to \infty$. We check that $f(x_n)$ converges to $\bar{f}(p)$.

Fix a basepoint $x \in \mathcal{X}$, so that $p = \sum_{U \in \mathrm{Supp}(p)} a_U p_U$ with $\sum_U a_U = 1$, each $a_U > 0$, and for all $U, U' \in \mathrm{Supp}(p)$,

$$\left| \frac{\mathsf{d}_U(x, x_n)}{\mathsf{d}_{U'}(x, x_n)} - \frac{a_U}{a_{U'}} \right| \to 0 \quad \text{and} \quad \frac{\mathsf{d}_V(x, x_n)}{\mathsf{d}_U(x, x_n)} \to 0$$

whenever $U \in \mathrm{Supp}(p)$ and $V \in \mathrm{Supp}(p)^\perp$, and finally $\pi_U(x_n) \to p_U$ for all $U \in \mathrm{Supp}(p)$.

Consider the sequence $(w_n) = (f(x_n))$. For each $U \in \mathrm{Supp}(p)$ and $W \in \pi(f)(U)$, let $c_W\colon \prod_{V \in \pi(f)(U)} \overline{\mathcal{C}V} \to \overline{\mathcal{C}W}$ be the canonical projection. By hypothesis, for each such W we have $|\mathsf{d}_W(f(x), w_n) - \lambda_W \mathsf{d}_U(x, x_n)| \le \lambda'_W$, where $\lambda_W \in [\lambda_1, \lambda_2]$ and $\lambda'_W \ge 0$. Hence for each $U \in \mathrm{Supp}(p)$ and $W \in \pi(f)(U)$, we have that $\pi_W(w_n) = c_W \circ \bar{f}(\pi_U(x_n)) \to c_W \circ \bar{f}(p_U)$ and $\bar{f}(\pi_U(x_n)) \to \sum_{W \in \pi(f)(U)} \beta_U \alpha_W c_W \cdot \bar{f}(p_U)$ as required. Moreover, if $V \in \mathfrak{S}'$ does not belong to $\pi(f)$, then $\mathsf{d}_V(f(x), w_n)$ is uniformly bounded by Definition 5.5(2).

Finally, if $V \in \mathfrak{S} - \mathrm{Supp}(p)$, then $\mathsf{d}_V(x, x_n)$ is dominated by $\mathsf{d}_U(x, x_n)$ for any $U \in \mathrm{Supp}(p)$. Hence, for such V, we have that $\mathsf{d}_W(f(x), f(x_n))$ is dominated by $\mathsf{d}_Z(f(x), f(x_n))$ whenever $W \in \pi(f)(V)$ and $Z \in \pi(f)(U)$ for some $U \in \mathrm{Supp}(p)$, since each $\rho(f, U)$ is a uniform quasiisometric embedding. Thus $f(x_n)$ converges to $\bar{f}(p)$.

More generally, given any sequence (z_k) in $\overline{\mathcal{X}}$ converging to $p \in \partial X$, we can use the ideas in the proof of Theorem 3.4 to build a sequence of internal sequences $(x_{k,i})$ such that $\lim_i x_{k,i} = z_k$ for each k. Namely, for each k, we can take a sequence $(x_{k,i}) \to z_k$ (if $z_k \in \mathcal{X}$, then we choose $x_{k,i} = z_k$ to be constant), and then we choose $N_k > 0$ large enough that if $n > N_k$, then the sequence $(x_{k,n})$ will satisfy conditions (1)–(7) from the proof of Theorem 3.4. This will force that $\lim_i x_{k,i} = z_k$, and then since $\lim_k z_k = p$, the above conditions will force $\lim_k x_{k,n} = p$.

Now, since $\lim_n x_{k,n} = p$ and $\lim_i x_{k,i} = z_k$, the internal case above implies

$$\lim_n \bar{f}(x_{k,n}) = \bar{f}(p) \quad \text{and} \quad \lim_i \bar{f}(x_{k,i}) = z_k.$$

Together, these imply that $\lim_k \bar{f}(z_k) = \bar{f}(p)$. Thus $\bar{f}$ is continuous at boundary points.

When $\mathcal{X}$ is proper Assertion (3) combines with Theorem 3.4 and Proposition 2.17(1) to imply that $\bar{f}$ is an embedding; compactness of $\partial \mathcal{X}$ implies that its image is closed. If in addition, f is an embedding, then $\bar{f}\colon \overline{\mathcal{X}} \to \overline{\mathcal{X}}'$ is an embedding, since assertion (3) again combines with Proposition 2.17(1) and Theorem 3.4 to imply that $\bar{f}$ is a continuous injection from a compact space to a Hausdorff space. ◻

Remark 5.7 Theorem 5.6 holds under slightly more general conditions: condition (3) of Definition 5.5 need only be imposed on $U \in \mathfrak{S}$ in cases where either there exists $V \in \mathfrak{S}$ with $U \perp V$ or $|\pi(f)(U)| > 1$ or both. For any U with empty orthogonal complement and for which $\pi(f)(U) = \{V\}$ for some $V \in \mathfrak{S}'$, it suffices to require that $\rho(f,U)\colon \mathcal{C}U \to \mathcal{C}V$ is a uniform quasiisometric embedding.

5.1 Limit sets of hierarchically quasiconvex sets

Let $(\mathcal{X},\mathfrak{S})$ be a proper hierarchically hyperbolic space and let $\mathcal{Y} \subseteq \mathcal{X}$ be hierarchically quasiconvex. Let $\Lambda\mathcal{Y}$ be the set of boundary points $p = \sum_{U \in \mathrm{Supp}(p)} a_U p_U \in \partial\mathcal{X}$ such that for all $U \in \mathrm{Supp}(p)$, there is a sequence $p_U^n \in \pi_U(\mathcal{Y})$ converging to p_U.

Proposition 5.8 (hierarchically quasiconvex subspaces have limit sets) *The set $\mathcal{Y} \cup \Lambda\mathcal{Y}$ is a closed subset of $\overline{\mathcal{X}}$, and $\mathcal{Y}$ is dense in $\mathcal{Y} \cup \Lambda\mathcal{Y}$. Hence $\mathcal{Y}$ has an HHS structure such that $\mathcal{Y} \cup \Lambda\mathcal{Y} = \overline{\mathcal{Y}}$.*

Proof This is a definition chase and an application of Proposition 2.17. □

Remark 5.9 When $\pi_U|_{\mathcal{Y}}$ is either surjective or uniformly bounded for each U, Theorem 5.6, together with the HHS structure on $\mathcal{Y}$ inherited from $\mathcal{X}$, implies that $\Lambda\mathcal{Y}$ is homeomorphic to the HHS boundary $\partial\mathcal{Y}$. This holds in particular for the main examples of hierarchically quasiconvex subspaces that we use, namely product regions:

Remark 5.10 (boundaries of standard product regions) Let $U \in \mathfrak{S}$, and recall from Section 1.3 that there is a quasiisometric embedding $F_U \times E_U \to \mathcal{X}$ coming from the standard hieromorphisms. By definition, ∂F_U consists of exactly those $\sum_V a_V p_V \in \partial\mathcal{X}$ where the support set $\{V\}$ consists entirely of elements of $\mathfrak{S}_U$, while ∂E_U consists of linear combinations of the same form, but with each $V \in \mathfrak{S}_U^{\perp}$. In particular, under the map $F_U \times E_U \to \mathcal{X}$, we see that the images of

$$\partial(F_U \times \{e_1\}),\ \partial(F_U \times \{e_2\}) \to \partial\mathcal{X}$$

are identical. Moreover, the subspace $\partial F_U \subset \partial\mathcal{X}$ is closed. Finally, $\partial P_U \subset \mathcal{X}$ is a closed subset homeomorphic to $\partial F_U \star \partial E_U$, where $\star$ denotes the spherical join.

5.2 Geometrically finite subgroups of mapping class groups

In this subsection, we will show that certain interesting subgroups of mapping class groups have a well-defined limit set in the boundary. Before doing so, we give a quick sketch of relevant facts about mapping class groups and Teichmüller spaces. For more details about the HHG structure of the mapping class group, the reader is referred to [6, Section 11].

Fix a finite-type surface S. The mapping class group $\mathcal{MCG}(S)$ of S acts properly and cocompactly on the marking graph $\mathcal{M}(S)$ of S [60]. The vertices of the marking graph, called markings, are isotopy classes of certain collections of curves on S (pants decomposition together with certain transverse curves). $\mathcal{MCG}(S)$ and $\mathcal{M}(S)$ are quasiisometric via the orbit map, and we will identify $\mathcal{MCG}(S)$ with an orbit in $\mathcal{M}(S)$ from now on. The mapping class group can be given a hierarchically hyperbolic structure by considering the collection $\mathfrak{S}$ of all its (isotopy classes of essential) subsurfaces and associating to each $Y \in \mathfrak{S}$ its curve graph $\mathcal{C}Y$, a graph whose vertices are isotopy classes of essential simple closed curves on Y, except when Y is an annulus (a case that will be more subtle to deal with later, and which we will hence explain in more detail here). When Y is an annulus, $\mathcal{C}Y$ has vertices the isotopy classes of arcs connecting the two boundary components, and two such vertices are adjacent if they can be represented by disjoint arcs. The maps $\pi_Y\colon \mathcal{MCG}(S) \to 2^{\mathcal{C}Y}$ are called subsurface projections and, when Y is not an annulus, they are defined more or less by intersecting the curves in the marking with Y. When Y is an annulus π_Y is defined in the following way. Let $\widehat{Y}$ be the annular cover of S where the core of the annulus lifts to a simple closed curve. There is a natural compactification $\overline{Y}$ of $\widehat{Y}$ which is a closed annulus, and that can be identified with Y. Given a marking m, lift to $\widehat{Y}$ all the curves in m, except possibly the (only) one which is isotopic to the core of Y. Each such lift can be compactified to an arc in $\overline{Y}$, and we can finally define $\pi_Y(m)$ to be the collection of all such arcs that connect distinct boundary components of $\overline{Y}$.

We now comment briefly on Teichmüller space $\mathcal{T}(S)$ endowed with the Teichmüller metric. A point on Teichmüller space corresponds to a hyperbolic metric on S, and we can hence consider the systole map $\mathrm{Sys}\colon \mathcal{T}(S) \to 2^{\mathcal{C}S}$ that maps points in Teichmüller space to the shortest curves in the corresponding hyperbolic metric. The set of systoles is nonempty and pairwise disjoint, thus giving a bounded subset of $\mathcal{C}S$.

5.2.1 Subsurface mapping class groups For any nonpants subsurface $Y \subset S$ there is a natural embedding $\iota_Y\colon \mathcal{MCG}(Y) \hookrightarrow \mathcal{MCG}(S)$ which takes any mapping class $f_Y \in \mathcal{MCG}(Y)$ to a mapping class $f \in \mathcal{MCG}(S)$ such that $f|_Y \equiv Y$ and $f|_{S\setminus Y} \equiv \mathrm{id}_{S\setminus Y}$; if Y is an annulus, we take $\mathcal{MCG}(Y)$ to be the cyclic subgroup generated by the Dehn (half) twist about the core of Y.

We can also see this map in terms of markings: For each component $X \subset S \setminus Y$ (including annuli with core curves in ∂Y), fix a marking $\mu_X \in \mathcal{M}(X)$; if X is an annulus, then $\mu_X \in \mathcal{C}X$. Define a map $\iota_Y\colon \mathcal{M}(Y) \to \mathcal{M}(S)$ by

$$\iota_Y(\mu_Y) = \mu_Y \sqcup \bigsqcup_{\alpha\in\partial Y} \alpha \sqcup \bigsqcup_{X\in S\setminus Y} \mu_X$$

for any marking $\mu_Y \in \mathcal{M}(Y)$.

The map ι_Y extends to a hieromorphism in the obvious way and it follows from the distance formula that it is a quasiisometric embedding. Since $\mathrm{diam}_Z\big(\iota_Y(\mathcal{M}(Y))\big)$ is uniformly bounded for each $Z \in \mathfrak{S} \setminus \mathfrak{S}_Y$ and ι_Y is surjective for each $W \in \mathfrak{S}_Y$, it is, moreover, easy to see that $\iota_Y(\mathcal{M}(Y))$ is a hierarchically quasiconvex subspace of $\mathcal{M}(S)$. Hence we have, by Proposition 5.8:

Theorem 5.11 *The natural inclusion $\iota_Y\colon \mathcal{MCG}(Y) \hookrightarrow \mathcal{MCG}(S)$ equivariantly extends to a continuous embedding $\partial\iota_Y\colon \partial\mathcal{MCG}(Y) \hookrightarrow \partial\mathcal{MCG}(S)$ for any nonpants subsurface $Y \subset S$.*

5.3 Convex cocompactness subgroups

Convex cocompact subgroups of mapping class groups are a much-studied class of hyperbolic subgroups of mapping class groups, mainly because they are precisely the class of subgroups of $\mathcal{MCG}(S)$ whose corresponding surface subgroup extensions are hyperbolic. Importantly, they satisfy several strong equivalent characterizations, which we state in the following theorem-definition with parts due variously to Farb and Mosher [31], Hamenstädt [40], Kent and Leininger [48], and the first author with Taylor [27]:

Theorem 5.12 *A subgroup $H < \mathcal{MCG}(S)$ is convex cocompact if it satisfies any of the following equivalent conditions:*

1. *Any orbit of H in $\mathcal{T}(S)$ is quasiconvex.*
2. *Any orbit of H in $\mathcal{C}S$ is quasiisometrically embedded.*
3. *Any orbit of H in $\mathcal{M}(S)$ is quasiisometrically embedded and has uniformly bounded subsurface projections.*
4. *H is a stable subgroup of $\mathcal{MCG}(S)$.*
5. *The corresponding extension Γ_H of $\pi_1(S)$ is Gromov-hyperbolic.*

The following is a corollary of Proposition 5.8 and Theorems 4.3 and 5.12:

Corollary 5.13 *If $H < \mathcal{MCG}(S)$ is a convex cocompact subgroup of $\mathcal{MCG}(S)$, then the inclusion map $H \hookrightarrow \mathcal{MCG}(S)$ H–equivariantly extends to a continuous embedding $\partial_{\mathrm{Gr}} H \hookrightarrow \partial\mathcal{MCG}(S)$.*

Proof It follows immediately from properties (2) and (3) of Theorem 5.12 that H is a hierarchically quasiconvex subgroup of $\mathcal{MCG}(S)$. Since H is hyperbolic, Theorem 4.3 implies that the boundary of the induced HHS structure on H inside of $\mathcal{MCG}(S)$ is homeomorphic to $\partial_{\mathrm{Gr}} H$. The result then follows from Proposition 5.8. □

In the rest of the section, we will consider finitely generated Veech subgroups and the Leininger–Reid combination subgroups of $\mathcal{MCG}(S)$, which are generally not hierarchically quasiconvex. Recall that for both classes of groups, their actions on $\mathcal{T}(S)$ do not extend continuously everywhere to embeddings of their boundaries into $\mathbb{P}\mathcal{ML}(S)$. The main goal of the remainder of this section is to prove that such an extension does exist for both classes of groups into $\partial\mathcal{MCG}(S)$.

5.3.1 Veech subgroups The construction of Veech and Leininger–Reid subgroups involves holomorphic quadratic differentials. We will not work with them directly, so we do not need to define them, but we will rather work with the q–metric associated to a holomorphic quadratic differential q on the surface S. This is a singular flat metric on S which is locally isometric to $\mathbb{R}^2$ except at finitely many points called *singularities*.

Given a holomorphic quadratic differential q on S, there exists a convex subset $\mathrm{TD}(q) \subset \mathcal{T}(S)$ with $\mathrm{TD}(q) \cong \mathbb{H}^2$ called a *Teichmüller disk*. Let $\mathrm{Aff}^+(q)$ denote the affine group of q. Following [52], we call any subgroup $G(q) \leq \mathrm{Aff}^+(q) \leq \mathcal{MCG}(S)$, with $G(q)$ acting properly on $\mathrm{TD}(q)$, a *Veech subgroup*, except that we will also ask that $G(q)$ be finitely generated. Veech subgroups have the property that every element of $G(q)$ is either pseudo-Anosov or a multitwist about some annular decomposition A of q [73], where this annular decomposition comes from a finite measured foliation with only closed leaves naturally associated to q.

Consider the Veech subgroup $G = G(q) \leq \mathcal{MCG}(S)$. Let $\mathcal{X}_G$ be the orbit of G of a fixed marking μ in the marking graph $\mathcal{M}(S)$. Given a multitwist $g \in G$ with annular decomposition $A_g = \{\alpha_1, \dots, \alpha_{n_g}\}$, let

$$\pi_g\colon \mathcal{X}_G \to \prod_{1 \leq i \leq n_g} \mathcal{C}\alpha_i$$

be given by $\pi_g(\nu) = (\pi_{\alpha_1}(\nu), \dots, \pi_{\alpha_{n_g}}(\nu))$ for $\nu \in \mathcal{X}_G$. If $g = T_{\alpha_1}^{k_1} \cdots T_{\alpha_{n_g}}^{k_{n_g}}$, let

$$L_g = \langle g \rangle \cdot \pi_g(\mu) \subset \prod_{1 \leq i \leq n_g} \mathcal{C}\alpha_i .$$

Note that $L_g \cong \mathbb{R}$, and in fact L_g is the projection of the g–orbit of μ and thus coarsely the line in $\mathbb{R}^{n_g}$ with slope $(k_1, \dots, k_{n_g})$, where we identify the origin of $\mathbb{R}^{n_g}$ with the projection of μ. For each L_g, let $\pi_{L_g}\colon \prod_{1 \leq i \leq n_g} \mathcal{C}\alpha_i \to L_g$ be the standard projection onto L_g, considered as a subspace of $\mathbb{R}^{n_g}$ identified as above.

We now define an HHS structure $(G, \mathfrak{S}_G)$ on G as follows:

Domains S is the unique nest-maximal domain in $\mathfrak{S}_G$, and for every primitive multitwist $g \in G$ with corresponding annular decomposition $A_g = \{\alpha_{g,1}, \dots, \alpha_{g,n_g}\}$, we include a domain $U_g \in \mathfrak{S}_G$.

The spaces To S, we associate $\pi_S(G \cdot \mu) \subset \mathcal{C}S$ and to each U_g, we set $\mathcal{C}U_g = L_g$ and declare $U_g \sqsubseteq S$ for each g; moreover, we specify that $U_g \pitchfork U_{g'}$ for each primitive $g \neq g'$.

Projections $\pi_S \colon \mathcal{X}_G \to \mathcal{C}S$ is the standard projection; for each U_g, we define $\pi_{U_g} \colon \mathcal{X}_G \to L_g$ by $\pi_{U_g}(\nu) = \pi_{L_g}(\pi_g(\nu))$ for each $\nu \in \mathcal{X}_G$.

Relative projections Given $U, V \in \mathfrak{S}_G$, we define $\rho^U_V \colon \mathcal{C}U \to \mathcal{C}V$ by:

($U \sqsubseteq V$) In this case, $V = S$ and $U = U_g$ for some primitive g; then $\rho^V_U = \pi_{L_g} \circ \pi_g$.

($U \pitchfork V$) If $U = U_g$ and $V = U_{g'}$, then

$$\rho^{U_g}_{U_{g'}} = \pi_{U_{g'}}(\langle g \rangle \cdot \mu).$$

Lemma 5.14 *If G is finitely generated, then $(G, \mathfrak{S}_G)$ is an HHS structure on G, and $G < \mathrm{Aut}(\mathfrak{S}_G)$.*

Proof We need to prove that $(G, \mathfrak{S}_G)$ satisfies the axioms; since it clearly satisfies projections, nesting, orthogonality, and finite complexity, it suffices to prove it satisfies the consistency, large link, bounded geodesic image, partial realization, and uniqueness axioms. Hyperbolicity of the associated spaces uses Lemma 5.15 (the only part for which we need finite generation of G).

There is no nontrivial orthogonality, so partial realization holds by construction. Bounded geodesic image holds by the bounded geodesic image axiom in $(\mathcal{MCG}(S), \mathfrak{S})$ and the definition of $\rho^S_{U_g}$. The consistency and large link axioms hold for a similar reason. Uniqueness follows from uniqueness in $(\mathcal{MCG}(S), \mathfrak{S})$ together with Lemma 5.16. □

Lemma 5.15 *The projection $\pi_S(G \cdot \mu)$ is quasiconvex in $\mathcal{C}S$.*

Proof Consider the action of G on the corresponding Teichmüller disk $\mathrm{TD}(q)$. Since the action is proper, this makes G a finitely generated Fuchsian group. Hence, G is geometrically finite [57], so that it acts with cofinite volume on a convex subspace $C_G \subseteq \mathrm{TD}(q)$. Consider now the image of C_G and $\mathrm{TD}(q)$ in $\mathcal{C}S$. Since geodesics in $\mathcal{T}(S)$ map to quasigeodesics in $\mathcal{C}S$ [59] and C_G is a convex subspace of $\mathcal{T}(S)$, it follows that $\pi_S(C_G)$ is quasiconvex in $\mathcal{C}S$.

Now, it is not hard to see that $\pi_S(C_G)$ coarsely coincides with $\pi_S(G\cdot\mu)$. In fact, C_G contains a G–equivariant collection of horodisks such that the action on the complement C'_G is cocompact, and cocompactness implies that $\pi_S(G\cdot\mu)$ coarsely coincides with the image in $\mathcal{C}S$ of C'_G. Moreover, each horodisk is stabilized by a multitwist, and the corresponding curves are short in all hyperbolic metrics corresponding to points in the horodisk. This implies that the whole horodisk maps to a uniformly bounded subset of $\mathcal{C}S$ under the systole map, namely a neighborhood of the aforementioned curves. To sum up, the projection of the Teichmüller disk to $\mathcal{C}S$ is quasiconvex and coarsely coincides with the projection of C'_G, which in turn coarsely coincides with the projection of $G\cdot\mu$, and we are done. □

Lemma 5.16 *There exists $V>0$ such that for any $U\in\mathfrak{S}-\{S\}$, either*

$$\operatorname{diam}_U(\pi_U(G\cdot\mu))\le V$$

or $U=\alpha_i\in A_g$ for some annular decomposition A_g. In the latter case, π_U is (uniformly) coarsely surjective.

Proof Let $U\subsetneq S$ be a subsurface and let $\Delta\subset U$ be its spine, which is obtained by pulling tight ∂U with respect to the q–metric, so that vertices of Δ are singular points and edges are saddle connections (ie geodesics connecting singularities and intersecting the singular set only at the endpoints). There exists a natural retraction $r\colon U\to\Delta$ and for each edge e of Δ, let $\delta_e=r^{-1}(m_e)$, where m_e is the midpoint of e. Each δ_e is either a curve or an arc in $(U,\partial U)$. We now divide into three cases.

***U* is nonannular** In this case, Δ has a degree-3 vertex v. Suppose that μ has a base curve α that traverses each saddle connection in Δ at most once. Then v has some incident edge e such that δ_e is disjoint from α. Now, for any $g\in\operatorname{Aff}^+(q)$, we have that $g\cdot\Delta$ is the spine of $g\cdot U$, with vertices that are singular points and edges saddle connections. In particular, $g\cdot\alpha$ is a curve using each saddle connection of Δ at most once, so $\mathsf{d}_{\mathcal{AC}U}(\alpha,g\cdot\alpha)\le 3$, where $\mathcal{AC}U$ denotes the arc-and-curve graph of U. Since there is a 2–Lipschitz retraction $\mathcal{AC}(U)\to\mathcal{C}U$ [60, Lemma 2.2], it follows that $\operatorname{diam}_U(G\cdot\mu)$ is uniformly bounded.

Since $G(q)$ preserves the set of all singularities, saddle connections, and geodesic representatives of curves, we are done provided we choose the marking μ in such a way that each of its base curves traverses each saddle connection at most once.

U* ∈ *A

$U\in A_g$* for some *g Let $g\in G(q)$ be a multitwist about curves $\alpha_1,\dots,\alpha_n$, with $g=\prod_{i=1}^n T_{\alpha_i}^{k_i}$, where $k_i\in\mathbb{Z}-\{0\}$. Hence π_U is k_i–surjective (where $U=\alpha_i$). Indeed, $\pi_U(g\cdot\mu)=\pi_U(T_{\alpha_i}^{k_i}\cdot\mu)$, and the k_i are uniformly bounded since the action of $G(q)$ on the corresponding Teichmüller disc is geometrically finite, and thus there are finitely many conjugacy classes of multitwists in $G(q)$; see the proof of Lemma 5.15.

U an annulus and $U \notin A_g$ for any g The spine Δ of U contains at least one singularity, and the angle at the singularity is greater than π on both sides. Let $\widehat{U}$ be the annular cover of S corresponding to U. The lift $\widehat{\Delta}$ of Δ disconnects $\widehat{U}$ into two connected components, and we will refer to the closure of each such connected component as a *side* of $\widehat{\Delta}$. Consider a singularity along $\widehat{\Delta}$ and a saddle connection entering the singularity. Then, for any side of $\widehat{\Delta}$ there exists a unique geodesic ray emanating from the given singularity, forming an angle of π with the given saddle connection and contained in the given side of $\widehat{\Delta}$. We let $\{\alpha_i\}$ be the open arcs in $\widehat{U}$ that can be formed by concatenating two such rays lying in opposite sides of $\widehat{\Delta}$. It is readily seen that any two α_i have intersection number at most 1. The bound on the diameter of the projection onto $\mathcal{C}U$ now follows from the fact that any arc in the subsurface projection onto $\mathcal{C}U$ of some curve in S can be represented either by a geodesic transverse to a saddle connection in $\widehat{\Delta}$, which is easily seen to be disjoint from some α_i, or a geodesic containing one of the singularities, which is easily seen to intersect an appropriate α_i containing that singularity at most once. □

Lemma 5.17 *There exists a G–equivariant extensible slanted hieromorphism*

$$(G, \mathfrak{S}_G) \to (\mathcal{MCG}(S), \mathfrak{S}).$$

Proof At the level of spaces, the map $G \to \mathcal{MCG}(S)$ is the inclusion. Define $\pi(f)\colon \mathfrak{S}_G \to 2^{\mathfrak{S}}$ as follows: let $\pi(f)(S) = \{S\}$, and for each primitive multitwist g, let $\pi(f)(U_g) = A_g$, where A_g is the set of pairwise-disjoint annuli corresponding to the multicurve supporting g. This is G–equivariant since $hA_g = A_{hgh^{-1}}$ for each multitwist g and each $h \in G$.

The map $\rho(f, S)\colon \mathcal{C}S \to \mathcal{C}S$ is the identity. For each primitive multitwist $g = T_{\alpha_1}^{k_1} \cdots T_{\alpha_{n_g}}^{k_{n_g}}$, the map $\rho(f, U)\colon L_g \to \prod_i \mathcal{C}\alpha_i$ was specified above. Observe that the composition of this map with any of the canonical projections to $\mathcal{C}\alpha_i$ is a coarse similarity with multiplicative constants determined by $\{k_1, \ldots, k_{n_g}\}$. These constants are uniformly bounded since there are finitely many conjugacy classes of multitwists in $G(q)$. □

Combining Lemma 5.17 and Theorem 5.6, Remark 5.7, and Theorem 4.3 yields:

Corollary 5.18 *For any Veech subgroup $G < \mathcal{MCG}(S)$, the inclusion $G \to \mathcal{MCG}(S)$ extends continuously to an equivariant embedding $\partial_{\mathrm{Gr}} G \to \partial \mathcal{MCG}(S)$ with closed image.*

Remark 5.19 Corollary 5.18 does not follow from Proposition 5.8 because the Veech subgroup G is not hierarchically quasiconvex in $\mathcal{MCG}(S)$ whenever it contains a multitwist supported on a multicurve with more than one component; indeed, in this case there are realization points in $\mathcal{MCG}(S)$ whose images in each curve graph lie in the image of G, but which are arbitrarily far from G.

5.3.2 Leininger–Reid surface subgroups We now turn to the Leininger–Reid surface subgroups constructed in [52, Theorem 6.1]. Again, we show that these are nonhierarchically quasiconvex subgroups of $\mathcal{MCG}(S)$ that nonetheless have well-defined limit sets in $\partial\mathcal{MCG}(S)$. The setup is as follows:

(1) Let $q_1,\ldots,q_n$ be holomorphic quadratic differentials, with $A_0\in\mathcal{C}S$ the core of the annular decomposition of each q_i such that each complementary component has negative Euler characteristic.
(2) Suppose $G_0=G_0(q_i)$ for all $i\le n$.
(3) Suppose $h\in\mathcal{MCG}(S)$ centralizes G_0 and is pure and pseudo-Anosov on all components of $S-A_0$.

Then, for

$$H=G(q_1)*_{G_0}h^{k_2}G(q_2)h^{-k_2}*_{G_0}\cdots*_{G_0}h^{k_n}G(q_n)h^{-k_n},$$

the map $H\to\mathcal{MCG}(S)$ is an embedding whenever

$$N=\min\bigl\{|k_i-k_j|:i,j\in\{1,\ldots,n\},\ i\neq j\bigr\}$$

(where we set $k_1=0$) is large enough. Moreover, every element of $\operatorname{im}(H\to\mathcal{MCG}(S))$ (which we denote by H) is either pseudo-Anosov or conjugate into an elliptic or parabolic subgroup of some $h^{k_i}G(q_i)h^{-k_i}$. In particular, the $G(q_i)$ can be chosen so that H fails to be hierarchically quasiconvex for the reason explained in Remark 5.19.

In the remainder of this section, we prove:

Theorem 5.20 *The inclusion $H\to\mathcal{MCG}(S)$ extends continuously to an equivariant embedding $\partial H\to\partial\mathcal{MCG}(S)$ with closed image.*

Proof This follows from Theorem 5.6, Remark 5.7, and Proposition 5.25 below. □

Our goal is now to state and prove Proposition 5.25, which says that the inclusion of H into $\mathcal{MCG}(S)$ is a slanted hieromorphism. We need control over various projections, which we achieve in the following preliminary lemmas.

Lemma 5.21 *There exists a constant Q such that $\pi_S(h^kG(q_i)h^{-k})$ is Q–quasiconvex for any i and any k.*

Proof Apply quasiconvexity of the $\pi_S(G(q_i))$ and boundedness of $\{\pi_S(1,h^k)\}_{k\in\mathbb{Z}}$. □

Denote by $\mathcal{Y}$ the set of connected components in S of the complement of the annuli in the annular decomposition of the multitwists in G_0.

Lemma 5.22 *There exists K such that for any Y transverse to some $Y_0 \in \mathcal{Y}$ we have $\mathsf{d}_Y(\rho_Y^{Y_0}, 1) \leq K$.*

Proof This is because $\rho_Y^{Y_0}$ coarsely coincides with $\pi_Y(P_{Y_0})$, and the fact that π_Y is coarsely Lipschitz (note that there are finitely many Y_0). □

Lemma 5.23 *For each $g \in G(q_i) - G_0$ for some i and each $Y \in \mathcal{Y}$, there exists $Y' \in \mathcal{Y}$ such that $g \cdot Y'$ is transverse to Y.*

Proof This is a restatement of [52, Lemma 4.1]. □

Lemma 5.24 *There exist C and M with the following property. For any $g = g_1 h^{m_1} \cdots g_k h^{m_k}$ with $g_i \in G(q_{j(i)}) - G_0$ and $|m_i| \geq M$ for each $i \leq k$, we have $\mathsf{d}_{Y_0}(1, g) \leq C$ for each $Y_0 \in \mathcal{Y}$.*

Proof Let K be as in Lemma 5.22. Proceed by induction on k, with C to be determined. If $k = 0$, there is nothing to prove.

Suppose $k \geq 1$. Fix $Y_0 \in \mathcal{Y}$ and let $Y = g_1 Y'$ with $Y' \in \mathcal{Y}$ chosen via Lemma 5.23, so that $Y' \pitchfork Y_0$. By induction, $\mathsf{d}_Y(g_1 h^{m_1}, g) = \mathsf{d}_{Y'}(1, g_2 h^{m_2} \cdots g_k h^{m_k}) \leq C$, since $hY = Y$ for any $Y \in \mathcal{Y}$ by hypothesis, so that $g_1 h^{m_1} \cdot Y' = g_1 \cdot Y' = Y$.

By Lemma 5.16, $\mathsf{d}_Y(1, g_1)$ is uniformly bounded by some V. Hence $\mathsf{d}_Y(1, g) \geq \mathsf{d}_Y(g_1, g_1 h^{m_1}) - C - V = \mathsf{d}_{Y'}(1, h^{m_1}) - C - V$. If $|m_1|$ is large enough, then this quantity is larger than $K + 10E$. Since $Y_0 \pitchfork Y$, consistency implies that we have $\mathsf{d}_{Y_0}(\rho_{Y_0}^Y, g) \leq E$. Also,

$$\mathsf{d}_{Y_0}(\rho_{Y_0}^Y, 1) \leq \mathsf{d}_{Y_0}(\rho_{Y_0}^Y, g_1) + V = \mathsf{d}_{g_1^{-1} Y_0}(\rho_{g_1^{-1} Y_0}^{Y'}, 1) + V \leq V + K,$$

hence $\mathsf{d}_{Y_0}(1, g) \leq 2E + V + K$. Thus we set $C = 2E + V + K$, which determines M. □

Proposition 5.25 *The subgroup $H \leq \mathcal{MCG}(S)$ admits a hierarchically hyperbolic space structure $(H, \mathfrak{S}_H)$ such that there is an extensible slanted hieromorphism $(H, \mathfrak{S}_H) \to (\mathcal{MCG}(S), \mathfrak{S})$ induced by the inclusion $H \hookrightarrow \mathcal{MCG}(S)$.*

Proof We follow a very similar procedure to that used for individual Veech subgroups. In particular, $\mathfrak{S}_H$ is defined exactly as $\mathfrak{S}_G$ was, except that there is now a domain U_g for each primitive multitwist in H. To verify that this yields an HHS structure, we must check that:

(1) $\pi_S(H)$ is quasiconvex.

(2) $\pi_U(H)$ is uniformly bounded unless $U \in A_g$ for some $g \in H$.

Once the properties above are proven, arguing exactly as in the proof of Lemma 5.14 and Lemma 5.17 yields the desired slanted hieromorphism and completes the proof.

We now set conventions and notations that we use throughout the proof. When some $g = g_1 \cdots g_k \in H$ with $g_i \in h^{k_{j(i)}} G(q_{j(i)}) h^{-k_{j(i)}} - G_0$ is any fixed element of H, we write $p_l = \pi_S(g_1 \cdots g_l)$ (with $p_0 = \pi_S(1)$), and let γ_l be a geodesic in $\mathcal{C}S$ from p_{l-1} to p_l, so that the concatenation of the γ_l is a path from $\pi_S(1)$ to $\pi_S(g)$. Furthermore, notice that we can write $g = h^{m_0} g'_1 h^{m_1} \cdots g'_k h^{m_k}$ for some $g'_i \in G(q_{j(i)}) - G_0$ (more specifically, $g'_i = h^{-k_{j(i)}} g_i h^{k_{j(i)}}$), and that $|m_l|$ for $l < k$ is bounded below by N (recall that this is the minimal value of $|k_i - k_j|$ for $i \neq j$). We set $h_l = h^{m_0} g'_1 h^{m_1} \cdots g'_l$.

In the following claim, we study geodesics connecting $\pi_S(1)$ to $\pi_S(g)$ for arbitrary $g \in G$. The claim easily implies that geodesics from $\pi_S(1)$ to $\pi_S(g)$ stay close to $\pi_S(H)$ for any $g \in H$ because each γ_l is contained in a coset of some $h^{k_{j(i)}} G(q_{j(i)}) h^{-k_{j(i)}}$ and such cosets are uniformly quasiconvex byLemma 5.21. Hence, the claim proves that $\pi_S(H)$ is quasiconvex, which is item (1) above.

Claim 2 There exists a constant R with the following property. For any $g \in H$, the Hausdorff distance between $\bigcup_l \gamma_l$ and $[\pi_S(1), \pi_S(g)]$ is bounded by R, where $[\pi_S(1), \pi_S(g)]$ is any geodesic in $\mathcal{C}S$ from $\pi_S(1)$ to $\pi_S(g)$. Moreover, for any $Y \in \mathcal{Y}$ we have that $\mathsf{d}_{h_l Y}(1, h_l), \mathsf{d}_{h_l Y}(g, h_l h^{m_l}) \le C$.

Proof We first show $\bigcup_l \gamma_l$ is uniformly close to $[\pi_S(1), \pi_S(g)]$.

It suffices to show that the endpoints of all γ_l lie within controlled distance of $[\pi_S(1), \pi_S(g)]$. Any such endpoint x coarsely coincides with both $\pi_S(h_l)$ and $\pi_S(h_l h^{m_l})$, for some l (since $\{\pi_S(h^m)\}_{m \in \mathbb{Z}}$ is a bounded set). Pick any $Y \in \mathcal{Y}$, and set $Z = h_l \cdot Y$. By Lemma 5.24 we have $\mathsf{d}_Z(h_l h^{m_l}, g) \le C$ and $\mathsf{d}_Z(1, h_l) \le C$. Hence, if m_l is large enough, we get $\mathsf{d}_Z(1, g) \ge \mathsf{d}_Y(1, h^{m_l}) - 2C \ge 100E$. Notice that by bounded geodesic image ρ^Z_S needs to be within $10E$ of geodesics from $\pi_S(h_l)$ and $\pi_S(h_l h^{m_l})$, which both coarsely coincide with the endpoint x we are interested in. If geodesics from $\pi_S(1)$ to $\pi_S(g)$ did not pass close to x we could then conclude that they do not pass close to ρ^Z_S, which would imply by bounded geodesic image that $\mathsf{d}_Z(1, g) \le 5E$. But this is not the case, and hence we get a bound on the distance from x to $[\pi_S(1), \pi_S(g)]$, as required.

Let us now prove that points on $[\pi_S(1), \pi_S(g)]$ are close to $\bigcup_l \gamma_l$. Suppose by contradiction that there exists $x \in [\pi_S(1), \pi_S(g)]$ with $\mathsf{d}_S\big(x, \bigcup_l \gamma_l\big) \ge 2C + 1$. Let

$x_1, x_2 \in [\pi_S(1), \pi_S(g)]$ lie on distinct sides of x (in the natural order of $[\pi_S(1), \pi_S(g)]$) with x_1 closer to $\pi_S(1)$ than x, and satisfy $\mathrm{d}_S(x_i, x) = C+1$. Then any $y \in \bigcup \gamma_l$ lies in $\mathcal{N}_C([\pi_S(1), x_1]) \cup \mathcal{N}_C([x_2, \pi_S(g)])$. However, the two neighborhoods are disjoint and the connected set $\bigcup \gamma_l$ contains points in both, a contradiction. $\triangleleft$

Let us now take $U \in \mathfrak{S} - \{S\}$ and $g \in H$ with $\mathrm{d}_U(1, g) \geq 100E$. We need to show that either U belongs to some $A_{g'}$ or $\mathrm{d}_U(1, g)$ is bounded independently of U and g.

We proved in the claim that, for any $Y \in \mathcal{Y}$, the projections of 1 and g on $h_l \cdot Y$ coarsely coincide with the projections of h_l and $h_l h^{m_l}$, respectively, and hence that $d_{h_l \cdot Y}(1, g) > 100E$ if $|m_l| \geq N$ is large enough. Since m_l can take finitely many values, we therefore get the desired bound whenever U is of the form $h_l \cdot Y$. We now assume that U is neither belongs to some $A_{g'}$ nor it is of the form $h_l \cdot Y$. Hence, for any l there exists Y such that $h_l \cdot Y \pitchfork U$ overlap, and hence are comparable in the partial order $\preceq$; see Proposition 2.8 of [6].

Another fact about $\preceq$ is that whenever $Y, Y' \in \mathcal{Y}$ and l are such that $h_l \cdot Y \pitchfork h_{l+1} \cdot Y'$, we have $h_l \cdot Y \preceq h_{l+1} \cdot Y'$, again provided $|m_l| \geq N$ is large enough. In fact,

$$\rho^{h_{l+1}Y'}_{h_l Y} = h_{l+1} \rho^{Y}_{h^{-1}_{l+1} h_l Y'}$$

coarsely coincides with $\pi_{h_l \cdot Y}(h_{l+1})$ (Lemma 5.22), which in turn coarsely coincides with $\pi_{h_l \cdot Y}(h_l h^{m_l})$ by Lemma 5.16 since $h_{l+1} = h_l h^{m_l} g'_{l+1}$. Finally, $\pi_{h_l \cdot Y}(h_l h^{m_l})$ coarsely coincides with $\pi_{h_l \cdot Y}(g)$ by what we said above.

By looking at a predecessor and a successor of U, we then see that the projections of $1, g$ onto U coarsely coincide with those of $h_l \cdot Y, h_{l+1} \cdot Y'$ for some l and Y and Y'. But these latter projections coarsely coincide with those of h_l and $h_l h^{m_l} g'_{l+1}$. The projections of h_l and $h_l h^{m_l}$ are uniformly close by boundedness of m_l, while the projections of $h_l h^{m_l}$ and $h_l h^{m_l} g'_{l+1}$ are uniformly close by Lemma 5.16. This concludes the proof. $\square$

6 Automorphisms of HHSs and their actions on the boundary

The most important special case of an extensible hieromorphism is an automorphism of $(\mathcal{X}, \mathfrak{S})$. For any automorphism $f\colon (\mathcal{X}, \mathfrak{S}) \to (\mathcal{X}, \mathfrak{S})$, each isometry $f\colon \mathcal{C}U \to \mathcal{C}(f(U))$ extends to a homeomorphism $\hat{f}\colon \partial\mathcal{C}U \to \partial\mathcal{C}(f(U))$, yielding an application of Theorem 5.6:

Corollary 6.1 (extensions of automorphisms to the boundary) *Any $f \in \mathrm{Aut}(\mathfrak{S})$ extends to a bijection $\overline{\mathcal{X}} \to \overline{\mathcal{X}}$ which restricts to a homeomorphism on $\partial\mathcal{X}$.*

Proof Let $f\colon (\mathcal{X},\mathfrak{S}) \to (\mathcal{X},\mathfrak{S})$ be an automorphism. Let $p \in \partial\mathcal{X}$, with $p = \sum_{i=1}^{n} a_{T_i}^{p} p_{T_i}$, where the T_i are pairwise orthogonal and $p_{T_i} \in \partial\mathcal{C}T_i$. Define a map $\hat{f}\colon \partial\mathcal{X} \to \partial\mathcal{X}$ by

$$\hat{f}(p) = \sum_{i=1}^{n} a_{T_i}^{p} \hat{f}(p_{T_i}),$$

where $\hat{f}\colon \partial\mathcal{C}T_i \to \partial\mathcal{C}(f(T_i))$ is induced by $f\colon \mathcal{C}T_i \to \mathcal{C}T_i$. Let $\overline{f}\colon \overline{\mathcal{X}} \to \overline{\mathcal{X}}$ be the extension of f that is $\hat{f}$ on $\partial\mathcal{X}$; extend f^{-1} similarly. Since f is an automorphism, $\overline{f}$ is clearly a bijection. Continuity of $\overline{f}$ and $\overline{f}^{-1}$ on the boundary follows from Theorem 5.6. □

When $(G,\mathfrak{S})$ is a hierarchically hyperbolic group, ∂G is defined. In general, if $\mathcal{X}$ and $\mathcal{X}'$ are hierarchically hyperbolic with respect to the same collection $\mathfrak{S}$, then there is a quasiisometry $\mathcal{X} \to \mathcal{X}'$ extending to the identity on the boundary. Indeed, the definition of $\partial\mathcal{X}$ depends only on $\mathfrak{S}$ and the attendant hyperbolic spaces.

Corollary 6.2 *Let $(G,\mathfrak{S})$ be a hierarchically hyperbolic group. Then the action of G on itself by left multiplication extends to an action of G on $\overline{G}$ by homeomorphisms.*

Section 6.1 is devoted to automorphisms, whose fixed points in $\partial\mathcal{X}$ we study in Section 6.2.

6.1 Classification of HHS automorphisms

In this subsection, we will classify HHS automorphisms by their actions on $\mathfrak{S}$. Let $g \in \mathrm{Aut}(\mathfrak{S})$ and fix a basepoint $X \in \mathcal{X}$. Set

$$\mathrm{Big}(g) = \{U \in \mathfrak{S} \mid \mathrm{diam}_{\mathcal{C}U}(\langle g\rangle \cdot X) \text{ is unbounded}\}.$$

Observe that $g \cdot U \in \mathrm{Big}(g)$ if $U \in \mathrm{Big}(g)$, since $g\colon \mathcal{C}U \to \mathcal{C}(gU)$ is an isometry.

Lemma 6.3 *There exists $M = M(\mathfrak{S}) > 0$ such that for all $g \in \mathrm{Aut}(\mathfrak{S})$ and $U \in \mathrm{Big}(g)$, we have $g^M \cdot U = U$.*

Proof Consider the orbit $\langle g\rangle \cdot U$ in $\mathfrak{S}$.

If there exists $n \geq 1$ such that $g^n \cdot U \subsetneq U$, then $g^{kn} \cdot U \subsetneq g^{(k-1)n} \cdot U \subsetneq \cdots \subsetneq g^n \cdot U \subsetneq U$ for all $k \geq 1$, so we either contradict finite complexity (if $\langle g\rangle \cdot U$ is infinite) or the fact that $\sqsubseteq$ is a partial order (if $\langle g\rangle \cdot U$ is finite). Hence $g^n \cdot U \not\sqsubseteq U$ unless $n = 0$. Similarly, $U \not\sqsubseteq g^n U$ unless $n = 0$.

Next, consider the case where $U \in \mathrm{Big}(g)$ and $g^n \cdot U \pitchfork U$ for some $n \geq 1$. Then, since $U \in \mathrm{Big}(g)$, we can choose arbitrarily large $m \in \mathbb{N}$ such that $\mathsf{d}_U(X, g^m \cdot X) >$

$T = 100E + \mathsf{d}_U(g^{-1} \cdot X, X) + f(m)$, where $f\colon \mathbb{N} \to \mathbb{N}$ is increasing. Hence $\mathsf{d}_{g^n U}(g^{m+1} \cdot X, g \cdot X) > T$, since $g\colon \mathcal{C}U \to \mathcal{C}gU$ is an isometry. The triangle inequality shows that $\mathsf{d}_{g^n U}(g^M \cdot X, X) > T - 2\mathsf{d}_{gU}(X, g^n \cdot X) = 100E + f(m)$. By considering at least two such values of m, we see that consistency is contradicted (specifically, we contradict Lemma 2.3 of [6]).

It follows that if $U \in \mathrm{Big}(g)$, then, for all $n \in \mathbb{Z}$, either $g^n \cdot U = U$ or $g^n \cdot U \perp U$. Hence $\langle g \rangle \cdot U$ is a pairwise-orthogonal collection. Hence there exists a global M, depending only on the complexity and Lemma 1.4, such that $g^M \cdot U = U$ for each $U \in \mathrm{Big}(g)$, establishing the first assertion. □

Proposition 6.4 *The automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is elliptic if and only if* $\mathrm{Big}(g) = \varnothing$.

Proof If $\langle g \rangle \cdot X$ is bounded, then $\mathrm{Big}(g) = \varnothing$ since projections are coarsely Lipschitz.

Conversely, suppose that $\mathrm{Big}(g) = \varnothing$. We will show that there exists $D = D(g)$ such that $\mathrm{diam}_V(\pi_V(\langle g \rangle \cdot X)) \le D$ for all $V \in \mathfrak{S}$. From this and the distance formula (Theorem 1.9), it follows that g is elliptic. Hence suppose that no such D exists.

We need two facts:

(a) For each $N \ge 0$, there exists $P = P(N, \mathfrak{S})$ such that for all $U \in \mathfrak{S}$ and $h \in \mathrm{Aut}(\mathfrak{S})$, either some positive power of h fixes U or $\{U, g \cdot U, \dots, g^P \cdot U\}$ contains a set of N pairwise-transverse elements. Indeed, as in the proof of Lemma 6.3, for any p, the elements of $\{U, g \cdot U, \dots, g^{p-1} \cdot U\}$ are pairwise $\sqsubseteq$–incomparable, and any pairwise-orthogonal subset has cardinality bounded by the complexity χ of $\mathfrak{S}$. Hence, if p exceeds the Ramsey number $\mathrm{Ram}(\chi + 1, N)$, we have by Ramsey's theorem that $\{U, g \cdot U, \dots, g^{p-1} \cdot U\}$ contains a set of N pairwise-transverse elements, so we can take $P = \mathrm{Ram}(\chi + 1, N) - 1$.

(b) For each $C \ge 0$ there exists $Q \in \mathbb{N}$ with the following property. Let $x, y \in \mathcal{X}$ and suppose $\{V_i\}_{i \in I}$ satisfies $\mathsf{d}_{V_i}(x, y) > E$ for all i, and that $|I| \ge Q$. Then there exists $V \in \mathfrak{S}$ such that $V_i \sqsubsetneq V$ for some $i \in I$, and $\mathsf{d}_V(x, y) > C$. This is a slight strengthening of Lemma 3.2; this exact statement is [7, Lemma 1.8].

Recall that χ denotes the complexity — ie the maximum level — in $\mathfrak{S}$, so that S is the unique element of level χ. Since $\mathrm{Big}(g) = \varnothing$ but there are arbitrarily large projections, by assumption, there exists a level $\ell < \chi$ and a constant $R < \infty$ such that:

- $\mathrm{diam}_U(\pi_U(\langle g \rangle \cdot X)) \le R$ when U has level greater than ℓ.
- For each $D < \infty$, there exists $U \in \mathfrak{S}$, of level ℓ, with $\mathrm{diam}_U\big(\pi_U(\langle g \rangle \cdot X)\big) > D$.

Let $U \in \mathfrak{S}$ be chosen so that $\mathsf{d}_U(X, g^n \cdot U) > \ddot{R}$, where $\ddot{R}$ is a constant to be determined. We can and shall assume that our U has been chosen at level ℓ, and we emphasize that such a U can be chosen so as to make $\ddot{R}$ arbitrarily large.

Let $Q = Q(R)$ be the constant provided by setting $C = R$ in fact (b) and let $P = \mathrm{Ram}(\chi + 1, Q)$. Fact (a) provides $U_1, \dots, U_Q \in \{U, g \cdot U, \dots, g^P \cdot U\}$ such that $U_i \pitchfork U_j$ when $i \neq j$. Now, for $1 \leq j \leq Q$, we have $\mathsf{d}_{U_j}(X, g^n \cdot X) \geq \ddot{R} - 100KEQ$. So, provided $\ddot{R}$ — which can be chosen *independently* of R and hence of Q — satisfies $\ddot{R} > 100KEQ + 10E$, fact (b) provides $T \in \mathfrak{S}$ such that $U_j \sqsubsetneq T$ for some j and such that $\mathsf{d}_T(X, g^n \cdot X) > R$. Now, since U_j is a translate of U and $\mathrm{Aut}(\mathfrak{S})$ preserves the levels, the level of U_j is ℓ, and hence T has level strictly greater than ℓ, which is a contradiction since $\mathsf{d}_T(X, g^n \cdot X) > R$. □

Remark 6.5 In the case where $\mathcal{X}$ is proper, there is a quick proof of Proposition 6.4 relying on the more powerful tools from Section 9.

Lemma 6.6 *Let $g \in \mathrm{Aut}(\mathfrak{S})$. Then there exists $D = D(g, E)$ such that*

$$\mathrm{diam}_U(\pi_U(\langle g \rangle \cdot X)) \leq D$$

for all $U \in \mathfrak{S} - \mathrm{Big}(g)$.

Proof Let $\mathrm{Big}(g) = \{U_i\}_{i \in I}$. Note that it suffices to prove the lemma for some positive power of g, so by Lemma 6.3, we may assume that $g \cdot U_i = U_i$ for all $i \in I$.

If $\mathrm{Big}(g) = \emptyset$, then g is elliptic by Proposition 6.4, from which the lemma follows immediately: for each $V \in \mathfrak{S}$, we have $\mathrm{diam}_V(\pi_V(\langle g \rangle \cdot X)) \leq K \,\mathrm{diam}_{\mathcal{X}}(\langle g \rangle \cdot X)$, which is bounded independently of V.

Next, suppose that $\mathrm{Big}(g) \neq \emptyset$ and $S \notin \mathrm{Big}(g)$ (as usual, $S \in \mathfrak{S}$ is the unique $\sqsubseteq$–maximal element). Then, for each $i \in I$, the element U_i is maximal in an HHS $(F_{U_i}, \mathfrak{S}_{U_i})$ admitting a g–equivariant hieromorphism to $(\mathcal{X}, \mathfrak{S})$. Since $U_i \neq S$, the complexity of $(F_{U_i}, \mathfrak{S}_{U_i})$ is strictly lower than that of $(\mathcal{X}, \mathfrak{S})$, so it follows by induction that $\mathrm{diam}_V(\pi_V(\langle g \rangle \cdot X))$ is bounded independently of V when $V \sqsubseteq U_i$. Indeed, in the base case, when the complexity is 1, $\mathcal{X}$ is itself a hyperbolic space and the lemma follows from the usual elliptic/parabolic/loxodromic classification of isometries of hyperbolic spaces [35].

Now, let $\mathfrak{T}$ be the set of all $U \in \mathfrak{S}$ such that $U \sqsubseteq U_i$ for some $i \in I$. Observe that $\mathfrak{T}$ is g–invariant and downward-closed under nesting. Then Proposition 2.4 of [7] provides an HHS $(\widehat{\mathcal{X}}_{\mathfrak{T}}, \mathfrak{S} - \mathfrak{T})$ with the same associated nesting and orthogonality relations, hyperbolic spaces, and projections. Since $\mathfrak{T}$ was g–invariant, g descends to an automorphism of $(\widehat{\mathcal{X}}_{\mathfrak{T}}, \mathfrak{S} - \mathfrak{T})$ such that the action of g on $\mathfrak{S} - \mathfrak{T}$ is the restriction of the original action on $\mathfrak{S}$ and, for each $V \in \mathfrak{S} - \mathfrak{T}$, the isometry $\mathcal{C}V \to \mathcal{C}gV$ is the original one. Now g has $\mathrm{Big}(g) = \emptyset$ with respect to $(\widehat{\mathcal{X}}_{\mathfrak{T}}, \mathfrak{S} - \mathfrak{T})$ and hence we are done by the proof of Proposition 6.4.

The preceding two analyses prove the lemma except in the case where $S \in \mathrm{Big}(g)$. Hence, suppose $S \in \mathrm{Big}(g)$, so that g acts either loxodromically or parabolically on $\mathcal{C}S$. In this case, we cannot induct on complexity, so we argue directly using consistency, bounded geodesic image, and simple properties of isometries of hyperbolic spaces.

If $U \in \mathfrak{S} - \{S\}$, then $U \subsetneq S$, and $\rho^U_S \subset \mathcal{C}S$ is a well-defined subset of diameter $\leq E$.

First suppose that g acts loxodromically on $\mathcal{C}S$. Then there exists $N = N(g)$ such that $\leq N$ elements of $\pi_S(\langle g\rangle \cdot X)$ lie in the $100E$–neighborhood of ρ^U_S. Let $\{g^i \cdot X\}_{i=n}^{n'}$ be the points in $\langle g\rangle \cdot X \subset \mathcal{X}$ projecting into $\mathcal{N}^S_{100E}(\rho^U_S) \subset \mathcal{C}S$, so that $n' - n \leq N$. Then for all $i, j \in \mathbb{Z}$, consistency and bounded geodesic image imply that

$$\begin{aligned} \mathsf{d}_U(g^i \cdot X, g^j \cdot X) &\leq E + \max_{n \leq k, k \leq n'} \mathsf{d}_S(g^k \cdot X, g^{k'} \cdot X) \\ &\leq E + \max_{0 \leq k, k' \leq N} K \mathsf{d}_{\mathcal{X}}(g^k \cdot X, g^{k'} \cdot X) + K, \end{aligned}$$

which is independent of U (here K is the coarse Lipschitz constant from Definition 1.1).

Next, suppose that g acts parabolically on $\mathcal{C}S$. By definition, $\langle g\rangle \cdot X$ has a unique limit point in the Gromov boundary of $\mathcal{C}S$, so there is an increasing function $f\colon \mathbb{N} \to \mathbb{N}$ such that $(g^n \cdot \pi_S(X) | g^m \cdot \pi_S(X))_{\pi_S(X)} > f(k)$ whenever $\min\{|m|, |n|\} \geq k$. In particular, there exists k, independent of U, such that no $\mathcal{C}S$–geodesic from $\pi_S(g^n \cdot X)$ to $\pi_S(g^m \cdot X)$ passes $100E$–close to ρ^U_S provided $|m| \geq k$ and $|n| \geq k$. We now argue exactly as in the loxodromic case to bound $\mathrm{diam}_U(\pi_U(\langle g\rangle \cdot X))$ independently of U. This completes the proof. □

Lemma 6.7 *For any distinct $U, V \in \mathrm{Big}(g)$, we have $U \perp V$.*

Proof Lemma 6.3 shows that by passing to a uniformly bounded power, if necessary (which does not affect the big-set), we can assume that $gU = U$ and $gV = V$. Hence g acts as an isometry of both of the (not necessarily proper) hyperbolic spaces $\mathcal{C}U, \mathcal{C}V$. Since $U, V \in \mathrm{Big}(g)$, the isometry g cannot be elliptic on either $\mathcal{C}U$ or $\mathcal{C}V$. Hence, by eg [35, Section 8.1], g is either parabolic or loxodromic on $\mathcal{C}U$ and $\mathcal{C}V$.

If $U \subsetneq V$ or $U \pitchfork V$, then ρ^U_V is a uniformly bounded subset of $\mathcal{C}V$, and, since $g^n \cdot \rho^U_V \asymp \rho^{g^n U}_{g^n V} = \rho^U_V$ for all $n \in \mathbb{Z}$, we have that $\langle g\rangle$–orbits in $\mathcal{C}V$ are bounded, contradicting that $U \in \mathrm{Big}(g)$. □

Definition 6.8 (elliptic) An automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is *elliptic* if some (hence any) orbit of $\langle g\rangle$ in $\mathcal{X}$ is bounded.

Definition 6.9 (axial) An automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is *axial* if some (hence any) orbit of $\langle g\rangle$ in $\mathcal{X}$ is quasiisometrically embedded.

Definition 6.10 (distorted) An element $g \in \mathrm{Aut}(\mathfrak{S})$ is *distorted* if it is not elliptic or axial.

Example 6.11 (distorted automorphisms in familiar examples) Let S be a surface of finite type and α a simple closed curve. In $\mathcal{MCG}(S)$, the subgroup $\langle \tau_\alpha \rangle$ generated by the Dehn twist about α is quasiisometrically embedded [30], but in $(\mathcal{T}(S), d_T)$, the orbit of τ_α is distorted. In fact, $\mathcal{MCG}(S)$ has no distorted automorphisms, as is the case for cube complexes with factor systems, since cubical automorphisms are combinatorially semisimple [39]. In Theorem 7.1 below, we prove that HHGs have no distorted elements. A simple example of an HHS with a distorted automorphism is obtained by gluing a combinatorial horoball to $\mathbb{Z}$; this encapsulates the difference between the HHS structures of $\mathcal{MCG}(S)$ and $(\mathcal{T}(S), d_T)$, where annular curve graphs are replaced by horoballs over annular curve graphs.

Proposition 6.12 *The automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is axial if and only if there exists $U \in \mathrm{Big}(g)$ such that $n \to g^n \cdot \pi_U(X)$ is a quasiisometric embedding $\mathbb{Z} \to \mathcal{C}U$ for any $X \in \mathcal{X}$.*

Proof Suppose that there exists $U \in \mathrm{Big}(g)$ such that $n \to g^n \cdot \pi_U(X)$ is a quasiisometric embedding. Then the distance formula (Theorem 1.9) yields a lower bound on $\mathsf{d}_{\mathcal{X}}(g^m \cdot X, g^n \cdot X)$ which is (at least) linear in $|m-n|$, ie g is axial.

Conversely, suppose that g is axial. Lemma 6.7 bounds the number of $U \in \mathrm{Big}(g)$ by the complexity of $\mathfrak{S}$. Lemma 6.6 ensures that $\mathrm{diam}_V(\pi_V(\langle g \rangle \cdot X))$ is bounded independently of V for $V \notin \mathrm{Big}(g)$. Since g acts axially on $\mathcal{X}$, the distance formula (Theorem 1.9) now implies that there exists at least one $U \in \mathrm{Big}(g)$ such that g acts axially on $\mathcal{C}U$. $\square$

The next proposition is an immediate consequence of Propositions 6.4 and 6.12:

Proposition 6.13 *The automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is distorted if and only if there exists $U \in \mathrm{Big}(g)$ such that $\langle g \rangle \cdot \pi_U(X)$ is unbounded but, for all $U \in \mathrm{Big}(g)$, we have*

$$\mathsf{d}_{\mathcal{C}U}(X, g^n \cdot X) = o(n).$$

Definition 6.14 (reducible) The automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is *irreducible* if $\mathrm{Big}(g) = \{S\}$, where $S \in \mathfrak{S}$ is the unique $\sqsubseteq$–maximal element. Otherwise, $S \notin \mathrm{Big}(g)$ and g is *reducible*.

Finally, we have the following strong characterization of irreducible axials:

Theorem 6.15 *Let $G \leq \mathrm{Aut}(\mathfrak{S})$ act properly and coboundedly on the hierarchically hyperbolic space $(\mathcal{X}, \mathfrak{S})$. Suppose that $g \in G$ is irreducible axial. Then g is Morse.*

Proof By [5, Corollary 14.4], G acts acylindrically on $\mathcal{C}S$, where S is $\sqsubseteq$–maximal in $\mathfrak{S}$, while g acts hyperbolically on $\mathcal{C}S$. By [70, Proposition 3.8], g is *weakly contracting* for the *path system* consisting of all geodesics in $\mathcal{C}S$, so g is Morse, by [70, Lemma 2.9]. □

Remark 6.16 (reducible Morse elements) The converse of Theorem 6.15 does not hold, as can be seen be examining a Morse element of an appropriately chosen right-angled Artin group whose support does not include all generators.

6.2 Dynamics of action on the boundary

In the remainder of this section, we impose the standing assumption that $\mathcal{X}$ is proper. We will analyze the action of an infinite-order automorphism g on $\partial(\mathcal{X}, \mathfrak{S})$, according to whether g is irreducible or reducible and according to whether g is axial or distorted.

6.2.1 Irreducible automorphisms

Lemma 6.17 *Let the irreducible* $g \in \mathrm{Aut}(\mathfrak{S})$ *fix some* $\lambda \in \partial\mathcal{X}$*. Then* $\mathrm{Supp}(\lambda) = \{S\}$.

Proof Suppose $U \in \mathrm{Supp}(\lambda) - \{S\}$. Since g is irreducible, its orbit in $\mathcal{C}S$ is unbounded. In particular, this means that the orbit of ρ^U_S is unbounded. By definition, $g \cdot \rho^U_S \asymp \rho^{g \cdot U}_S$ and thus U could not be fixed by g, completing the proof. □

Proposition 6.18 (irreducible axials act with north–south dynamics) *If* $g \in \mathrm{Aut}(\mathfrak{S})$ *is irreducible axial, then* g *has exactly two fixed points* $\lambda_+, \lambda_- \in \partial\mathcal{X}$*. Moreover, for any boundary neighborhoods* $\lambda_+ \in U_+$ *and* $\lambda_- \in U_-$*, there exists an* $N > 0$ *such that* $g^N(\partial\mathcal{X} - U_-) \subset U_+$.

Proof Let $g \in \mathrm{Aut}(\mathfrak{S})$ be irreducible axial. For the rest of the proof, fix a basepoint $X \in \mathcal{X}$.

Existence of $\boldsymbol{\lambda_+, \lambda_- \in \partial\mathcal{X}}$ For any n, let $X_n = g^n \cdot X$. We will show that (X_n) converges to some point in $\partial\mathcal{X}$; a similar argument will show that (X_{-n}) converges to some other point, and then we will prove they are distinct. By compactness (Theorem 3.4), there exists a subsequence $(X_{n_k}) \subset (X_n)$ which converges to some point $\lambda_+ \in \partial\mathcal{X}$. By irreducibility of g, we must have that $\lambda_+ \in \partial\mathcal{C}S \subset \partial\mathcal{X}$. By irreducibility and the definition of convergence, we have that $\pi_{\mathcal{C}S}(X_{n_k}) \to \lambda_+ \in \partial\mathcal{C}S$. Axiality of g then implies that, for any other subsequence $(X_{n_l}) \subset (X_n)$, the Gromov product $(X_{n_k}, X_{n_l})_X \to \infty$ in $\mathcal{C}S$ as $k, l \to \infty$. This implies that $\pi_{\mathcal{C}S}(X_n) \to \lambda_+ \in \partial\mathcal{C}S$, which implies that $X_n \to \lambda_+ \in \partial\mathcal{X}$.

Similarly, we define $X_{-n} \to \lambda_- \in \partial\mathcal{X}$. Observe that $(\pi_{\mathcal{C}S}(X_n), \pi_{\mathcal{C}S}(X_{-n}))_{\pi_{\mathcal{C}S}(X)}$ is uniformly bounded by Proposition 6.12, implying that $\lambda_+ \neq \lambda_-$. Since g stabilizes the orbit, it obviously fixes λ_+ and λ_-. Note that λ_+ and λ_- are independent of our choice of $X \in \mathcal{X}$.

Uniqueness of $\boldsymbol{\lambda_+, \lambda_- \in \partial\mathcal{X}}$ By Lemma 6.17, any point $\lambda \in \partial\mathcal{X}$ fixed by g has $\operatorname{Supp}(\lambda) = S$. If g fixes three points in $\partial\mathcal{X}$, then it fixes three points in $\partial\mathcal{C}S$. As such, g coarsely fixes the coarse median of those points, producing a bounded orbit, a contradiction.

North–south dynamics on $\boldsymbol{\partial\mathcal{X}}$ Fix boundary neighborhoods $\lambda_+ \in U_+$ and $\lambda_- \in U_-$ with $U_+ \cap U_- = \varnothing$.

Claim 1 For any $p \in \partial\mathcal{X} - \{\lambda_-\}$, the sequence $(g^n(p))$ does not converge to λ_-.

Proof of Claim 1 If $\operatorname{Supp}(p) \neq \{S\}$, then $(g^n(p))$ cannot converge to a point in $\partial\mathcal{X}$ supported on S, as g does not alter the coefficients of the pieces of p supported on proper subdomains. In particular, since $\operatorname{Supp}(\lambda_-) = \{S\}$, as shown above, $(g^n(p))$ cannot converge to λ_-. Thus we may assume that $\operatorname{Supp}(p) = \{S\}$.

Let $[X, p]$ be a hierarchy ray in $\overline{\mathcal{X}}$. Since $\operatorname{Supp}(p) = \{S\}$, $[X, p]$ projects to a D–quasigeodesic, $[X, p]_S \subset \overline{\mathcal{C}S}$. Let $[X, \lambda_-]$ be the orbit $(g^{-n}(X))$, which is a quasigeodesic with quality depending on g.

Consider $m \in \mathcal{C}S$, the coarse median of (λ_-, p, X). By hyperbolicity, there exist points $Y \in [X, p]_S$ and $Z \in [X, \lambda_-]$ sufficiently far out along $[X, p]_S$ and $[X, \lambda_-]$ such that any geodesic $[Y, Z]$ between Y and Z comes uniformly close to m, independent of Y and Z; in particular, the coarse median of (X, Y, Z) is uniformly close to m. Moreover, there is a uniform constant $\delta' > 0$ (depending on D, g, and the hyperbolicity constant, $\delta > 0$) such that each of $[Y, Z]$, $[X, Y]$, and $[X, Z]$ is δ'–close to m.

Let $m_{Y,Z} \in [Y, Z]$ and $m_{X,Z} \in [X, Z]$ be points δ'–close to m. Then there exists a uniform $\delta'' > 0$ such that $[m_{Y,Z}, Z]$ and $[m_{X,Z}, Z]$ must δ''–fellow-travel. By axiality, there exists $N > 0$ such that, for all $n > N$, $g^n(m_{X,Z})$ is between X and $g^n(X)$ along the quasigeodesic axis of g in $\mathcal{C}S$. This implies that the coarse median of $(X, g^n(Y), g^n(Z))$ is uniformly close to X. Thus $(g^n(p), \lambda_-)_X$ is uniformly bounded and $(g^n(p))$ cannot converge to λ_- in $\partial\mathcal{C}S$ and thus not in $\partial\mathcal{X}$ as well. $\triangleleft$

Since the limit of $(g^n(p))$ is a fixed point, uniqueness of λ_-, λ_+ and Claim 1 imply that $g^n(p) \to \lambda_+$ for any $p \in \partial\mathcal{X} - \{\lambda_-\}$.

Now consider the function $f \colon \partial\mathcal{X} - U_- \to \mathbb{N}$, where $f(p)$ is the least power N_p such that $g^{N_p}(p) \in U_+$. Since λ_+ and λ_- are the unique fixed points of g, such a power exists, otherwise the sequence $(g^n(p)) \subset \partial\mathcal{X}$ would subconverge to another fixed point. Since $\partial\mathcal{X}$ is compact (Theorem 3.4) the function f attains a maximum, N_f. By definition, $g^{N_f}(\partial\mathcal{X} - U_-) \subset U_+$, completing the proof. □

We now treat the irreducible distorted case:

Proposition 6.19 (irreducible distorteds act parabolically) *If $g \in \mathrm{Aut}(\mathfrak{S})$ is irreducible distorted, then g has exactly one fixed point $\lambda_g \in \partial\mathcal{X}$, and $g^n \cdot X, g^{-n} \cdot X \to \lambda_g$ for any $X \in \overline{\mathcal{X}}$.*

Proof Let $S \in \mathfrak{S}$ be the unique $\sqsubseteq$–maximal element, so that $gS = S$ and $g\colon \mathcal{C}S \to \mathcal{C}S$ is an isometry. By the definition of irreducibility, $\mathrm{Big}(g) = \{S\}$, so g has unbounded orbits in the δ–hyperbolic space $\mathcal{C}S$. We now apply the classification of isometries of hyperbolic spaces, as summarized in [17, Section 3], emphasizing that these results do *not* rely on properness of the space in question.

First, by Proposition 3.2 of [17] and the fact that $\langle g\rangle \cdot \pi_X(X)$ (which coarsely coincides with $\pi_S(\langle g\rangle \cdot X)$) is distorted — ie not quasiconvex — in $\mathcal{C}S$, we have that the action of $\langle g\rangle$ on $\mathcal{C}S$ is *not* lineal or focal. By Lemma 3.3, the action of $\langle g\rangle$ on $\mathcal{C}S$ is not of *general type*. Hence the action is *horocyclic*, ie the limit set of $\langle g\rangle$ on $\partial\mathcal{C}S$ consists of *exactly one point* λ_g with $g\lambda_g = \lambda_g$. Moreover, Proposition 3.1 of [17] implies that every $\lambda \neq \lambda_g$ in $\partial\mathcal{C}S$ has infinite $\langle g\rangle$–orbit. We also denote by λ_g the image of this limit point under the usual ($\mathrm{Aut}(\mathfrak{S})$–equivariant) embedding $\partial\mathcal{C}S \to \partial\mathcal{X}$. We thus have a fixed point $\lambda_g \in \partial\mathcal{X}$ for g. Now, suppose that $\lambda \in \partial\mathcal{X}$ is fixed by g. By Lemma 6.17, $\lambda \in \partial\mathcal{C}S \subset \partial\mathcal{X}$. If $\lambda \neq \lambda_g$, then (as a point of $\partial\mathcal{C}S$), λ cannot be fixed by g, so λ_g is the unique fixed point in $\partial\mathcal{X}$.

Finally, if $p \in \partial\mathcal{X} - \lambda_g$, then $g^n \cdot p \to \lambda_g$, for it subconverges to some point by compactness of $\overline{\mathcal{X}}$ (Theorem 3.4), which is fixed by g and thus must be λ_g by uniqueness. □

Proposition 6.20 *Let $g \in \mathrm{Aut}(\mathfrak{S})$ be irreducible distorted and fix $\lambda_g \in \partial\mathcal{X}$. For any neighborhood $U \subset \partial\mathcal{X}$ of λ_g, there exists $N > 0$ such that if $p \in \partial\mathcal{X} - U$, then $g^N \cdot p \in U$.*

Proof Fix a neighborhood $\lambda_g \in U \subset \partial\mathcal{X}$ and let $p \in \partial\mathcal{X} - U$. Let $F\colon \overline{\mathcal{X}} \to \mathbb{N}$ be the map which takes each $p \in \overline{\mathcal{X}}$ to the minimal $n \in \mathbb{N}$ such that $g^n \cdot p \in U$; note that F is defined by Proposition 6.19. We prove that F is bounded.

Assume not; then there exists a sequence $(p_i) \subset \partial\mathcal{X}$ such that $F(p_i) = n_i \to \infty$ as $i \to \infty$. By compactness of $\overline{\mathcal{X}}$, the sequence (p_i) accumulates on some point $\mu \in \partial\mathcal{X}$. If $N_\mu = F(\mu)$, then $g^{N_\mu} \cdot \mu \in U$. Choose an open neighborhood $g^{N_\mu} \cdot \mu \in V \subset U$.

By passing to a subsequence if necessary, we may assume $p_i \to \mu$ and continuity of the action of g on $\partial\mathcal{X}$ implies that $g^{N_\mu} \cdot p_i \to g^{N_\mu} \cdot \mu$. In particular, this implies that the sequence $(g^{N_\mu} \cdot p_i)$ eventually lies in $V \subset U$, a contradiction. □

6.2.2 Reducible automorphisms We now turn to nonelliptic reducible automorphisms. As before, we assume $\mathcal{X}$ is proper, $g \in \mathrm{Aut}(\mathfrak{S})$ has infinite order and is thus axial or distorted, and $\mathrm{Big}(g) \neq \varnothing$ denotes the set of (pairwise orthogonal) $U \in \mathfrak{S}$ where $\mathrm{diam}_{\mathcal{C}U}(\langle g\rangle \cdot X) = \infty$.

If g is reducible, then $\mathrm{Big}(g) = \{A_i\} \sqcup \{B_j\}$, where g acts axially on $\mathcal{C}A_i$ and distortedly on $\mathcal{C}B_j$ for all i and j, and $A_i, B_j \neq S$ for all i and j. Proposition 6.12 implies that g is axial if and only if $\{A_i\} \neq \varnothing$; otherwise g is distorted.

We must be careful with nontrivial finite orbits in $\mathfrak{S}$. To that end, recall that by Lemma 6.3 there exists $M = M(\mathfrak{S}) > 0$ such that g^M fixes $\mathrm{Big}(g)$ pointwise. The proof of that lemma shows that g^M in fact fixes $\{A_i\}$ and $\{B_i\}$ pointwise, since we cannot have $g \cdot A_i = B_j$ for any i and j. Let $h = g^M$, and note that $\mathrm{Big}(h) = \mathrm{Big}(g)$. Note that we can choose M so that *any* pairwise-orthogonal subset of $\mathfrak{S}$ stabilized by h is fixed by h pointwise.

Lemma 6.21 *Let $V \in \mathfrak{S}$ and suppose that $V \sqsubseteq U$ or $V \pitchfork U$ for some $U \in \mathrm{Big}(g)$. Suppose also that $p \in \partial\mathcal{X}$ is fixed by g. Then $V \notin \mathrm{Supp}(p)$.*

Proof By hypothesis, $h \cdot p = p$. Observe that $\langle h\rangle \cdot \rho^V_U$ is unbounded. Since $U \in \mathrm{Big}(g)$, we have that $h \cdot \rho^V_U = \rho^{h\cdot V}_U$ and $h \cdot U$ is infinite, implying $U \notin \mathrm{Supp}(p)$, as required. □

We denote by $\mathbb{S}^k$ a k–sphere and by $\mathbb{D}^k$ a k–ball. Given spaces X and Y, we denote by $X \star Y$ their join. For each i and j, let $F_i = F_{A_i}$ and $F'_j = F_{B_j}$ be the standard factors associated to A_i and B_j, so that there is a quasiconvex hieromorphism $\prod_i F_i \times \prod_j F'_j \to \mathcal{X}$, inducing an embedding $\bigstar_i \partial F_i \to \bigstar_j \partial F'_j \to \partial\mathcal{X}$ whose image is a closed g–invariant subset which we denote by $\mathfrak{E}(g)$. (Note: The image of $\prod_i F_i \times \prod_j F'_j$ need not be g–invariant, but since g stabilizes each standard product region $F'_j \times E_{B_j}$, the subspaces gF_i, F_i are parallel, and thus have the same boundary.)

For each i, the action of $h = g^M$ on $P_{F_i} \cong F_i \times E_{A_i}$ induces an action of h on F_i by applying the restriction homomorphism $\theta_{A_i}\colon \mathrm{Stab}_{\mathrm{Aut}(\mathfrak{S})}(A_i) \to \mathrm{Aut}(\mathfrak{S}_{A_i})$. For each A_i, let h_i be the image of h under this homomorphism, and let h_j be the image of h under the corresponding restriction homomorphism for B_j.

The following proposition says that, up to taking a power, a reducible automorphism can be decomposed into irreducible automorphisms on subdomains:

Proposition 6.22 *If g is nonelliptic reducible and $h = g^M$, then the following hold:*

(1) *For each i, h_i is an irreducible axial automorphism of F_i which fixes a unique pair of points $\lambda_{i,+}, \lambda_{i,-} \in \partial\mathcal{C}A_i$ and acts with north–south dynamics on $\partial\mathcal{C}A_i$.*

(2) *For each j, h_j is an irreducible distorted automorphism of F'_j and fixes a unique point $\lambda_{h_j} \in \partial\mathcal{C}B_j$.*

Hence, g stabilizes (and h fixes pointwise) a nonempty subspace $S(g) \star C(g) \subseteq \partial\mathcal{X}$, where $S(g) = \varnothing$ or $S(g) \cong \mathbb{S}^{|\{A_i\}|-1}$ and $C(g) = \varnothing$ or $C(g) \cong \mathbb{D}^{|\{B_j\}|}$. Moreover, for each $n > 0$, g^n does not fix any point in $\mathfrak{E}(g) - S(g) \star C(g)$.

Proof For each i, h_i acts on $\mathcal{C}A_i$ axially by the assumption on g and irreducibly by construction. Hence, Proposition 6.18 implies that h_i fixes two points $\lambda_{i,+}, \lambda_{i,-} \in \partial\mathcal{C}A_i$ and acts with north–south dynamics on $\partial\mathcal{C}A_i$. Similarly, for each j, h_j acts on $\mathcal{C}B_j$ distortedly by assumption and irreducibly by construction. Proposition 6.19 then implies that h_j fixes a unique point $\lambda_{h_j} \in \partial\mathcal{C}B_j$.

If $\{A_i\} \neq \varnothing$, then each A_i contributes a pair of points $\lambda_{i,+}, \lambda_{i,-} \in \partial\mathcal{C}A_i$ fixed by h, which we can think of as a copy of $\mathbb{S}^0$, namely $\mathbb{S}^0_i$. Moreover, h clearly fixes the join of these spheres, $\bigstar_i \mathbb{S}^0_i \cong \mathbb{S}^{|\{A_i\}|-1} = S(g)$, as required.

Similarly, if $\{B_i\} \neq \varnothing$, then each B_j contributes a point $\lambda_{h_j} \in \partial\mathcal{C}B_j$ fixed by h, and h fixes the join of these points, $\bigstar_j \lambda_{h_j} \cong \mathbb{D}^{|\{B_j\}|} = C(g)$, as required.

Since h fixes these $S(g)$ and $C(g)$, h clearly fixes $S(g) \star C(g)$. Now, if g^n fixes a point $\lambda \in \mathfrak{E}(g)$, then $h^n = (g^n)^M$ fixes λ. If $\lambda = \sum_i a_i p_i + \sum_j b_j q_j$, where $p_i \in \partial F_i$ and $q_i \in \partial F'_j$, then the uniqueness of the $\lambda_{i,+}, \lambda_{i,-}$ and λ_{h_j} implies that, for $a_i \neq 0$ and $b_j \neq 0$, we must have $q_j = \lambda_{h_j}$ and either $p_i = \lambda_{i,+}$ or $p_i = \lambda_{i,-}$. □

Remark 6.23 Set $\mathrm{Comp}(g) = \{p \in \partial\mathcal{X} \mid \mathrm{Supp}(p) \subset \{A_i, B_j\}^{\perp}_{i,j}\}$ and let $\mathrm{Fix}(h) \subset \partial\mathcal{X}$ be the set of fixed points of h. It is not difficult to show that

$$\mathrm{Fix}(h) \subseteq S(g) \star C(g) \star \mathrm{Comp}(g),$$

but proper containment can happen.

Lemma 6.24 *Let $U \in \mathrm{Big}(g)$ and $U \sqsubseteq V$. For all $p \in \partial\mathcal{X}$ such that $g^n(p) = p$ for some $n > 0$, we have $V \notin \mathrm{Supp}(p)$.*

Proof It suffices to prove the lemma for $h = g^M$. Suppose for a contradiction that $V \in \mathrm{Supp}(p)$. Since $U \in \mathrm{Big}(h)$, $\mathrm{diam}_V(\langle h\rangle \cdot \rho^U_V)$ is uniformly bounded. Take any sequence $X_k \to p$ in $\overline{\mathcal{X}}$; note that this implies $X_k \to p_V$ in $\overline{\mathcal{C}V}$. Thus, there exists $K > 0$ such that $d_V(X_k, \rho^U_V) > 100E$ if $k \geq K$.

Since h is unbounded on $\mathcal{C}U$, there exists $N > 0$ depending only on K such that $d_U(X_k, h^n(X_k)) > 100E$ if $n \geq N$ and $k \geq K$. If γ is a hierarchy path between X_k and $h^n(X_k)$ in $\mathcal{X}$, then the bounded geodesic image axiom (Definition 1.1(7)) implies that $\pi_V(\gamma) \cap N_E(\rho_V^U) \neq \varnothing$. In particular, this implies that $d_V(X_k, h^n(X_k)) > 100E$. Thus, for any $n > N$, we have that $(X_k, h^n(X_k))_{\rho_V^U}$ is uniformly bounded as $k \to \infty$, which implies that no power of h could fix p, a contradiction. □

Proposition 6.25 *Let $p \in \partial\mathcal{X}$ be such that $g^M(p) = p$ for some $M > 0$. Then*

$$p \in S(g) \star C(g) \star \Bigl(\bigcap_i \partial E_{A_i} \cap \bigcap_j \partial E_{B_j}\Bigr).$$

Proof Lemmas 6.21 and 6.24 imply

$$\operatorname{Supp}(p) \subset \bigcup_{i,j} \bigl(\mathfrak{S}_{A_i} \cup \mathfrak{S}_{B_j} \cup (\{A_i\}^\perp \cap \{B_j\}^\perp)\bigr),$$

which, together with Proposition 6.22 and g–invariance of $\operatorname{Big}(g)$, gives the claim. □

6.3 Dynamics on boundaries of HHGs

Fix a hierarchically hyperbolic group $(G, \mathfrak{S})$.

Definition 6.26 (stable boundary points) A point $p \in \partial G$ is a *stable boundary point* if p is a fixed point of some irreducible axial element of $\operatorname{Aut}(\mathfrak{S})$.

The next lemma states that irreducible axials have cobounded orbits.

Lemma 6.27 *Let $g \in G$ be an irreducible axial. Then, given any $X \in \mathcal{X}$, there exists $N > 0$ such that $\operatorname{diam}_{\mathcal{C}U}(\langle g\rangle \cdot X) < N$ for any $U \in \mathfrak{S} - \{S\}$.*

Proof If not, then there is a sequence of domains $U_n \in \mathfrak{S}$ such that $\operatorname{diam}_{\mathcal{C}U_n}(\langle g\rangle \cdot x) \geq n$ for each n. Since g is irreducible axial, $\langle g\rangle \cdot X$ projects to a uniform quasigeodesic in $\mathcal{C}S$.

By the bounded geodesic image axiom and hyperbolicity of $\mathcal{C}S$, for each $n > 100E$, there exists a sequence $(k_n) \subset \mathbb{Z}$ such that $\rho_S^{U_n} \in \mathcal{N}_E([g^{k_n} \cdot X, g^{k_n+1} \cdot X]) \subset \mathcal{C}S$, where $[g^{k_n} \cdot X, g^{k_n+1} \cdot X]$ is any geodesic between $g^{k_n} \cdot X$ and $g^{k_n+1} \cdot X$ in $\mathcal{C}S$. Moreover, since $\langle g\rangle \cdot X$ is a uniform quasigeodesic in $\mathcal{C}S$, it follows that $d_{U_n}(g^{k_n} \cdot X, g^{k_n+1} \cdot X) \asymp \operatorname{diam}_{U_n}(\langle g\rangle \cdot X) \geq n$.

It follows that there exists a sequence of domains $U_n' = g^{-k_n} \cdot U_n \in \mathfrak{S}$ with $\rho_S^{U_n'} \in \mathcal{N}_E([X, g \cdot X])$ and $d_{U_n'}(X, g \cdot X) \asymp \operatorname{diam}_{U_n'}(\langle g \rangle \cdot X) \geq n$, which is impossible by the distance formula. This completes the proof. □

Proposition 6.28 *If G has an irreducible axial element, then the set of stable boundary points is dense in ∂G.*

Proof Let $p \in \partial G$ be any point and let $\lambda \in \partial G$ be a stable boundary point for some irreducible axial $g \in G$. Choose $X \in \mathcal{X}$ and let $\gamma_n = [X, g^n \cdot X]$ be a D–hierarchy path between X and $g^n \cdot X$. Let $\gamma = [X, \lambda]$ be the limiting D–hierarchy ray as $n \to \infty$. Since $\gamma_n \to \gamma$ uniformly on compact sets and $\langle g \rangle \cdot X$ is uniformly cobounded by Lemma 6.27, it follows that γ is uniformly cobounded.

By coboundedness of the action of G and density of the interior (Proposition 2.17), there exists a sequence $(g_n) \subset G$ and $N > 0$ such that $g_n(X) \to p$ and thus $g_n \cdot \lambda \to p$. Since G acts on itself by automorphisms, we have that $g_n \cdot [X, \lambda]$ projects to an infinite quasigeodesic in $\mathcal{C}S$, implying that $g_n \cdot \lambda \in \partial \mathcal{C}S \subset \partial G$, which completes the proof. □

Theorem 6.29 (topological transitivity of the G–action on ∂G) *Let $(G, \mathfrak{S})$ be a hierarchically hyperbolic group with G not virtually cyclic and containing an irreducible axial element. For any $p \in \partial G$, the orbit $G \cdot p$ is dense in ∂G.*

Proof Let $U \subseteq \partial G$ be an open set. By Proposition 6.28, there exists an irreducible axial $g \in G$ with stable boundary points $\lambda_{g,+}, \lambda_{g,-} \in \partial G$, one of which is contained in U. Suppose that $\lambda_{g,+} \in U$ and $\lambda_{g,-} \neq p$. Then, since ∂G is Hausdorff, it follows from Proposition 6.18 that some power of g moves p into U, as required. Hence either we are done, or for every irreducible axial g with $\lambda_{g,+} \in U$, we have $\lambda_{g,-} = p$.

Now, suppose that there exists $q \in \partial G - U \cup \{p\}$. Then, by Proposition 6.28, and the fact that ∂G is Hausdorff, we may argue as above, using Proposition 6.18, that some irreducible axial element takes p arbitrarily close to q, and thus that some power of g takes a translate of p into U, as required, unless p is a stable point for *every* irreducible axial element of G. But then G does not contain two independent irreducible axial elements whence, since G acts acylindrically on $\mathcal{C}S$ by [5, Theorem 14.3], a theorem of Osin (see Theorem 9.3 below) implies that G is virtually cyclic. □

Corollary 6.30 *If $(G, \mathfrak{S})$ is an HHG with an irreducible axial, then $\partial \mathcal{C}S$ is dense in ∂G.*

Remark 6.31 In Section 9, we investigate the question of when groups of HHS automorphisms contain irreducible axial elements. In that section, we consider a more general class, so-called "rank-one" elements, of which irreducible axial elements are the main examples.

7 Coarse semisimplicity in hierarchically hyperbolic groups

Theorem 7.1 *If $(G, \mathfrak{S})$ is a hierarchically hyperbolic group, then each $g \in G$ is either elliptic or axial, and $\pi_U(\langle g \rangle)$ is a quasiisometrically embedded copy of $\mathbb{Z}$ for each $U \in \mathrm{Big}(g)$.*

Proof of Theorem 7.1 This follows from Lemmas 7.3 and 7.4 below. □

Our main tool here is the following result of Bowditch:

Lemma 7.2 [11, Lemma 2.2] *If G acts acylindrically by isometries on a hyperbolic space M, then each element of G acts either elliptically or loxodromically on M.*

Lemma 7.2 and [5, Theorem 14.3] combine to yield:

Lemma 7.3 *If $g \in G$ is irreducible, then g is either elliptic or axial.*

Recall that for any reducible $g \in G$, we have $\mathrm{Big}(g) = \{A_i\} \cup \{B_j\}$, where g acts axially on each $\mathcal{C}A_i$ and distortedly on each $\mathcal{C}B_j$. It remains to prove:

Lemma 7.4 *If $g \in G$ is reducible, then $\{B_j\} = \varnothing$.*

For each $U \in \mathfrak{S}$, let $G_U = \mathcal{A}_U \cap G$ be the subgroup of G fixing $U \in \mathfrak{S}$ and let $\bar{G}_U = \theta_U(G_U)$, where $\mathcal{A}_U = \mathrm{Stab}_{\mathrm{Aut}(\mathfrak{S})}(U)$ and $\theta_U \colon \mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U)$ is the restriction homomorphism.

Lemma 7.5 *Let $U \in \mathfrak{S}$. Then $\bar{G}_U$ acts acylindrically on $\mathcal{C}U$.*

Proof of Lemma 7.5 By definition, $\bar{G}_U$ acts by automorphisms on the hierarchically hyperbolic space $(F_U, \mathfrak{S}_U)$. We first establish:

Claim 1 For each $R \geq 0$, there exists $K = K(R)$ such that any R–ball $B \subseteq F_U$ intersects gB for at most K elements $g \in \bar{G}_U$.

Proof of Claim 1 Since the inclusion hieromorphism $(F_U, \mathfrak{S}_U) \to (G, \mathfrak{S})$ is a quasiisometric embedding (with constants independent of U), it suffices to bound the number of cosets $g(\ker \theta_U)$ in G_U for which $g(\ker \theta_U) \cdot (B' \times E_U) = (\bar{g}B') \times E_U$ intersects $B' \times E_U$, where B' is a ball in $F_U \subset P_U \subset \mathcal{X}$ of radius depending on R and the quasiisometry constants. Such a bound exists because G acts on itself geometrically. ◁

We now follow the proof of Theorem 14.3 of [5]. Let $\epsilon > 0$ be given and let $R \geq 1000\epsilon$. Consider the set $\mathfrak{H}$ of $g \in \overline{G}_U$ such that $\mathsf{d}_U(x, gx), \mathsf{d}_U(y, gy) < \epsilon$, where $x, y \in F_U$. Choose s_0 as in the distance formula for $(F_U, \mathfrak{S}_U)$ and, for each $r \geq 0$, consider the set $\mathfrak{L}(r)$ of $\sqsubseteq$–maximal $V \in \mathfrak{S}_U - \{U\}$ such that $\mathsf{d}_V(x, y) > s_0$ and $\left|\mathsf{d}_U(x, \rho^V_U) - \frac{1}{2}R\right| < r\epsilon$. Arguing exactly as in the proof of Theorem 14.3 of [5] yields a uniform bound on $|\mathfrak{L}(11)|$. We then divide into two cases.

First, if $\mathfrak{L}(10) \neq \varnothing$, then we again argue as in the proof of [5, Theorem 14.3], reaching the conclusion that, if $V \in \mathfrak{L}(10)$ and $g \in \mathfrak{H}$, then $\mathfrak{g}_{P_V}(x)$ coarsely coincides with $g \cdot \mathfrak{g}_{P_V}(x)$, from which it follows from Claim 1 that $\mathfrak{H}$ has uniformly bounded cardinality. The argument in [5] uses only the $\overline{G}_U$–equivariance of the gate construction and Definition 1.1 and thus goes through.

Similarly, if $\mathfrak{L}(10) = \varnothing$, then the argument in [5] uses only the existence of hierarchy paths, large links, bounded geodesic image, the distance formula, and a bound on the cardinalities of stabilizers of balls in F_U. The latter comes from Claim 1, and thus the argument works verbatim in the present context. □

Proof of Lemma 7.4 Let $U \in \mathrm{Big}(g)$. Let $M > 0$ be as in Lemma 6.3 and set $h = g^M$; note that $h \cdot U = U$, ie $h \in \mathcal{A}_U$. Let $h_U = \theta_U(h) \in \overline{G}_U$. By Lemma 7.5, $\overline{G}_U$ acts acylindrically on $\mathcal{C}U$, so by Lemma 7.2, h_U is either elliptic or loxodromic on $\mathcal{C}U$. Since $U \in \mathrm{Big}(h)$, it must be the case that h_U is loxodromic on $\mathcal{C}U$. Since h acts like h_U on $\mathcal{C}U$, the claim follows. □

8 Essential structures, essential actions and product HHSs

8.1 Product HHSs

It is shown in [6] that, if $\mathcal{X}_0, \mathcal{X}_1$ admit hierarchically hyperbolic structures, then $\mathcal{X}_0 \times \mathcal{X}_1$ admits a hierarchically hyperbolic structure making the inclusions $\mathcal{X}_i \to \mathcal{X}_0 \times \mathcal{X}_1$ into hieromorphisms with hierarchically quasiconvex image. Rather than recall the construction, we now give a more streamlined (equivalent) definition.

Definition 8.1 Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space. Then $(\mathcal{X}, \mathfrak{S})$ is a *product HHS* if there exists $K < \infty$ and $U \in \mathfrak{S}$ such that for all $V \in \mathfrak{S}$, either $V \sqsubseteq U$, or $V \perp U$, or $\mathrm{diam}(\mathcal{C}V) \leq K$. If, in addition, for each $n \in \mathbb{N}$ there exist $V, W \in \mathfrak{S}$ with $V \sqsubseteq U$, $W \perp U$ and $\mathrm{diam}(\pi_V(\mathcal{X})), \mathrm{diam}(\pi_W(\mathcal{X})) > n$, then $(\mathcal{X}, \mathfrak{S})$ is a *product region with unbounded factors*. Observe that $(\mathcal{X}, \mathfrak{S})$ is a product HHS if and only if there exists $U \in \mathfrak{S}$ such that $P_U \to \mathcal{X}$ is coarsely surjective, and that $(\mathcal{X}, \mathfrak{S})$ is a product region with unbounded factors if in addition F_U and E_U are both unbounded.

8.2 Essential structures and cores

Definition 8.2 (essential HH structures) Let $(\mathcal{X}, \mathfrak{S})$ be an HHS and let $G \leq \operatorname{Aut}(\mathfrak{S})$. Then $(\mathcal{X}, \mathfrak{S})$ is G*–essential* if, for any G–invariant hierarchically quasiconvex $\mathcal{Y} \subset \mathcal{X}$, all of $\mathcal{X}$ is contained in some regular neighborhood of $\mathcal{Y}$.

Remark 8.3 Compare Definition 8.2 to the definition of a G–essential cube complex from [18], which requires that the cube complex be the cubical convex hull of a G–orbit (but actually requires something stronger).

Proposition 8.4 (essential core) *Let $(\mathcal{X}, \mathfrak{S})$ be an HHS and let $G \leq \operatorname{Aut}(\mathfrak{S})$ be a subgroup. Suppose that one of the following holds:*

1. *G acts properly and cocompactly on $\mathcal{X}$ and with finitely many orbits on $\mathfrak{S}$, ie $(G, \mathfrak{S})$ is an HHG.*
2. *G acts on $\mathcal{X}$ with unbounded orbits and with no fixed point in $\partial\mathcal{X}$.*

Then there exists a G–invariant, G–essential, hierarchically quasiconvex subspace $\mathcal{Y} \subset \mathcal{X}$ such that whichever of (1) or (2) held for $G \curvearrowright \mathcal{X}$ holds for the action of G on $\mathcal{Y}$.

Proof If $(\mathcal{X}, \mathfrak{S})$ is an HHG, the claim follows immediately with $\mathcal{Y} = \mathcal{X}$. In the second case, we will build $\mathcal{Y} \subset \mathcal{X}$ so that $\mathcal{Y}$ is hierarchically quasiconvex and G–invariant, with the property that if $\mathcal{Y}' \subset \mathcal{X}$ is hierarchically quasiconvex and G–invariant, then there exists an $R > 0$ such that $\mathcal{Y} \subset \mathcal{N}_R(\mathcal{Y}')$. Given such a $\mathcal{Y}$, the fact that G does not fix a point in $\partial\mathcal{Y}$ follows from Proposition 5.8 and the hypothesis that G does not fix a point in $\partial\mathcal{X}$.

To construct $\mathcal{Y}$, for each $U \in \mathfrak{S}$, let $H_U \subseteq \mathcal{C}U$ be the union of all geodesics starting and ending in $\pi_U(G \cdot x)$ for some fixed basepoint $x \in \mathcal{X}$. A thin quadrilateral argument shows that H_U is uniformly quasiconvex. Let $\mathcal{Y}$ consist of all realization points y with $\pi_U(y) \in H_U$ for all $U \in \mathfrak{S}$; this subspace is easily seen to have the required properties. □

Recall that, by hierarchical quasiconvexity, $(\mathcal{Y}, \mathfrak{S})$ is normalized: for each $U \in \mathfrak{S}$, the associated hyperbolic space is uniformly quasiisometric to $\pi_U(\mathcal{Y}) \subseteq \mathcal{C}U$.

9 Coarse rank-rigidity and its consequences

Throughout this section, $(\mathcal{X}, \mathfrak{S})$ is a hierarchically hyperbolic space with $\mathcal{X}$ proper and $\mathfrak{S}$ countable; we always let S denote the $\sqsubseteq$–maximal element of $\mathfrak{S}$. In this section, we consider countable subgroups $G \leq \mathrm{Aut}(\mathfrak{S})$ (so that, by the distance formula, G acts discretely on $\mathcal{X}$). These standing hypotheses cover the case where $(G, \mathfrak{S})$ is an HHG. We emphasize our standing assumption that all HHSs are normalized.

Definition 9.1 (rank-one automorphism) The automorphism $g \in \mathrm{Aut}(\mathfrak{S})$ is *rank-one (on* $(\mathcal{X}, \mathfrak{S})$*)* if

- g is axial;
- $|\mathrm{Big}(g)| = 1$;
- if $U \in \mathfrak{S}$ is orthogonal to the domain in $\mathrm{Big}(g)$, then $\mathrm{diam}(\pi_U(\mathcal{X})) < \infty$.

Irreducible axial elements are rank-one.

Our first goal is to show that, under the above hypotheses, either G contains an irreducible axial element or the G–essential core of $\mathcal{X}$ is a product HHS (not necessarily with unbounded factors). This is done in Section 9.1, using tools from Sections 9.2, 9.3 and 9.4. In Section 9.5, we apply results of Section 9.1.

9.1 Irreducible axials or fixed domains

We now prove the following two parallel propositions (one covering the nonparabolic case, and one covering the HHG case):

Proposition 9.2 *Let $(\mathcal{X}, \mathfrak{S})$ be an HHS with $\mathcal{X}$ proper and $\mathfrak{S}$ countable. Let the countable group $G \leq \mathrm{Aut}(\mathfrak{S})$ act with unbounded orbits in $\mathcal{X}$ and without a global fixed point in $\partial \mathcal{C}S$. Then either G contains an irreducible axial element, or there exists $U \in \mathfrak{S} - \{S\}$ such that $|G \cdot U| < \infty$. Moreover, any G–essential hierarchically quasiconvex subspace $\mathcal{Y} \subset \mathcal{X}$ coarsely coincides with the standard product region $P_U \cap \mathcal{Y}$.*

Proof By Proposition 8.4, there exists a G–invariant hierarchically quasiconvex subspace $\mathcal{Y}$ with a hierarchically hyperbolic structure $(\mathcal{Y}, \mathfrak{S})$ admitting a G–equivariant hieromorphism $(\mathcal{Y}, \mathfrak{S}) \to (\mathcal{X}, \mathfrak{S})$ that is the inclusion on $\mathcal{Y}$ and the identity on $\mathfrak{S}$, and such that $(\mathcal{Y}, \mathfrak{S})$ is G–essential. Moreover, G continues to act without a global fixed point in $\partial \mathcal{C}S$. Hence, since $\mathcal{Y}$ is proper and $\mathfrak{S}$ is countable, Proposition 9.11 provides an irreducible axial isometry of $(\mathcal{Y}, \mathfrak{S})$ (hence of $(\mathcal{X}, \mathfrak{S})$) unless $\mathrm{diam}(\pi_S(\mathcal{Y})) < \infty$. If $\mathrm{diam}(\pi_S(\mathcal{Y})) < \infty$, then Proposition 9.10 completes the proof. □

The HHG version requires the following theorem of Osin, which we also use elsewhere:

Theorem 9.3 [65, Theorem 1.1] *Let G be a group acting acylindrically on a hyperbolic space. Then exactly one of the following holds:*

(1) *G has bounded orbits.*

(2) *G is virtually infinite cyclic and contains a loxodromic element.*

(3) *G contains infinitely many independent loxodromic elements.*

Proposition 9.4 *Let $(G, \mathfrak{S})$ be an HHG. Then either G contains an irreducible axial element or there exists $U \in \mathfrak{S}$ such that $|G \cdot U| < \infty$ and G coarsely coincides with P_U.*

Proof The G–action on $(G, \mathfrak{S})$ is essential. If $\operatorname{diam}(\mathcal{C}S) = \infty$, then, since G acts acylindrically on $\mathcal{C}S$, as proved in [5, Section 14], Theorem 9.3 implies that G contains an irreducible axial element. Hence we can assume that $\operatorname{diam}(\mathcal{C}S) < \infty$, and in particular that G has no fixed point in $\partial\mathcal{C}S = \varnothing$. The claim now follows from Proposition 9.10. □

9.2 Finding finite orbits in $\mathfrak{S}$

Let μ be a probability measure on G, whose support generates G. All spaces are equipped with their Borel σ–algebra, so every subset of G is measurable, while the measurable subsets of $\overline{\mathcal{X}}$ are determined by Definition 2.11.

Lemma 9.5 (stationary measure on $\overline{\mathcal{X}}$) *There exists a μ–stationary probability measure ν on $\overline{\mathcal{X}}$, ie for all ν–measurable $E \subseteq \overline{\mathcal{X}}$,*

$$\nu(E) = \sum_{g \in G} \mu(g)\nu(g^{-1}E) = \mu * \nu(E).$$

Proof This is a standard fact, relying on compactness of $\overline{\mathcal{X}}$, ie Theorem 3.4. See [34, Lemma 1.2], for example. □

Remark 9.6 (sampling $\mathcal{X}$) Since our aim in this section is to establish that, after passing if necessary to a G–essential core, G contains an irreducible axial element or $\mathcal{X}$ is a product HHS, and these properties are insensitive to modifications of $\mathcal{X}$ within its quasiisometry type, we now "discretize" $\mathcal{X}$, for convenience in the proof of Lemma 9.8.

Let $\mathcal{D} = G\backslash\mathcal{X}$, and let $\bar{\mathrm{d}}$ be the quotient pseudometric, so $(\mathcal{D}, \bar{\mathrm{d}})$ is proper since $\mathcal{X}$ is proper. Hence there exists $\epsilon > 0$ and a countable set $\{\bar{x}_n\}_{n \geq 0}$ in $\mathcal{D}$ such that $\mathcal{N}^{\mathcal{D}}_{\epsilon}(\{\bar{x}_n\}) = \mathcal{D}$. Thus $\mathcal{X}$ contains a countable, G–invariant set $\{x_n\}_{n \geq 0}$ for which the inclusion $\{x_n\} \hookrightarrow \mathcal{X}$ is a quasiisometry, and we replace $\mathcal{X}$ with $\{x_n\}$. We can thus assume that $\mathcal{X}$ is countable.

Lemma 9.7 *For each $\mathcal{U} \subset \mathfrak{S}$, the set $\{p \in \partial\mathcal{X} : \operatorname{Supp}(p) = \mathcal{U}\}$ is ν–measurable.*

Proof Either $\{p \in \partial\mathcal{X} : \mathrm{Supp}(p) = \mathcal{U}\} = \varnothing$, in which case we're done, or $\mathcal{U} = \{U_i\}$ is a set of pairwise-orthogonal domains. Let $\mathcal{X}_0$ be the set of points $q \in \partial\mathcal{X}$ such that, for all $V \in \mathrm{Supp}(q)$, there exists $U \in \mathcal{U}$ with $V \sqsubseteq U$. Note that

$$\mathcal{Y} = \{p \in \partial\mathcal{X} : \mathrm{Supp}(p) = \mathcal{U}\} \subseteq \mathcal{X}_0.$$

Let $\mathcal{X}_1$ be the subset of $\mathcal{X}_0$ consisting of those $q \in \mathcal{X}_0$ such that for some $V \in \mathrm{Supp}(q)$, we have $V \notin \mathcal{U}$ (so V is properly nested in some $U \in \mathcal{U}$ and orthogonal to the remaining elements).

$\mathcal{X}_0$ is closed in $\overline{\mathcal{X}}$ We will check that for any sequence $\{q_n\}$ with each $q_n \in \mathcal{X}_0$, if $q_n \to q$, then $q \in \mathcal{X}_0$. Suppose not, ie suppose that there exists $V \in \mathrm{Supp}(q)$ such that $V \not\sqsubseteq U$ for all $U \in \mathcal{U}$. Consider a basic neighborhood $\mathcal{N} = \mathcal{N}_{\epsilon,\{N_T\}}(q)$ of q. There are two cases.

First case This is the case where there exists $U \in \mathcal{U}$ such that $U \pitchfork V$ or $U \sqsubsetneq V$ and, for infinitely many n, there exists $W \in \mathrm{Supp}(q_n)$ such that $W \sqsubseteq U$ and $W \not\perp V$. Let $\mathcal{I}$ be the set of such n.

First, suppose that q_n is remote with respect to q. Suppose that the basic neighborhood $\mathcal{N}$ has been chosen so that N_V does not meet the 10^9E–neighborhood of ρ^U_V. Then for arbitrarily large $n \in \mathcal{I}$, the subsets ρ^U_V and ρ^W_V coarsely coincide, and hence $(\partial\pi_{\mathrm{Supp}(q)}(q_n))_V = \rho^W_V$ does not lie in N_V. It follows that for arbitrarily large $n \in \mathcal{I}$, we have $q_n \notin \mathcal{N}$, by the definition of the remote part of a basic set. This is a contradiction.

Second, suppose that q_n is nonremote with respect to q, where $n \in \mathcal{I}$. Exactly as before, suppose that N_V does not meet the 10^9E–neighborhood of ρ^U_V (which is still defined by assumption). We still have that ρ^W_V is defined and coarsely coincides with ρ^U_V for some $W \in \mathrm{Supp}(q_n)$, by assumption. Hence, again, we have that $(\partial\pi_{\mathrm{Supp}(q)}(q_n))_V = \rho^W_V$ does not lie in N_V. From the final condition in the definition of the nonremote part of a basic set, it follows that $q_n \notin \mathcal{N}$, which is again a contradiction.

Second case In this case, for all but finitely many n, we have $V \perp W$ for all $W \in \mathrm{Supp}(q_n)$. The point q_n is nonremote with respect to q. Indeed, there exists $V \in \mathrm{Supp}(q)$ which is orthogonal to every element of $\mathrm{Supp}(q_n)$. In particular, $V \in \mathrm{Supp}(q) - \mathrm{Supp}(q_n) \cap \mathrm{Supp}(q)$. Now, $\sum_{T \in \mathrm{Supp}(q_n) - \mathrm{Supp}(q)} a^{q_n}_T < \epsilon$, so

$$\sum_{T \in \mathrm{Supp}(q_n) \cap \mathrm{Supp}(q)} a^{q_n}_T > 1 - \epsilon,$$

while $|a^q_T - a^{q_n}_T| < \epsilon$ whenever $T \in \mathrm{Supp}(q_n) \cap \mathrm{Supp}(q)$. Hence

$$\sum_{T \in \mathrm{Supp}(q) \cap \mathrm{Supp}(q_n)} a^q_T > 1 - \epsilon\big(|\mathrm{Supp}(q_n) \cap \mathrm{Supp}(q)|\big),$$

which is impossible when ϵ is sufficiently small compared to a_V^q, since $V \notin \mathrm{Supp}(q_n)$. Hence $q_n \notin \mathcal{N}$, a contradiction.

Conclusion Let $\mathfrak{T}$ be the set of support sets $\mathcal{V} \neq \mathcal{U}$ such that for each $V \in \mathcal{V}$, there exists $U \in \mathcal{U}$ with $V \sqsubseteq U$. Then $\mathfrak{T}$ is countable, being a set of finite subsets of the countable set $\mathfrak{S}$. Now, $\mathcal{X}_1$ is the union over all $\mathcal{V} \in \mathfrak{T}$ of the set $\mathcal{X}_0(\mathcal{V})$ of $q \in \partial\mathcal{X}$ such that for each $W \in \mathrm{Supp}(q)$, there exists $V \in \mathcal{V}$ with $W \sqsubseteq V$. Hence, by the previous part of the proof, $\mathcal{X}_1$ is a countable union of closed sets. Thus $\mathcal{Y} = \mathcal{X}_0 - \mathcal{X}_1$ is Borel, and hence ν–measurable. □

Lemma 9.8 *If G has no finite orbit in $(\mathfrak{S} - \{S\}) \cup \partial\mathcal{C}S$, then ν is supported on $\partial\mathcal{C}S \subset \overline{\mathcal{X}}$.*

Proof Let D be the set of finite subsets of $\mathfrak{S}$, so that D is countable and G acts on D in the obvious way. By construction, $\{S\}$ and $\varnothing$ are the only elements of D whose G–orbits are finite. We first define a map $\mathcal{O}\colon \overline{\mathcal{X}} \to D$. Note that if $\mathfrak{S} = \{S\}$, then $\partial\mathcal{X} = \partial\mathcal{C}S$, and the claim follows, so we assume that there exists $U \sqsubsetneq S$.

Defining $\mathcal{O}$ on boundary points For each $p \in \partial\mathcal{X}$, let $\mathcal{O}(p) = \mathrm{Supp}(p)$. Observe that this assignment is G–equivariant and that $\mathcal{O}(p) = \{S\}$ if and only if $p \in \partial\mathcal{C}S$.

Defining $\mathcal{O}$ on interior points Let $\mathcal{B} \subset \mathcal{X}$ contain exactly one point from each G–orbit, and choose $F \in D - \{\{S\}, \varnothing\}$. For each $x \in \mathcal{B}$, let $\mathcal{O}(x) = F$. Then, for any $x \in \mathcal{B}$ and $g \in G$, let $\mathcal{O}(gx) = gF$. Then $\mathcal{O}$ is G–equivariant and, for all $x \in \mathcal{X}$, the nonempty finite set $\mathcal{O}(x)$ differs from $\{S\}$. For any $F' \in D$, either $\mathcal{O}^{-1}(F') = \varnothing$ or $F' = gF$ for some $g \in G$. Hence, for any subset D' of D, we can write $\mathcal{O}^{-1}(D') = \bigcup_{gF \in D'} g\mathcal{B}$. It follows that $\mathcal{O}^{-1}(D')$ is a countable union of translates of $\mathcal{B}$, which is a countable union of closed sets (singletons) by Remark 9.6, and thus $\mathcal{O}^{-1}(D')$ is Borel.

Measurability of $\overline{\mathcal{X}} - \partial\mathcal{C}S$ Since $\partial\mathcal{C}S = \big\{p \in \partial\mathcal{X} : \mathrm{Supp}(p) = \{S\}\big\}$, it follows from Lemma 9.7 that $\overline{\mathcal{X}} - \partial\mathcal{C}S$ is measurable.

Measurability of $\mathcal{O}$ There is a probability measure $\tilde{\nu}$ on D given by $\tilde{\nu}(A) = \nu(\mathcal{O}^{-1}(A))$, for each $A \subseteq D$. A set $\mathcal{O}^{-1}(A)$ decomposes as

$$\{x \in \mathcal{X} : \mathcal{O}(x) \in A\} \cup \{p \in \partial\mathcal{X} : \mathrm{Supp}(p) \in A\}.$$

The set $\{p \in \partial\mathcal{X} : \mathrm{Supp}(p) \in A\} = \bigcup_{\mathcal{U} \in A}\{p : \mathrm{Supp}(p) = \mathcal{U}\}$, which is ν–measurable by Lemma 9.7. Since $A \subseteq D$ is countable, it suffices to show that $\mathcal{O}^{-1}(F) \cap \mathcal{X}$ is Borel for each $F \in D$, but this was established above.

Conclusion We have that $\mathcal{O}\colon \overline{\mathcal{X}} \to D$ is a measurable G–equivariant map. Since G preserves $\partial\mathcal{C}S$, it follows that $\overline{\mathcal{X}} - \partial\mathcal{C}S$ is a G–invariant ν–measurable set.

Suppose that $F' \in D$ has the property that $G \cdot F'$ is finite. Then $G \cdot U$ is a finite G–invariant subset of $\mathfrak{S}$ for each $U \in F'$ and, by our hypothesis that there is no finite G–orbit in $\mathfrak{S} - \{S\}$, we have that $F' = \{S\}$. Since $\mathcal{O}(e) \neq \{S\}$ for all $e \in \overline{\mathcal{X}} - \partial\mathcal{C}S$, it follows that $\mathcal{O}(\overline{\mathcal{X}} - \partial\mathcal{C}S)$ does not contain a finite G–orbit. As shown in eg [2; 45, Lemma 2.2.2; 75, Lemma 3.4; 44, Lemma 3.3], we must have $\nu(\overline{\mathcal{X}} - \partial\mathcal{C}S) = 0$. □

Corollary 9.9 *If* $\operatorname{diam}(\mathcal{C}S) < \infty$, *then* G *stabilizes a finite subset of* $\mathfrak{S} - \{S\}$.

Proof By hypothesis, $\partial\mathcal{C}S = \varnothing$, so ν cannot be supported on $\partial\mathcal{C}S$. Hence G has a finite orbit in $\mathfrak{S} \cup \partial\mathcal{C}S$ by Lemma 9.8 and thus G must have a finite orbit in $\mathfrak{S} - \{S\}$. □

9.3 Finding product structures when $\operatorname{diam}(\mathcal{C}S) < \infty$

Proposition 9.10 *Suppose* $G \leq \operatorname{Aut}(\mathfrak{S})$ *is a countable subgroup with* $\operatorname{diam}(\mathcal{C}S) < \infty$. *Then there exists* $U \in \mathfrak{S} - \{S\}$ *and a finite-index subgroup* G' *such that* $G' \cdot U = U$ *and* $\mathcal{X}$ *coarsely coincides with* P_U. *Hence either* $(\mathcal{X}, \mathfrak{S})$ *is a product HHS with unbounded factors or* $\mathcal{X}$ *coarsely coincides with* F_U *or* E_U.

Proof By Corollary 9.9, there exists $U \in \mathfrak{S} - \{S\}$ and a finite-index subgroup $G' \leq G$ such that $G' \cdot U = U$. Note that G' continues to act essentially on $(\mathcal{X}, \mathfrak{S})$, coarsely stabilizing P_U. Since P_U is hierarchically quasiconvex, $\mathcal{X}$ coarsely equals P_U by essentiality. The last assertion is immediate. □

9.4 Finding irreducible axial elements when $\operatorname{diam}(\mathcal{C}S) = \infty$

Proposition 9.11 *Let* $(\mathcal{X}, \mathfrak{S})$ *be a hierarchically hyperbolic space. Let* $G \leq \operatorname{Aut}(\mathfrak{S})$ *act essentially and suppose that* G *acts on* $\mathcal{X}$ *with no global fixed point in* $\partial\mathcal{C}S$ *and that* $\mathcal{C}S$ *is unbounded. Then* G *contains an irreducible axial automorphism of* $(\mathcal{X}, \mathfrak{S})$.

Proof Suppose that every orbit of G in $\mathcal{C}S$ is bounded, so that, fixing $x_0 \in \mathcal{X}$, there exist $Q, R < \infty$ such that $\operatorname{diam}_S(G \cdot \pi_S(x_0)) \leq R$ and $G \cdot \pi_S(x_0)$ is Q–quasiconvex. Consider the set of all E–consistent tuples $(b_U)_{U \in \mathfrak{S}}$ such that $b_S \in G \cdot \pi_S(x_0)$. Let $\mathcal{Y}$ be the set of realization points in $\mathcal{X}$ corresponding to such tuples, provided by Theorem 1.7, and note that G acts on $\mathcal{Y}$. By definition, $\mathcal{Y}$ is hierarchically quasiconvex in $\mathcal{X}$ provided $\pi_U(\mathcal{Y})$ is uniformly quasiconvex in $\mathcal{C}U$ for each $U \in \mathfrak{S}$, which we now verify.

If $\vec{b}$ is such a tuple, with $d_S(b_S, \rho_S^U) \leq E$, then consistency puts no constraint on the U–coordinate of $\vec{b}$, ie for any such U, the map $\pi_U \colon \mathcal{Y} \to \mathcal{C}U$ is uniformly coarsely surjective, and in particular $\pi_U(\mathcal{Y})$ is uniformly quasiconvex in $\mathcal{C}U$. On the other hand, if $d_S(\rho^U_S, G \cdot \pi_S(x_0)) > E$, then consistency and bounded geodesic image imply that $\pi_U(\mathcal{Y})$ is uniformly bounded, and hence uniformly quasiconvex.

The existence of $\mathcal{Y}$ contradicts G–essentiality of $\mathcal{X}$. Hence G has an unbounded orbit in $\mathcal{C}S$, so either there exists $g \in G$ acting loxodromically on $\mathcal{C}S$, so g is irreducible axial, or there exists a unique fixed point $p \in \partial\mathcal{C}S$, which is impossible. □

9.5 Coarse rank-rigidity

Recall that a metric space $\mathcal{X}$ is *wide* if no asymptotic cone of $\mathcal{X}$ has a cut-point. The following lemma is well-known and elementary:

Lemma 9.12 *Let $\mathcal{X}$ be a metric space quasiisometric to the product $\mathcal{X}_0 \times \mathcal{X}_1$, where each $\mathcal{X}_i$ is unbounded. Then $\mathcal{X}$ is wide, ie no asymptotic cone of $\mathcal{X}$ has a cut-point.*

We now prove the main theorems of this section. Much of the work was done in proving Propositions 9.2 and 9.4; the remaining work is largely in sorting out technical issues that arise when attempting to induct on complexity; these issues mainly stem from the fact that, given $U \in \mathfrak{S}$, the induced HHS structure on E_U does not have a uniquely determined $\sqsubseteq$–maximal element.

Theorem 9.13 (coarse rank-rigidity for nonparabolic actions) *Let $(\mathcal{X}, \mathfrak{S})$ be an HHS with $\mathcal{X}$ proper and $\mathfrak{S}$ countable. Let the countable group $G \leq \mathrm{Aut}(\mathfrak{S})$ act essentially with unbounded orbits in $\mathcal{X}$ and without a fixed point in $\partial(\mathcal{X}, \mathfrak{S})$. Then one of the following holds:*

(1) *$\mathcal{X}$ is a product HHS with unbounded factors; specifically, $\mathcal{X}$ is coarsely equal to P_U for some $U \in \mathfrak{S}$ with $|GU| < \infty$.*

(2) *There exists $g \in G$ such that g is rank-one.*

If conclusion (1) holds, then $\mathcal{X}$ is wide.

Proof By Proposition 9.2, either G contains an irreducible axial element, which is rank-one by definition, so conclusion (2) holds, or there is a finite-index subgroup $G' \leq G$ fixing some $U \in \mathfrak{S} - \{S\}$, so that by essentiality, $\mathcal{X}$ coarsely coincides with the standard product region P_U. This implies that $\mathcal{X}$ is a product HHS. Choose U of minimal level with this property, ie no domain of lower level has a finite G–orbit in $\mathfrak{S}$.

Since G has unbounded orbits in $\mathcal{X}$, at least one of E_U and F_U is unbounded. If F_U and E_U are both unbounded, then conclusion (1) holds, and we are done. The statement about wideness follows from Lemma 9.12.

If F_U is unbounded and E_U is bounded, then $(F_U, \mathfrak{S}_U)$ is an HHS with F_U proper and $\mathfrak{S}_U$ countable, on which G' acts by HHS automorphisms with no fixed point in $\partial\mathcal{C}U$ (for otherwise G would have a fixed point in $\partial\mathcal{X}$). By minimality, G' has no finite orbit in $\mathfrak{S}_U - \{U\}$, so Proposition 9.11 provides $g \in G'$ acting as an irreducible axial element of $\mathrm{Aut}(\mathfrak{S}_U)$. As an element of $\mathrm{Aut}(\mathfrak{S})$, we see that g is rank-one, for otherwise there would be some $V \perp U$ with $\mathrm{diam}(\mathcal{C}V) = \infty$, contradicting that E_U is bounded.

Finally, suppose that E_U is unbounded and F_U is bounded. Let $\mathfrak{C}$ be a minimal G'–invariant set of $\sqsubseteq$–minimal elements C of $\mathfrak{S} - \{S\}$ such that $W \sqsubseteq C$ whenever $W \perp U$.

Suppose that there exists $C \in \mathfrak{C}$ with $C \perp U$. Then $g \cdot C \perp g \cdot U = U$ for all $g \in G'$, so $g \cdot C \sqsubseteq C$, from which it follows that (passing if necessary to a further finite-index subgroup if necessary) $G' \cdot C = C$. Then $(E_U, \mathfrak{S}_C)$ is an HHS satisfying the hypotheses of the theorem, and $G' \leq \mathrm{Aut}(\mathfrak{S}_C)$ acts without a fixed point in ∂E_U (since it stabilizes $\partial E_U \subset \partial\mathcal{X}$). In this case, the claim follows by induction on complexity. Indeed, in the base case, $|\mathfrak{S}| = 1$ and the theorem is obvious. Otherwise, induction shows that either conclusion (1) holds, or there exists $g \in G$ that acts as a rank-one element of $\mathrm{Aut}(\mathfrak{S}_C)$. Since G' preserves P_U and P_U coarsely equals $\mathcal{X}$, this implies that g is rank-one on $(\mathcal{X}, \mathfrak{S})$, as required.

The definition of $\mathfrak{C}$ and Definition 1.1(3) imply that $C \not\sqsubseteq U$ and $U \not\sqsubseteq C$ for all $C \in \mathfrak{C}$. Hence it remains to consider the case where each $C \in \mathfrak{C}$ satisfies $C \pitchfork U$; fix such a C. Since G' stabilizes U, it coarsely stabilizes the image $\overline{P}_U$ of $P_U = F_U \times E_U \to \mathcal{X}$. In other words, for any basepoint $x \in \mathcal{X}$, the orbit $G' \cdot x$ lies in a neighborhood of $\overline{P}_U$. Now, since $C \pitchfork U$, the definition of P_U implies that $\pi_C(gx)$ uniformly coarsely coincides with ρ^U_C for all $g \in G'$, whence $\mathrm{diam}(\pi_C(G' \cdot x)) < \infty$, so, by essentiality, $\mathrm{diam}(\pi_C(\mathcal{X})) < \infty$.

In this case, form a new index set $\mathfrak{S}_U^{\perp}$ by appending to the set of domains orthogonal to U a new domain $\boldsymbol{C}$. In $\mathfrak{S}_U^{\perp} \cap \mathfrak{S}$, the associated hyperbolic spaces, projections from E_U, and relative projections are defined as in $\mathfrak{S}$. The hyperbolic space $\mathcal{C}\boldsymbol{C}$ is a single point, so the projections $\pi_{\boldsymbol{C}}\colon \mathcal{X} \to \mathcal{C}\boldsymbol{C}$ and $\rho^V_{\boldsymbol{C}}$ for $V \perp U$ are defined in an obvious way. We thus have an HHS structure $(E_U, \mathfrak{S}_U^{\perp})$ with $G' \leq \mathrm{Aut}(\mathfrak{S}_U^{\perp})$, of complexity less than that of $\mathfrak{S}$, and we can argue as above by induction. Observe that, if $g \in \mathrm{Aut}(\mathfrak{S}_U^{\perp})$ is rank-one on E_U, then $\mathrm{Big}(g)$ consists of some element of $\mathfrak{S}_U^{\perp} \cap \mathfrak{S}$, and since $\pi_C(\mathcal{X})$ is bounded for all $C \in \mathfrak{C}$, and we can argue as above that g is rank-one on $(\mathcal{X}, \mathfrak{S})$. □

Theorem 9.14 (coarse rank-rigidity for HHG) *Let $(G, \mathfrak{S})$ be an infinite hierarchically hyperbolic group. Then exactly one of the following holds:*

(1) *$(G, \mathfrak{S})$ is a product HHS with unbounded factors, and G is wide.*

(2) *G contains a rank-one element, and is thus not wide.*

Moreover, conclusion (1) holds if and only if $\mathrm{diam}(\mathcal{C}S) < \infty$.

Proof By Proposition 9.4, either G contains an irreducible axial element, which is rank-one, or there exists $U \in \mathfrak{S} - \{S\}$ with $G' \cdot U = U$ for some finite-index $G' \leq G$, and G coarsely coincides with P_U. In the latter case, we argue as in the proof of Theorem 9.13, by induction on complexity, using the following observation: if $V \in \mathfrak{S} - \{S\}$ and a finite-index subgroup G' fixes V, then the action of G' on F_V is proper and cobounded. Moreover, G' acts with finitely many orbits on $\mathfrak{S}_V$, so $(G', \mathfrak{S}_V)$ is an HHG structure on G', enabling induction. □

9.6 Tits alternative for HHGs

The goal of this subsection is the following theorem:

Theorem 9.15 (Tits alternative for HHGs) *Let $(G, \mathfrak{S})$ be an HHG and let $H \leq G$. Then H either contains a nonabelian free group or is virtually abelian.*

Before we proceed with the proof, we need some supporting results:

Proposition 9.16 *Let $(G, \mathfrak{S})$ be a hierarchically hyperbolic group. Then any $H \leq G$ containing an irreducible axial element is virtually $\mathbb{Z}$ or contains a nonabelian free group.*

Proof Since G acts on $\mathcal{C}S$ acylindrically [5], and hence $H \leq G$ does, Theorem 9.3 implies that either H is virtually cyclic or H contains irreducible axial elements g and h such that $\{h^{\pm}\} \cap \{g^{\pm}\} = \varnothing$. Propositions 6.18 and 2.17(1) enable an application of the ping-pong lemma, showing that g^N and h^N freely generate a free subgroup F for some $N > 0$. Or, one can apply [5, Corollary 14.6], which uses [32, Proposition 2.4]. □

Lemma 9.17 *Let $(G, \mathfrak{S})$ be an HHG with $S \in \mathfrak{S}$ $\sqsubseteq$–maximal. Suppose that $H \leq G$ has bounded orbits in $\mathcal{C}S$ and fixes some $p \in \partial\mathcal{C}S$. Then $|H| < \infty$.*

Proof By Theorem 14.3 of [5], G acts acylindrically on $\mathcal{C}S$, ie for each $\epsilon > 0$, there exists $R \geq 0$ and $N \in \mathbb{N}$ such that whenever $s, s' \in \mathcal{C}S$ satisfy $\mathrm{d}_S(s, s') \geq R$, there are at most N elements $g \in G$ for which $\mathrm{d}_S(s, g \cdot s), \mathrm{d}_S(s', g \cdot s') \leq \epsilon$.

Fix $s \in \mathcal{C}S$ and let ϵ_1 bound the diameter of the orbit $H \cdot s$. Let γ be a $(1, 20\delta)$–quasigeodesic ray with endpoint p and initial point s, where $\mathcal{C}S$ is δ–hyperbolic. Then, for all $h \in H$, the ray $h \cdot \gamma$ emanates from $h \cdot s$ and has endpoint $h \cdot p = p$. This fact, together with a thin quadrilateral argument, shows that there exists $k = k(\delta)$

and R_0 such that for all $h \in H$, we have $\mathsf{d}_S(t, h\cdot t) \le k\delta$ whenever $t \in \gamma$ satisfies $\mathsf{d}_S(s,t) \ge R_0$. Let $\epsilon = \max\{\epsilon_1, k\delta\}$ and let R and N be the associated constants coming from acylindricity. Then we can choose $t \in \gamma$ so that $\mathsf{d}_S(s,t) > R$ while $\mathsf{d}_S(s, h\cdot s)$, $\mathsf{d}_S(t, h\cdot t) \le \epsilon$ for all $h \in H$, and hence $|H| \le N$. □

Proof of Theorem 9.15 Note that H is a countable subgroup of $\mathrm{Aut}(\mathfrak{S})$, since G is finitely generated. We divide into cases, according to whether H fixes some $p \in \partial G$.

***H* fixes $p \in \partial \mathcal{C}S$** In this case, by Proposition 9.16, H is either virtually cyclic, contains a nonabelian free group, or, by Theorem 9.3, H has a bounded orbit in $\mathcal{C}S$. Lemma 9.17 implies that H is finite in the latter case.

***H* has no fixed boundary point** Suppose there is an irreducible axial $g \in H$. Then either H contains a nonabelian free group or H is virtually $\mathbb{Z}$, by Proposition 9.16.

Otherwise, Proposition 9.2 provides $U \in \mathfrak{S} - \{S\}$ such that $H \cdot U$ is finite and the H–essential core $\mathcal{Y}$ of in G coarsely coincides with $P_U \cap \mathcal{Y}$. By replacing H with a finite-index subgroup if necessary, we can assume that $H \cdot U = U$.

Thus we have an H–essential product HHS $(\mathcal{X}_0 \times \mathcal{X}_1, \mathfrak{S}^\times)$ with $H \le \mathrm{Aut}(\mathfrak{S}^\times)$ acting on $\mathcal{X}_0 \times \mathcal{X}_1$. Here $\mathfrak{S}^\times$ consists of two disjoint subsets $\mathfrak{S}_0$ and $\mathfrak{S}_1$, together with various domains whose associated spaces are uniformly bounded, with the property that $U_0 \perp U_1$ for all $U_0 \in \mathfrak{S}_0$ and $U_1 \in \mathfrak{S}_1$, and each $\mathfrak{S}_i$ gives $\mathcal{X}_i$ an HHS structure (for more on product decompositions, see [6]). Let $H_i \le H$ be the stabilizer of some (hence any) parallel copy of $\mathcal{X}_i$.

Observe that $H_i \le \mathrm{Aut}(\mathfrak{S}_i)$ is an action on an HHS of strictly lower complexity for $i \in \{0,1\}$, namely $(\mathcal{X}_i, \mathfrak{S}_i)$. If H_i contains no irreducible axial element, then $\mathcal{X}_i$ decomposes as a product HHS, by Theorem 9.13. Otherwise, applying Lemma 7.5 and Theorem 9.3, we see that either H_0 or H_1 (hence H) contains a nonabelian free group, or H_i is virtually $\mathbb{Z}$ for $i \in \{0,1\}$. Hence, either H contains a nonabelian free subgroup, or by induction on complexity, we have a product HHS $(\prod_j L_j^i, \mathfrak{S}_i)$ such that $H_i \le \mathrm{Aut}(\mathfrak{S}_i)$ and each $L_j^i \cong_{\mathrm{qi}} \mathbb{R}$. In the latter case, we conclude that H virtually acts geometrically by HHS automorphisms on $\left(\prod_{ij} L_j^i, \mathfrak{S}^\times\right)$. Hence, for some n, a finite-index subgroup of H acts by uniform quasiisometries on $\mathbb{R}^n$, so H is virtually abelian.

***H* fixes $p \in \partial G - \partial \mathcal{C}S$** In this case, H has a finite-index subgroup fixing some $U \in \mathrm{Supp}(p)$ (so $U \subsetneq S$). We now argue by induction on complexity as above. □

9.7 The "omnibus subgroup theorem"

Our next result generalizes the Handel–Mosher "omnibus subgroup theorem" from [42]. Theorem 9.20 below implies the omnibus subgroup theorem in the case where $\mathcal{X}$ is the mapping class group of a connected, oriented surface of finite type. In order to state the theorem, we need to restrict the class of HHSs we consider, and give some definitions.

Definition 9.18 (hierarchical acylindricity) Given an HHS $(\mathcal{X}, \mathfrak{S})$, we say that $G \leq \mathrm{Aut}(\mathfrak{S})$ is *hierarchically acylindrical* if, for each $U \in \mathfrak{S}$, the image of $G \cap \mathcal{A}_U$ under the restriction homomorphism $\theta_U \colon \mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U)$ acts acylindrically on $\mathcal{C}U$.

Lemma 7.5 implies that every group of automorphisms of an HHG is hierarchically acylindrical. Moreover, hierarchical acylindricity passes to subgroups. For the rest of this subsection, fix $G \leq \mathrm{Aut}(\mathfrak{S})$ to be hierarchically acylindrical.

Definition 9.19 (active domains) Let $G \leq \mathrm{Aut}(\mathfrak{S})$ be a group of HHS automorphisms. We say $U \in \mathfrak{S}$ is an *active domain* for G if $\mathrm{diam}_U(\pi_U(G \cdot x))$ is unbounded for some (hence any) $x \in \mathcal{X}$. Let $\mathfrak{A}(G)$ be the set of $\sqsubseteq$–maximal active domains for G. Note that if $G = \langle g \rangle$, then $\mathfrak{A}(G) = \mathrm{Big}(g)$.

Theorem 9.20 (omnibus subgroup theorem) *Let $(\mathcal{X}, \mathfrak{S})$ be a hierarchically hyperbolic space with $\mathfrak{S}$ countable and $\mathcal{X}$ proper. Let $G \leq \mathrm{Aut}(\mathfrak{S})$ be a countable hierarchically acylindrical subgroup. Then there exists an element $g \in G$ with $\mathfrak{A}(G) = \mathrm{Big}(g)$. Moreover, for any $g' \in G$ and each $U \in \mathrm{Big}(g')$, there exists $V \in \mathrm{Big}(g)$ with $U \sqsubseteq V$.*

Before we prove Theorem 9.20, we prove a lemma related to fixed boundary points of G. Throughout, $\xi(\mathfrak{S})$ denotes the complexity of $(\mathcal{X}, \mathfrak{S})$, ie the length of a longest $\sqsubseteq$–chain.

Definition 9.21 (fixed-point set) Given an arbitrary HHS $(\mathcal{X}, \mathfrak{S})$ and $G \leq \mathrm{Aut}(\mathfrak{S})$, let $\mathrm{Fix}(G) = \{p \in \partial(\mathcal{X}, \mathfrak{S}) \mid G \cdot p = p\}$.

Given $p \in \mathrm{Fix}(G)$, let $G' \leq_{\text{f.i.}} G$ be a finite-index subgroup of G which fixes each $U \in \mathrm{Supp}(p)$. Let $U \in \mathrm{Supp}(p)$ and suppose that G is hierarchically acylindrical. Since G' fixes U, the restriction homomorphism θ_U gives a group G'_U which (coarsely) acts on F_U and acts acylindrically on $\mathcal{C}U$. The next lemma relates supports of fixed points to active domains.

Lemma 9.22 *If $p \in \mathrm{Fix}(G)$, $U \in \mathrm{Supp}(p)$, and $V \in \mathfrak{A}(G)$, then either $U \perp V$ or $U = V$. Moreover, in the latter case, there exists $g'_U \in G'_U$ such that $U \in \mathrm{Big}(g'_U)$ and $\langle g'_U \rangle \leq_{\text{f.i.}} G'_U$.*

Proof We separately analyze two cases.

The case $U \pitchfork V$ or $U \sqsubsetneq V$ Suppose that $U \pitchfork V$ or $U \sqsubsetneq V$, ie ρ^U_V is a well-defined coarse point. Since $G' \cdot U = U$, we have that G' coarsely stabilizes the image of $P_U = F_U \times E_U \to \mathcal{X}$, which we denote by $\overline{P}_U$. In other words, $G' \cdot x_0$ is uniformly close to $\overline{P}_U$ for all $x_0 \in \overline{P}_U$.

By definition of the standard embedding, if $V \pitchfork U$ or $U \sqsubseteq V$, then $\pi_V(\overline{P}_U) \asymp \rho^U_V \in \mathcal{C}V$ (see Section 1.3). Thus for any $x_0 \in \overline{P}_U$ and $V \in \mathfrak{S}$ with $U \pitchfork V$ or $U \sqsubsetneq V$, we have

$$\operatorname{diam}_V(G' \cdot x_0) \asymp 1,$$

which implies that any orbit of G' projects to a bounded subset of $\mathcal{C}V$. Hence $V \notin \mathfrak{A}(G)$, a contradiction. Thus either $V \sqsubseteq U$ or $V \perp U$.

The case $V \sqsubseteq U$ Now suppose $V \sqsubseteq U$. Since $U \in \operatorname{Supp}(p)$, it follows that G'_U fixes a point $p_U \in \partial F_U$, where $p_U \in \partial \mathcal{C}U$. Since G is hierarchically acylindrical, G'_U acts acylindrically on $\mathcal{C}U$. By Theorem 9.3 and the fact that G'_U fixes a point in $\partial\mathcal{C}U$, one of the following holds:

(1) G'_U has bounded orbits in $\mathcal{C}U$.

(2) G'_U contains an element g'_U which acts axially on $\mathcal{C}U$, and $\langle g'_U \rangle \leq_{\text{f.i.}} G'_U$.

If (1) holds, then, since G'_U fixes a point of $\partial\mathcal{C}U$, Lemma 9.17 implies that $|G'_U| < \infty$. In this case, since $V \sqsubseteq U$, we have $\pi_V(G' \cdot x) = \pi_V(G'_U \cdot x)$ is finite, so $V \notin \mathfrak{A}(G)$, a contradiction.

If (2) holds, then we have found the desired element g'_U. Moreover, the existence of this element shows that U is nested into some element of $\mathfrak{A}(G)$. On the other hand, $V \sqsubseteq U$ and $V \in \mathfrak{A}(G)$, so $U = V$ by maximality of V.

Thus the only possibilities are that either $V \perp U$ or $U = V$ and the desired g'_U exists. $\square$

We are now ready for the proof of Theorem 9.20:

Proof of Theorem 9.20 The "moreover" part of the statement follows automatically from the first assertion and the definition of $\mathfrak{A}(G)$, for if $g' \in G$ and $U \in \operatorname{Big}(g')$, then U is an active domain for G and thus U must nest into some domain in $\mathfrak{A}(G) = \operatorname{Big}(g)$.

We now prove the main part of the statement. By Proposition 8.4, we can assume that G acts essentially on $\mathcal{X}$. Let $S \in \mathfrak{S}$ to be the unique $\sqsubseteq$–maximal domain in $\mathfrak{S}$. Note that if G contains an irreducible axial element or has finite order, then we are done. Moreover, by acylindricity of the action of G on $\mathcal{C}S$, either G contains an irreducible axial or has bounded orbits in $\mathcal{C}S$ (so $S \notin \mathfrak{A}(G)$).

In particular, if G fixes a point of $\partial\mathcal{C}S$, then Lemma 9.17 implies that $|G| < \infty$, and we are done. We may therefore assume that G does not fix a point in $\partial\mathcal{C}S$ and $S \notin \mathfrak{A}(G)$.

We now argue by induction on complexity of $\mathfrak{S}$. Suppose that $\xi(\mathfrak{S}) = 1$. Then either there is an irreducible axial element, and we are done, or G acts with bounded orbits on $\mathcal{C}S$, in which case $\mathfrak{A}(G) = \varnothing$ since $\mathfrak{S} = \{S\}$, and we are done.

Now assume that the statement holds for any group of automorphisms of an HHS that satisfies the hypotheses of the theorem and has complexity less than $\xi(\mathfrak{S})$.

There are two main cases, depending on whether or not G has a fixed point in $\partial\mathcal{X}$.

First consider the case where G fixes no point of $\partial\mathcal{X}$. Proposition 9.2 implies that either G contains an irreducible axial, in which case we are done, or there exists $U \in \mathfrak{S} - \{S\}$ such that $|G \cdot U| < \infty$ and $\mathcal{X}$ is coarsely equal to $P_U \subset \mathcal{X}$. In the latter case, after passing to a finite-index subgroup if necessary, we have G acting by automorphisms on the HHS $(P_U, \mathfrak{S})$ (with complexity $\xi(\mathfrak{S})$).

The remaining possibility is that G fixes some $p \in \partial\mathcal{X} - \partial\mathcal{C}S$. In this case, after passing if necessary to a finite-index subgroup, we again find $U \in \mathfrak{S} - \{S\}$ with $GU = U$ and G acting by automorphisms on the HHS $(P_U, \mathfrak{S})$ (with complexity $\xi(\mathfrak{S})$).

In either case, let $P_U = F_U \times E_U$, so that $\mathfrak{S}$ contains orthogonal subsets $\mathfrak{S}_U$ and $\mathfrak{S}_U^{\perp}$ such that $(F_U, \mathfrak{S}_U)$ and $(E_U, \mathfrak{S}_U^{\perp})$ are HHSs of complexity at most $\xi(\mathfrak{S}) - 1$. By replacing G with an index-2 subgroup if necessary, we can assume that G stabilizes $\mathfrak{S}_U$. Moreover, G stabilizes $\mathfrak{S}_U^{\perp o} := \{V \in \mathfrak{S} : V \perp U\}$, ie $\mathfrak{S}_U^{\perp o}$ is obtained from $\mathfrak{S}_U^{\perp}$ by removing W if $W \not\perp U$, where $W \sqsubsetneq S$ is the (arbitrarily chosen) $\sqsubseteq$–minimal "container" domain containing everything orthogonal to U.

Recall that $\mathfrak{S}_U^{\perp}$ consists of all domains $V \in \mathfrak{S}$ with $V \perp U$ along with a $\sqsubseteq$–minimal domain $W \in \mathfrak{S}$ such that $V \sqsubseteq W$ for all $V \perp U$. If W is the unique such domain, then $G \cdot W = W$, and thus G admits a natural restriction homomorphism to $\mathrm{Aut}(\mathfrak{S}_U^{\perp})$.

Otherwise, $W \notin \mathfrak{A}(G)$. Since $\mathrm{diam}_W(\pi_W(P_U)) \asymp 1$, we may replace W with a single point W^* such that $\mathcal{C}W^* = \{*\}$. From this we obtain a new HHS structure on $(E_U, \mathfrak{S}_U^{\perp o})$, where $\mathfrak{S}_U^{\perp o} = \mathfrak{S}_U^{\perp} - W \cup \{W^*\}$, by making the obvious alterations to the projection and domain maps associated to W.

In either case, let G_U be the image of G under the usual restriction homomorphism $\mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U)$. Let $G_U^{\perp}$ be the image of G under the restriction map $\psi\colon \mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U^{\perp})$ or, if W is not unique, we take $G_U^{\perp}$ be the image of $\psi\colon \mathcal{A}_U \to \mathrm{Aut}(\mathfrak{S}_U^{\perp o})$ defined as follows: for all $g \in \mathcal{A}_U$, the map $\psi(g)$ acts like g on $\mathfrak{S}_U^{\perp o}$ and acts as the identity on $\mathcal{C}W^*$.

Hence we have HHSs $(F_U, \mathfrak{S}_U)$ and $(E_U, \mathfrak{S}_U^{\perp})$, of complexity at most $\xi(\mathfrak{S}) - 1$, and groups $G_U \le \mathrm{Aut}(\mathfrak{S}_U)$ and $G_U^{\perp} \le \mathrm{Aut}(\mathfrak{S}_U^{\perp})$ or $\mathrm{Aut}(\mathfrak{S}_U^{\perp o})$ that satisfy the hypotheses of the theorem.

We now show that $\mathfrak{A}(G) = \mathfrak{A}(G_U) \sqcup \mathfrak{A}(G_U^{\perp})$. The inclusions $\mathfrak{A}(G_U), \mathfrak{A}(G_U^{\perp}) \to \mathfrak{A}(G)$ are obvious. Conversely, suppose that $V \in \mathfrak{A}(G)$. If $U \in \operatorname{Supp}(p)$ for some $p \in \operatorname{Fix}(G)$ (as we can assume is the case whenever $\operatorname{Fix}(G) \neq \varnothing$), then Lemma 9.22 implies that $V = U$ or $V \perp U$, ie $V \in \mathfrak{S}_U \sqcup \mathfrak{S}_U^{\perp}$ (and, if $V = W$, then W is the unique container and hence G–invariant). Otherwise, the proof of Lemma 9.22 shows that $V \perp U$ or $V \sqsubseteq U$. Hence $V \in \mathfrak{A}(G_U) \sqcup \mathfrak{A}(G_U^{\perp})$.

By induction on complexity, either $\mathfrak{A}(G_U) = \varnothing$, or there exists $\bar{h} \in G_U$ with $\operatorname{Big}(\bar{h}) = \mathfrak{A}(G_U)$. Likewise, either $\mathfrak{A}(G_U^{\perp}) = \varnothing$, or there exists $\bar{h}^{\perp} \in G_U^{\perp}$ with $\operatorname{Big}(\bar{h}^{\perp}) = \mathfrak{A}(G_U^{\perp})$. If $\mathfrak{A}(G_U) = \varnothing$ (resp. $\mathfrak{A}(G_U^{\perp}) = \varnothing$), we take $\bar{h} = 1$ (resp. $\bar{h}^{\perp} = 1$). Since $\mathfrak{A}(G) = \mathfrak{A}(G_U) \sqcup \mathfrak{A}(G_U^{\perp})$, we must use $\bar{h}$ and $\bar{h}^{\perp}$ to find $g \in G$ with $\operatorname{Big}(g) = \mathfrak{A}(G_U) \sqcup \mathfrak{A}(G_U^{\perp})$.

Choose $h, h^{\perp} \in G$ stabilizing $\mathfrak{S}_U$ and $\mathfrak{S}_U^{\perp}$ and mapping to $\bar{h} \in G_U$ and $\bar{h}^{\perp} \in G_U^{\perp}$, respectively, under the above restriction maps. Let k be the image of h in $G_U^{\perp}$ and let $k^{\perp}$ be the image of $h^{\perp}$ in G_U, so we are considering the action of h and $k^{\perp}$ on $\mathfrak{S}_U$ and $h^{\perp}$ and k on $\mathfrak{S}_U^{\perp}$.

Let $\{U_1, \dots, U_\ell\} = \operatorname{Big}(\bar{h}) \subset \mathfrak{S}_U$ and let $\{V_1, \dots, V_k\} = \operatorname{Big}(\bar{h}^{\perp}) \subset \mathfrak{S}_U^{\perp}$. By passing to powers, we can assume that $hU_i = U_i$ and $h^{\perp}V_j = V_j$ for all i and j. Since the action of G_U on $\mathfrak{S}_U$ preserves $\mathfrak{A}(G_U)$, and the action of $G_U^{\perp}$ on $\mathfrak{S}_U^{\perp}$ preserves $\mathfrak{A}(G_U^{\perp})$, we can, by passing to powers, assume that $k^{\perp}$ preserves each U_i and k preserves each V_j.

Let $N \gg 0$ and consider $F = \langle h^N, (h^{\perp})^{10N} \rangle \leq G$. The image of F in G_U is $\bar{F} = \langle \bar{h}^N, (k^{\perp})^{10N} \rangle$, and the image of F in $G_U^{\perp}$ is $\bar{F}^{\perp} = \langle k^N, (\bar{h}^{\perp})^{10N} \rangle$. The above discussion shows that $\bar{F}$ acts acylindrically on each $\mathcal{C}U_i$ and $\bar{F}^{\perp}$ acts acylindrically on each $\mathcal{C}V_j$. Examining the various cases that arise according to how k acts on the $\mathcal{C}V_i$ and how $k^{\perp}$ acts on the $\mathcal{C}U_i$ shows that, in each case, there exists $g \in F$ whose image in $\bar{F}$ is loxodromic on each $\mathcal{C}U_i$ and whose image in $\bar{F}^{\perp}$ is loxodromic on each V_j. Hence $\operatorname{Big}(g) = \mathfrak{A}(G_U) \sqcup \mathfrak{A}(G_U^{\perp})$, as required. □

The following is an immediate but useful corollary of Theorem 9.20:

Corollary 9.23 *If $G \leq \operatorname{Aut}(\mathfrak{S})$ is hierarchically acylindrical, then $\mathfrak{A}(G)$ is pairwise orthogonal.*

9.8 Rank-rigidity for some CAT(0) cube complexes

We now use Theorems 9.14 and 9.13 to reprove the rank-rigidity theorem of Caprace and Sageev [18], in the case where the cube complex in question contains a *factor system*. See Section 10 for a discussion of the definition, and the definition of the *simplicial boundary* $\partial_{\triangle}\mathcal{X}$ of the cube complex $\mathcal{X}$.

Corollary 9.24 (rank-rigidity for cube complexes with factor systems) *Let $\mathcal{X}$ be an unbounded* CAT(0) *cube complex with a factor system $\mathfrak{F}$. Let G act on $\mathcal{X}$ and suppose that one of the following holds:*

(1) *G acts on $\mathcal{X}$ properly and cocompactly.*

(2) *G acts on $\mathcal{X}$ with no fixed point in $\mathcal{X} \cup \partial_{\triangle}\mathcal{X}$.*

Then $\mathcal{X}$ contains a G–invariant convex subcomplex $\mathcal{Y}$ such that either G contains a rank-one isometry of $\mathcal{Y}$ or $\mathcal{Y} = \mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are unbounded convex subcomplexes.

We remark that in view of [36, Remark 5.3], we could have stated the corollary in terms of fixed points in the CAT(0) boundary rather than the simplicial boundary, but we have opted for the latter because of the close relationship between the simplicial and HHS boundaries discussed in Section 10.

Proof of Corollary 9.24 First suppose that G acts on $\mathcal{X}$ essentially, in the sense that every halfspace contains points of some G–orbit arbitrarily far from the associated hyperplane (in particular, $\mathcal{X}$ does not contain a G–invariant proper convex subcomplex). Recall from [5] that $\mathcal{X}$ is equipped with a hierarchically hyperbolic structure $(\mathcal{X}, \mathfrak{S})$, where $\mathfrak{S}$ is the set of *factored contact graphs* of elements of $\mathfrak{F}$, and that $G \leq \mathrm{Aut}(\mathfrak{S})$. If G acts on $\mathcal{X}$ properly and cocompactly, then $(G, \mathfrak{S})$ is an HHG; if G acts on $\mathcal{X}$ with no fixed point in $\partial_{\triangle}\mathcal{X}$, then G does not fix a point in $\partial(\mathcal{X}, \mathfrak{S})$, by Theorem 10.1 below.

Depending on which hypothesis we invoke, one of Theorem 9.14 or Theorem 9.13 implies that either there exists $g \in G$ which is rank-one (in the HHS sense) or there exists $U \in \mathfrak{S}$ such that $\mathcal{X}$ coarsely coincides with P_U, which has unbounded factors, and $G'U = U$ for some finite-index $G' \leq G$. In the former case, elements that are rank-one in the HHS sense (with respect to this particular HHS structure on $\mathcal{X}$) are rank-one isometries of $\mathcal{X}$ in the usual sense, by [36, Proposition 5.1] and the definition of a factor system [5, Section 8].

In the latter case, $P_U = F_U \times E_U$ is a genuine convex product subcomplex with unbounded factors (see [5]). Let $g \in G$ and suppose that H is a hyperplane intersecting P_U but not gP_U. Since P_U is coarsely equal to $\mathcal{X}$ and $\mathcal{X}$ is essential, the halfspace of P_U separated from gP_U by H contains points arbitrarily far from H, whence P_U and gP_U cannot lie at finite Hausdorff distance. This contradicts that P_U is invariant under a finite-index subgroup of G. Hence P_U and gP_U are *parallel* for all $g \in G$, ie they are crossed by exactly the same hyperplanes. Thus $\mathcal{X} = P_U \times Y$ for some compact cube complex Y, whence Y is a single point, by essentiality. It follows that

P_U is G–invariant, so $\mathcal{X} = P_U$ by essentiality. Hence $\mathcal{X}$ decomposes as a product with unbounded factors. In general, we first replace $\mathcal{X}$ by its G–essential core in either preceding argument, using Proposition 3.5 of [18]. □

Remark 9.25 Question A of [6] asks whether the existence of a proper cocompact action of G on the CAT(0) cube complex $\mathcal{X}$ ensures that $\mathcal{X}$ contains a factor system. By a result in [5], the answer is affirmative provided $\mathcal{X}$ embeds as a convex subcomplex in the universal cover of the Salvetti complex of some right-angled Artin group. Although it is a strong condition, we believe that such embeddings always exist (although there is in general no algebraic relationship between G and the RAAG).

9.8.1 The Poisson boundary of an HHG Results in [5] show that, if G is an HHG with $\operatorname{diam} \mathcal{C}S = \infty$, then, given a nonelementary probability measure μ on G, the boundary $\partial\mathcal{C}S$ admits a μ–stationary measure making it the Poisson boundary. As a topological model of the Poisson boundary, $\partial\mathcal{C}S$ is unsatisfactory since it need not be compact. However:

Theorem 9.26 (the HHS boundary is the Poisson boundary) *Let $(G, \mathfrak{S})$ be an HHG with $\operatorname{diam} \mathcal{C}S = \infty$, μ be a nonelementary probability measure on G with finite entropy and finite first logarithmic moment, and ν the resulting μ–stationary measure on ∂G. Then $(\partial G, \nu)$ is the Poisson boundary for (G, μ).*

We use acylindricity of the action of G on $\mathcal{C}S$ and a result of Maher and Tiozzo [54]:

Theorem 9.27 [54, Theorem 1.5] *Let G be a countable group which acts acylindrically on a separable Gromov-hyperbolic space X. If μ is a nonelementary probability measure on G with finite entropy and finite first logarithmic moment with corresponding stationary measure ν, then $(\partial X, \nu)$ is the Poisson boundary for (G, μ).*

Proof of Theorem 9.26 Let μ be a nonelementary probability measure on G with finite entropy and finite first logarithmic moment. Since G acts on $\mathcal{C}S$ acylindrically [5, Theorem 14.3], Theorem 9.27 implies that there exists a μ–stationary measure ν' on $\partial\mathcal{C}S$ such that $(\partial\mathcal{C}S, \nu')$ is the Poisson boundary for (G, μ).

Let $f\colon \partial\mathcal{C}S \hookrightarrow \partial G$ be the embedding from Proposition 2.13. By Lemma 9.7, $f(\partial\mathcal{C}S)$ is Borel, so for any Borel subset $V \subset \partial G$, the set $V \cap f(\partial\mathcal{C}S)$ is Borel. Define a new measure ν on ∂G by $\nu(V) = \nu'\big(f^{-1}(V \cap f(\partial\mathcal{C}S))\big)$.

Since f is G–equivariant, it follows that ν is μ–stationary. By definition, $f(\partial\mathcal{C}S)$ has full ν–measure. Moreover, $(\partial G, \nu)$ is a μ–boundary by measurability of f and it is maximal since $(\partial\mathcal{C}S, \nu')$ is maximal. Thus $(\partial G, \nu)$ models the Poisson boundary for (G, μ). □

10 Case study: CAT(0) cube complexes

Throughout this section, $\mathcal{X}$ is a locally finite CAT(0) cube complex in which each collection of pairwise-intersecting hyperplanes is (not necessarily uniformly) finite. In [5], it is shown that CAT(0) cube complexes can often be given HH structures using certain collections of convex subcomplexes called *factor systems*. We recall the definition in Section 10.2. When $\mathfrak{F}$ is a factor system for $\mathcal{X}$, denote the resulting HH structure by $(\mathcal{X}, \overline{\mathfrak{F}})$.

The *simplicial boundary* of $\mathcal{X}$ was introduced in [36]; we recall the definition below. The simplicial boundary and the HH structure are closely related by the following theorem:

Theorem 10.1 (simplicial and HHS boundaries) *Let $\mathcal{X}$ be a* CAT(0) *cube complex with a factor system $\mathfrak{F}$. There is a topology $\mathcal{T}$ on the simplicial boundary $\partial_\triangle \mathcal{X}$ such that:*

(1) *There is a homeomorphism $b\colon (\partial_\triangle \mathcal{X}, \mathcal{T}) \to \partial(\mathcal{X}, \overline{\mathfrak{F}})$.*

(2) *For each component C of the simplicial complex $\partial_\triangle \mathcal{X}$, the inclusion $C \hookrightarrow (\partial_\triangle \mathcal{X}, \mathcal{T})$ is an embedding.*

In particular, if $\mathfrak{F}$ and $\mathfrak{F}'$ are factor systems on $\mathcal{X}$, then $\partial(\mathcal{X}, \overline{\mathfrak{F}})$ is homeomorphic to $\partial(\mathcal{X}, \overline{\mathfrak{F}'})$.

We prove Theorem 10.1 in Section 10.3.

Remark 10.2 Proposition 3.37 of [36] relates $\partial_\triangle \mathcal{X}$ to its Tits boundary $\partial_T \mathcal{X}$. There is an analogous relationship between the HHS boundary and the visual boundary when the former is defined (ie when $\mathcal{X}$ has a factor system). Specifically, one can show that there is a commutative diagram

$$\begin{tikzcd} \partial_\triangle \mathcal{X} \arrow[r, "I"] \arrow[d, "b"'] & \partial_T \mathcal{X} \arrow[d, "\mathrm{id}"] \\ \partial(\mathcal{X}, \overline{\mathfrak{F}}) \arrow[r, "J"] & \partial_{\mathrm{vis}} \mathcal{X} \end{tikzcd}$$

where b is the bijection from Theorem 10.1, I and J are embeddings, J is $\frac{\pi}{2}$–quasisurjective, and $\partial(\mathcal{X}, \overline{\mathfrak{F}})$ is a deformation retract of $\partial_{\mathrm{vis}} \mathcal{X}$. The CAT(0) metric on $\mathcal{X}$ is far afield from our present discussion, since the HHS structure depends only on the combinatorics of $\mathcal{X}$ and is insensitive to changes in the CAT(0) metric (unlike the visual boundary [22]), so we will not give a detailed proof of the above. The top part of the diagram comes from [36, Proposition 3.37]; the missing ingredient is to shown that J is an embedding, which is a tedious exercise in the definition of the topology on $\partial(\mathcal{X},\overline{\mathfrak{F}})$.

10.1 The simplicial boundary

We first recall the necessary definitions from [36].

Definition 10.3 (UBS, boundary equivalence, minimal UBS) A set $\mathcal{U}$ of hyperplanes in $\mathcal{X}$ is a *unidirectional boundary set (UBS)* if each of the following holds:

- $\mathcal{U}$ is infinite.
- If $U, U' \in \mathcal{U}$ and a hyperplane V separates U and U', then $V \in \mathcal{U}$.
- If $U, U', U'' \in \mathcal{U}$ are pairwise disjoint, then one of them separates the other two.
- For all hyperplanes W, at least one component of $\mathcal{X} - W$ contains at most finitely many elements of $\mathcal{U}$.

Given UBSs $\mathcal{U}$ and $\mathcal{V}$, let $\mathcal{U} \preceq \mathcal{V}$ if all but finitely many elements of $\mathcal{U}$ lie in $\mathcal{V}$. The UBSs $\mathcal{U}$ and $\mathcal{V}$ are *boundary equivalent* if $\mathcal{U} \preceq \mathcal{V}$ and $\mathcal{V} \preceq \mathcal{U}$, and $\mathcal{U}$ is *minimal* if $\mathcal{U}$ and $\mathcal{V}$ are boundary equivalent for all UBSs $\mathcal{V}$ with $\mathcal{V} \preceq \mathcal{U}$.

Remark 10.4 Any infinite set of hyperplanes which is closed under separation contains a minimal UBS [36, Lemma 3.7].

Proposition 3.10 of [36] shows that each UBS $\mathcal{U}$ is boundary equivalent to a UBS of the form $\bigsqcup_{i=0}^{k} \mathcal{U}_i$, where each $\mathcal{U}_i$ is a minimal UBS, and this decomposition is unique up to boundary equivalence. Up to reordering, for $0 \le i < j \le k$, for all but finitely many $U \in \mathcal{U}_j$, the hyperplane U intersects all but finitely many elements of $\mathcal{U}_i$. In this situation, $\mathcal{U}_j$ *dominates* $\mathcal{U}_i$. The number k is the *dimension* of $\mathcal{U}$.

Definition 10.5 (simplicial boundary) A *k–simplex at infinity* is a boundary equivalence class of k–dimensional UBSs. If v and v' are simplices at infinity, represented by boundary sets $\mathcal{V}$ and $\mathcal{V}'$, then $\mathcal{V} \cap \mathcal{V}'$ is, if infinite, a boundary set representing the simplex $v \cap v'$. The *simplicial boundary* $\partial_\triangle \mathcal{X}$ of $\mathcal{X}$ is the simplicial complex with a closed k–simplex for each k–dimensional simplex at infinity; the simplex u represented by the UBS $\mathcal{U}$ is a face of the simplex v, represented by $\mathcal{V}$, if $\mathcal{U} \preceq \mathcal{V}$.

Remark 10.6 (boundaries of convex subcomplexes) It is shown in [36] that if $\mathcal{Y} \subseteq \mathcal{X}$ is a convex subcomplex, then $\partial_\triangle \mathcal{Y} \subset \partial_\triangle \mathcal{X}$ in a natural way: each simplex at infinity in $\partial_\triangle \mathcal{Y}$ corresponds to a UBS in $\mathcal{X}$ consisting of hyperplanes that intersect $\mathcal{Y}$, and these hyperplanes intersect in $\mathcal{X}$ exactly when they intersect in $\mathcal{Y}$, by convexity.

10.1.1 Visibility

Definition 10.7 (visible simplex) The simplex u at infinity is *visible* if there exists a combinatorial geodesic ray γ in $\mathcal{X}^{(1)}$ such that the set $\mathcal{U}$ of hyperplanes intersecting γ represents the boundary–equivalence class u. Otherwise, the simplex u at infinity is *invisible*. If every simplex at infinity is visible, then $\mathcal{X}$ is *fully visible*.

Theorem 3.19 of [36] states that every maximal simplex of $\partial_\triangle \mathcal{X}$ is visible. Visibility is also related to a subtlety in the definition of $\partial_\triangle \mathcal{X}$:

Remark 10.8 (visibility and proper faces) Let $\bigsqcup_{i=0}^k \mathcal{U}_i$ be a UBS, with each $\mathcal{U}_i$ a minimal UBS, numbered so that for $0 \le i < j \le k$ and all $U \in \mathcal{U}_j$, we have that $U \cap V \neq \varnothing$ for all but finitely many $V \in \mathcal{U}_i$. If, up to modifying each $\mathcal{U}_i$ in its boundary equivalence class, $U \cap V \neq \varnothing$ whenever $U \in \mathcal{U}_i$, $V \in \mathcal{V}_j$, and $i \neq j$, then the simplex u represented by $\bigsqcup_{i=0}^k \mathcal{U}_i$ is visible. In this case, $\mathcal{X}$ contains an isometrically embedded (on the 1–skeleton) cubical orthant, the boundary of whose convex hull is u. Conversely, if we know that each $\mathcal{U}_i$ represents a visible 0–simplex, then $\bigsqcup_{i \in K} \mathcal{U}_i$ represents a visible simplex at infinity for any $K \subset \{0, \dots, k\}$, as is proved in [36]. If this does not occur, then there may be subsets $K \subset \{0, \dots, k\}$ such that $\bigsqcup_{i \in K} \mathcal{U}_i$ represents an invisible simplex at infinity, or is not even a UBS (by virtue of failing to satisfy the condition on separation). In other words, when $\mathcal{X}$ is not fully visible, simplices at infinity may have *proper faces* that are not genuine simplices at infinity represented by UBSs.

A visible simplex $v \subseteq \partial_\triangle \mathcal{X}$ is *represented* by the combinatorial geodesic ray $\gamma \subseteq \mathcal{X}^{(1)}$ if the UBS of hyperplanes intersecting γ represents the boundary equivalence class v.

Remark 10.9 (factor systems and visibility) Conjecture 2.8 of [4] states that if $\mathcal{X}$ is a CAT(0) cube complex on which some group acts geometrically, then $\mathcal{X}$ is fully visible. Also, the proof of Theorem 10.1 shows that, if $\mathcal{X}$ contains a factor system (see Definition 10.10), then every simplex of $\partial_\triangle \mathcal{X}$ is visible. This is another reason for interest in Question A of [6], which asks whether every CAT(0) cube complex on which some group acts geometrically contains a factor system.

10.2 Factor systems: hierarchical hyperbolicity of cube complexes

We now summarize results from [5] yielding hierarchically hyperbolic structures on $\mathcal{X}$. We refer the reader to Section 2 of [5] for discussion of convex subcomplexes and the gate map $\mathfrak{g}_F \colon \mathcal{X} \to F$ from $\mathcal{X}$ to any convex subcomplex F.

Recall that each hyperplane H of $\mathcal{X}$ lies in a *carrier*, $\mathcal{N}(H)$, which is the union of closed cubes intersecting H. For all H, there is a cubical isomorphism $\mathcal{N}(H) \cong H \times \left[-\frac{1}{2}, \frac{1}{2}\right]$; a subcomplex of $\mathcal{X}$ which is the image under the inclusion $\mathcal{N}(H) \to \mathcal{X}$

of either of the subcomplexes $H \times \{\frac{1}{2}\}$ or $H \times \{-\frac{1}{2}\}$ is a *combinatorial hyperplane*. We say that two convex subcomplexes F and F' of $\mathcal{X}$ are *parallel* if for any hyperplane H of $\mathcal{X}$, we have $H \cap F \neq \varnothing$ if and only if $H \cap F' \neq \varnothing$. We let $\overline{\mathfrak{F}}$ denote a choice of representatives for each parallelism class of elements of $\mathfrak{F}$.

Definition 10.10 A *factor system* $\mathfrak{F}$ is a set of convex subcomplexes such that:

(1) Each nontrivial combinatorial hyperplane of $\mathcal{X}$ belongs to $\mathfrak{F}$, as does each convex subcomplex parallel to a nontrivial combinatorial hyperplane.

(2) $\mathcal{X} \in \mathfrak{F}$.

(3) There exists $\xi > 0$ such that, for all $F, F' \in \mathfrak{F}$, either

$$\mathfrak{g}_F(F') \in \mathfrak{F} \quad \text{or} \quad \operatorname{diam}(\mathfrak{g}_F(F')) \leq \xi.$$

(4) There exists $\Delta \geq 1$ such that each point in $\mathcal{X}$ belongs to at most Δ elements of $\mathfrak{F}$.

We require that elements of $\mathfrak{F}$ are not single points. (This condition is only imposed to ensure that nesting and orthogonality are mutually exclusive: if F is a single point and $F' \in \mathfrak{F}$, then $F \perp F'$ and $F \sqsubseteq F'$, so we exclude this situation.)

The *contact graph* $\mathcal{C}\mathcal{X}$ of $\mathcal{X}$ (see [37]) has a vertex for each hyperplane, with two hyperplanes joined by an edge if no third hyperplane separates them. If $F \subseteq \mathcal{X}$ is a convex subcomplex, then F is a CAT(0) cube complex whose hyperplanes have the form $H \cap F$, where H is a hyperplane of $\mathcal{X}$, and, by convexity of F, this yields an embedding $\mathcal{C}F \hookrightarrow \mathcal{C}\mathcal{X}$ of $\mathcal{F}$ as a full subgraph.

Given a factor system $\mathfrak{F}$ on $\mathcal{X}$, we define the *factored contact graph* $\widehat{\mathcal{C}}F$ of each $F \in \mathfrak{F}$ as follows. Begin with $\mathcal{C}F$. For each parallelism class of subcomplexes $F' \in \mathfrak{F}$, parallel to a proper subcomplex of F that is not a single 0–cube, we have $\mathcal{C}F' \subsetneq \mathcal{C}F$, and we cone off $\mathcal{C}F'$ by adding a vertex $v_{F'}$ to $\mathcal{C}F$ and joining each vertex of $\mathcal{C}F' \subset \mathcal{C}F$ to $v_{F'}$. The resulting factored contact graph $\widehat{\mathcal{C}}F$ is uniformly quasiisometric to a tree [5, Proposition 8.24].

Let us now define the maps $\pi_F\colon \mathcal{X} \to 2^{\widehat{\mathcal{C}}F}$. For each $F \in \mathfrak{F}$, given $x \in \mathcal{X}$, let $\mathfrak{g}_F(x) \in F$ be its gate. There is a nonempty finite set of hyperplanes H of F that are not separated from x by any other hyperplane; these form a nonempty clique in $\mathcal{C}F$, to which we send x. We then compose with $2^{\mathcal{C}F} \hookrightarrow 2^{\widehat{\mathcal{C}}F}$ to yield $\pi_F\colon \mathcal{X} \to 2^{\widehat{\mathcal{C}}F}$ sending each point to a clique.

Let $F \sqsubseteq F'$ if F is parallel to a subcomplex of F', and $F \perp F'$ if there is a cubical isometric embedding $F \times F' \to \mathcal{X}$ (after possibly varying F and F' in their parallelism classes). Otherwise, F and F' are transverse. With these definitions, it is shown in [5; 6] that $(\mathcal{X}, \overline{\mathfrak{F}})$ is a hierarchically hyperbolic space.

10.3 Relating the simplicial and HHS boundaries

Fix $\mathcal{X}$ with a factor system $\mathfrak{F}$; necessarily, $\mathcal{X}$ is uniformly locally finite.

Proof of Theorem 10.1 We will first exhibit a bijection $b\colon \partial_\triangle \mathcal{X} \to \partial(\mathcal{X}, \overline{\mathfrak{F}})$. We then define $\mathcal{T} = \{b^{-1}(\mathcal{O})\}$, where $\mathcal{O}$ varies over all open sets in $\partial(\mathcal{X}, \overline{\mathfrak{F}})$, so as to make b a homeomorphism. It then suffices to verify that this topology agrees with the simplicial topology on each component of $\partial_\triangle \mathcal{X}$; the "in particular" statement then follows immediately.

Reduction to the single-simplex case Let m be a maximal simplex of $\partial_\triangle \mathcal{X}$. By the definition of the simplicial boundary, m is a simplex at infinity, ie it is represented by some UBS $\mathcal{M}$. Moreover, by [36, Theorem 3.19], we can take $\mathcal{M}$ to be the set of hyperplanes intersecting some combinatorial geodesic ray γ_m emanating from the (fixed) basepoint x_0. Let $\mathcal{Y}_m$ be the convex hull of γ_m.

By [5, Lemma 8.4], $\mathfrak{F}_m = \{F \cap \mathcal{Y}_m : F \in \mathfrak{F}\}$ is a factor system. (We emphasize that $\mathfrak{F}_m$ is a set, not a multiset: if $F, F' \in \mathfrak{F}$ satisfy $F \cap \mathcal{Y}_m = F' \cap \mathcal{Y}_m$, we count this subcomplex once.) We adopt the following convention: for each $F \cap \mathcal{Y}_m \in \mathfrak{F}_m$, we assume that F has been chosen so that F is $\sqsubseteq$–minimal among all $F' \in \mathfrak{F}$ with $F' \cap \mathcal{Y}_m = F \cap \mathcal{Y}_m$. (Note that there is a unique such minimal F: if $F \cap \mathcal{Y}_m = F' \cap \mathcal{Y}_m$, then $F \cap \mathcal{Y}_m = F \cap F' \cap \mathcal{Y}_m$, and $F \cap F' \sqsubseteq F, F'$.)

Also, if $F \sqsubseteq F'$, then $F \cap \mathcal{Y}_m \sqsubseteq F' \cap \mathcal{Y}_m$, obviously. Conversely, suppose that $F \cap \mathcal{Y}_m \sqsubseteq F' \cap \mathcal{Y}_m$. Let $F'' = \mathfrak{g}_F(F')$, so $F'' \sqsubseteq F'$ and $F'' \sqsubseteq F$. Then $F'' \cap \mathcal{Y}_m = F \cap \mathcal{Y}_m$, so $F'' = F$ by minimality, whence $F \sqsubseteq F'$.

If $F \perp F'$, then convexity of $\mathcal{Y}_m$ implies $(F \times \perp F') \cap \mathcal{Y}_m = (F \cap \mathcal{Y}_m) \times (F' \cap \mathcal{Y}_m)$, so $(F \cap \mathcal{Y}_m) \perp (F' \cap \mathcal{Y}_m)$. Conversely, suppose that $(F \cap \mathcal{Y}_m) \perp (F' \cap \mathcal{Y}_m)$. For brevity, let $A = F \cap \mathcal{Y}_m$ and $B = F' \cap \mathcal{Y}_m$, so that $\mathcal{X}$ contains $A \times B$. By Lemma 10.13, there exist $F_A, F_B \in \mathfrak{F}$ such that $A \subset F_A$, $B \subset F_B$ and $F_A \perp F_B$. Let $F'_A = F \cap F_A$ and $F'_B = F' \cap F_B$. Then $F'_A \cap \mathcal{Y}_m = F \cap \mathcal{Y}_m$ and $F'_A \sqsubseteq F$, so minimality of F implies $F'_A = F$; similarly $F'_B = F'$. But since $F_A \perp F_B$ and $F'_A \sqsubseteq F_A$ and $F'_B \sqsubseteq F_B$, we have $F \perp F'$.

It follows that there is a hieromorphism $(\mathcal{Y}_m, \overline{\mathfrak{F}}_m) \to (\mathcal{X}, \overline{\mathfrak{F}})$ defined as follows: the map $\mathcal{Y}_m \to \mathcal{X}$ is the inclusion; the map $\mathfrak{F}_m \to \mathfrak{F}$ is given by $F \cap \mathcal{Y}_m \mapsto F$ for each $F \cap \mathcal{Y}_m \in \mathfrak{F}_m$ (where F is $\sqsubseteq$–minimal in $\mathfrak{F}$ with the given intersection with $\mathcal{Y}$), and for each $F \cap \mathcal{Y}_m$, the map $\widehat{\mathcal{C}}(F \cap \mathcal{Y}_m) \to \widehat{\mathcal{C}}F$ is the inclusion on contact graphs and sends cone vertices to cone vertices in the obvious way.

We will see below that $\mathcal{Y}_m = \prod_{i=0}^{k} \mathcal{Y}_{m_i}$, where each $\mathcal{Y}_{m_i}$ has the property that $\partial\widehat{\mathcal{C}}(F \cap \mathcal{Y}_{m_i}) = \varnothing$ for all $F \in \mathfrak{F}$ except for a unique $\ddot{F}_i \in \mathfrak{F}$ for which $\partial\widehat{\mathcal{C}}(\ddot{F}_i \cap \mathcal{Y}_{m_i})$

consists of a single point p_i. Moreover, $\ddot{F}_i \perp \ddot{F}_j$ for $i \neq j$. Lemma 10.11 shows that for each $F \cap \mathcal{Y}_m$, the map $\widehat{\mathcal{C}}(F \cap \mathcal{Y}_m) \hookrightarrow \widehat{\mathcal{C}}F$ is a uniform quasiisometric embedding, inducing a boundary map, ie p_i may be regarded as a point in $\partial\widehat{\mathcal{C}}\ddot{F}_i$ for each i. We thus obtain an injective map $b_m\colon \partial(\mathcal{Y}_m, \overline{\mathfrak{F}}_m) \to \partial(\mathcal{X}, \overline{\mathfrak{F}})$ given by

$$b_m\left(\sum_{i=0}^{k} a_i m_i\right) = \sum_{i=0}^{k} a_i p_i.$$

Constructing *b* We will observe below that if m, m' are maximal simplices, then the associated collections $\{p_i\}_{i=0}^{k}$ and $\{p_i'\}_{i=0}^{k'}$ intersect in a set corresponding precisely to the set of 0–simplices of $m \cap m'$. It follows that the maps constructed above are *compatible*, ie $b_m|_{\mathcal{Y}_{m\cap m'}} = b_{m'}|_{\mathcal{Y}_{m\cap m'}}$ and that, if m and m' are disjoint maximal simplices of $\partial_\triangle\mathcal{X}$, then b_m and $b_{m'}$ have disjoint images. Pasting together the b_m thus yields an injection $b\colon \partial_\triangle\mathcal{X} \to \partial(\mathcal{X}, \overline{\mathfrak{F}})$.

Surjectivity of *b* Let $\{\ddot{F}_i\}_{i=1}^{k}$ be a support set in $\overline{\mathfrak{F}}$, choose for each i a point $p_i \in \partial\widehat{\mathcal{C}}\ddot{F}_i$ and let $p = \sum_{i=1}^{k} a_i p_i$. For each i, let σ_i be a geodesic ray in the quasitree $\widehat{\mathcal{C}}\ddot{F}_i$ joining $\pi_{\widehat{\mathcal{C}}\ddot{F}_i}(x_0)$ to p_i. Let $\{H_n^i\}$ be a sequence of hyperplanes of $\mathcal{X}$, each crossing $\ddot{F}_i$, corresponding to vertices of σ_i, ordered so that H_n^i separates H_{n+1}^i from x_0. Any $P \in \mathfrak{F}$ that crosses infinitely many of these hyperplanes satisfies $\ddot{F}_i \sqsubseteq P$, or else some element of $\mathfrak{F}$ nested into $\ddot{F}_i$ would "kill" the p_i direction in $\partial\widehat{\mathcal{C}}\ddot{F}_i$. Every simplex of $\partial_\triangle\big(\prod_{j=0}^{k} \ddot{F}_j\big) \subset \partial_\triangle\mathcal{X}$ is visible, from which it is easy to check that there is a unique (up to boundary–equivalence) minimal UBS $\mathcal{M}_i$ containing $\{H_n^i\}$ and representing a 0–simplex m_i of $\partial_\triangle\mathcal{X}$ such that $\{m_0, \dots, m_k\}$ span a simplex m. By definition, $b_m\big(\sum_i a_i m_i\big) = p$.

Analysis of components To prove that each component C of $\partial_\triangle\mathcal{X}$, with the simplicial topology, is embedded in $(\partial_\triangle\mathcal{X}, \mathcal{T})$, we must show that $b\circ\mathrm{id}\colon \partial_\triangle\mathcal{X} \to \partial(\mathcal{X}, \overline{\mathfrak{F}})$ restricts to an embedding on C, where $\mathrm{id}\colon \partial_\triangle\mathcal{X} \to (\partial_\triangle\mathcal{X}, \mathcal{T})$ is the identity. Let m be a maximal simplex of $\partial_\triangle\mathcal{X}$. Let $p = \sum_i a_i p_i \in b\circ\mathrm{id}(M)$ and let $\mathcal{N} = \mathcal{N}_{\{U_i\},\epsilon}(p) \cap \partial(\mathcal{Y}_m, \overline{\mathfrak{F}}_m)$ be a basic neighborhood of p, as defined in Section 1.1. Observe that $\mathcal{N}$ is completely nonremote, whence it is clear from the definition that $b_m^{-1}(\mathcal{N})$ is basic in the simplicial topology on $\partial_\triangle\mathcal{Y}_m = m$, so b_m is continuous. It follows that $b\circ\mathrm{id}$ is continuous. A similar argument shows that the restriction of $b\circ\mathrm{id}$ to C is an open map. To complete the proof, it now suffices to produce the F_i and analyze their factored contact graphs, which we do in the next several steps.

Visibility of faces of *m* Let m be a maximal simplex of $\partial_\triangle\mathcal{X}$ and observe that $\partial_\triangle\mathcal{Y}_m$ is exactly the simplex m. We now verify that each face of m is a visible simplex at infinity. Let $m_0, \dots, m_k$ be the 0–simplices of m; represent m_i by a minimal UBS $\mathcal{M}_i$ such that $\mathcal{M}_j$ dominates $\mathcal{M}_i$ when $i < j$ and $\mathcal{M} = \bigsqcup_{i=0}^{k} \mathcal{M}_i$. Recall from Remark 10.8 that if $\mathcal{M}_i$ dominates $\mathcal{M}_j$ for all i and j, then each subsimplex of m is visible.

By projecting γ_m to a combinatorial hyperplane on the carrier of some element of $\mathcal{M}_k$, we see that $\mathcal{M} - \mathcal{M}_k$ represents a visible codimension-1 face m' of m, represented by a ray $\gamma_{m'}$. The convex hull $\mathcal{Y}_{m'} \subset \mathcal{Y}_m$ of $\gamma_{m'}$ inherits a factor system from $\mathcal{Y}_m$ as above. Hence, by induction, for $i < k$, the 0–simplex represented by $\mathcal{M}_i$ is visible. Thus it suffices to show that the 0–simplex m_k represented by $\mathcal{M}_k$ is visible. (In the base case, m is a maximal 0–simplex, and is visible by maximality.) Suppose, for a contradiction, that m_k is not visible, so there exists $i < k$ such that $\mathcal{M}_i$ fails to dominate $\mathcal{M}_k$. In particular, $k \geq 1$.

The UBS $\mathcal{M}_k$ contains a sequence $\{M_n\}_{n\geq 0}$ of pairwise-disjoint hyperplanes such that M_n separates $M_{n\pm 1}$ for all $n \geq 1$. For each n, let M_n^+ be the combinatorial hyperplane in $\mathcal{N}(M_n)$ in the same component of $\mathcal{X} - M_n$ as M_{n+1}. For each n, let $P_n = \mathfrak{g}_{M_0^+}(M_n^+)$ be the projection of M_n^+ on M_0^+. The set of hyperplanes crossed by both M_0 and M_n contains all but finitely many elements of $\mathcal{M}_i$; hence each P_n is unbounded and thus belongs to the factor system $\mathfrak{F}_m$. Moreover, for all $N \geq 0$, the intersection $\bigcap_{n=0}^{N} P_n \neq \varnothing$. Hence, since $\mathfrak{P}_m$ has multiplicity $\Delta < \infty$, it must be the case that there exists N such that $P_n = P_N$ for all $N \geq n$. Thus, when $n, n' \geq N$, the set of elements of $\mathcal{M}_j$ crossed by M_n coincides with the set crossed by $M_{n'}$ for all $j \leq k-1$. Hence each $\mathcal{M}_j$ dominates $\mathcal{M}_k$, whence m_k is visible.

Structure of $\mathcal{Y}_m$ By [36, Theorem 3.23] and visibility of the m_i established above, after moving x_0 if necessary, $\mathcal{Y}_m = \prod_{i=0}^{k} \mathcal{Y}_{m_i}$, where $\mathcal{Y}_{m_i}$ is the convex hull in $\mathcal{X}$ of a combinatorial geodesic ray γ^i at the basepoint x_0 representing a 0–simplex m_i of m. Each point of $m = \partial_\triangle \mathcal{Y}_m$ can be uniquely written as $\sum_{i=0} a_i m_i$, where $a_i \geq 0$ and $\sum_{i=1}^{k} a_i = 1$.

For each i, let $\{H_n^i\}_{n\geq 0}$ be the set of hyperplanes crossing γ^i; this is a minimal UBS and is numbered according to the order in which γ^i crosses the H_n^i. Thus, if $n > m$, the hyperplane H_n^i does not separate H_m^i from x_0 (in fact, either $H_n^i \cap H_m^i \neq \varnothing$ or H_m^i separates H_n^i from x_0). Choose $F_i \in \overline{\mathfrak{F}}_m$ to be $\sqsubseteq$–minimal such that all but finitely many H_n^i cross F_i. Observe that $F_i \perp F_j$ for all $i \neq j$, and that $F_i \subseteq \mathcal{Y}_{m_i}$.

Suppose that m' is some other maximal simplex and $\mathcal{Y}_{m'} = \prod_{i=0}^{k'} \mathcal{Y}_{m'_i}$. For each i, let $F'_i \in \overline{\mathfrak{F}}_{m'}$ be $\sqsubseteq$–minimal among those factors crossing all but finitely many of the elements crossing $\mathcal{Y}_{m_i}$. Suppose that $\partial\widehat{\mathcal{C}}F_i = \partial\widehat{\mathcal{C}}F'_j$ for some $i \leq k$ and $j \leq k'$. Then the set of hyperplanes crossing $\mathcal{Y}_{m_i}$, which is boundary-equivalent to that crossing F_i, is boundary-equivalent to that crossing F'_j and hence that crossing $\mathcal{Y}_{m'_j}$, ie $m_i = m'_j$.

Orthogonality Each F_i has the form $F_i = \widehat{F}_i \cap \mathcal{Y}_m$, where $\widehat{F}_i \in \overline{\mathfrak{F}}$. While orthogonality of elements of $\overline{\mathfrak{F}}$ implies orthogonality of the corresponding elements of $\overline{\mathfrak{F}}_m$, the converse need not hold, but we will require that $\widehat{F}_i \perp \widehat{F}_j$ for all $i \neq j$, in order to construct points of $\partial(\mathcal{X},\overline{\mathfrak{F}})$. However, finitely many applications of Lemma 10.13 below show that for each i, there exists $\ddot{F}_i \in \overline{\mathfrak{F}}$ such that $F_i \subseteq \ddot{F}_i \subseteq \widehat{F}_i$ and such that $\ddot{F}_i \perp \ddot{F}_j$ for all $i \neq j$.

Factored contact graphs in $\overline{\mathfrak{F}}_m$ For any $F \in \overline{\mathfrak{F}}_m$, we have, by convexity and [18, Proposition 2.5], that $F = \prod_{i=0}^k \mathfrak{g}_{\mathcal{Y}_{m_i}}(F)$, whence $\mathcal{C}F$ decomposes as a join, so $\widehat{\mathcal{C}}F$ is obtained from a join by coning off certain subgraphs. Thus $\widehat{\mathcal{C}}F$ is bounded (and $\partial\widehat{\mathcal{C}}F = \varnothing$) unless F is parallel to a subcomplex of some $\mathcal{Y}_{m_i}$. We claim that $\partial\widehat{\mathcal{C}}F_i$ consists of exactly one point p_i for each i, and that, for all other $F \in \overline{\mathfrak{F}}_m$, we have $\partial\widehat{\mathcal{C}}F = \varnothing$.

Observe that $\mathcal{C}F_i$ coarsely coincides with $\mathcal{C}\mathcal{Y}_i$, the $\{H_n^i\}$ are partially ordered by the order in which γ_i crosses them, and that $\mathcal{C}F_i$ is coarsely equal to a maximal chain in this partial order (ie a combinatorial ray σ in $\mathcal{C}F_i$). By Theorem 2.4 of [36], σ is unbounded in $\mathcal{C}F_i$, since F_i is $\sqsubseteq$–minimal, and thus determines a point $p_i \in \partial\mathcal{C}F_i$. Moreover, p_i is unique, since $\widehat{\mathcal{C}}F_i$ lies in the 1–neighborhood of σ ($\widehat{\mathcal{C}}F_i$ is obtained from σ by adding edges reflecting intersections of elements of the $\{H_n^i\}$).

Hence, if $\sigma \subset \widehat{\mathcal{C}}F_i$ is unbounded, then $\partial\widehat{\mathcal{C}}F_i = \{p_i\}$. By $\sqsubseteq$–minimality of F_i, no hyperplane of F_i crosses infinitely many $\{H_n^i\}$, so hyperplanes of F_i are compact. By minimality of the UBS $\{H_n^i\}$, any element of $\overline{\mathfrak{F}}_m$ corresponding to a cone-vertex in $\widehat{\mathcal{C}}F_i$ crosses finitely many hyperplanes. It follows that for all $n \geq 0$, there exists $N \geq n$ such that H_n^i and H_m^i cannot be adjacent to the same cone-vertex of $\widehat{\mathcal{C}}F_i$ when $m \geq N$. Hence $\partial\widehat{\mathcal{C}}F_i = \{p_i\}$.

We have shown that if $F \in \overline{\mathfrak{F}}_m$ has unbounded factored contact graph, then F is (up to parallelism) contained in some $\mathcal{Y}_{m_i}$. If F intersects only finitely many elements of $\{H_i\}$, then F is compact and thus $\widehat{\mathcal{C}}F$ is bounded. If F intersects infinitely many, then it intersects all but finitely many, whence either F is parallel to F_i or $\widehat{\mathcal{C}}F$ contains a subgraph, containing all but finitely many hyperplane-vertices, whose vertices are all adjacent to the cone-point corresponding to $\mathfrak{g}_F(F_i)$; thus $\widehat{\mathcal{C}}F$ is bounded. This completes the description of the boundaries of the factored contact graphs of the elements of $\overline{\mathfrak{F}}_m$. □

Lemma 10.11 *Let $\mathfrak{F}$ be a factor system in $\mathcal{X}$, let $\mathcal{Y} \subseteq \mathcal{X}$ be a convex subcomplex, and let $\mathfrak{F}'$ be the factor system in $\mathcal{Y}$ consisting of all subcomplexes of the form $F' \cap \mathcal{Y}$, where $F' \in \mathfrak{F}$. Let $F \cap \mathcal{Y} \in \mathfrak{F}'$, and suppose that if $F' \in \mathfrak{F}$ satisfies $F' \cap \mathcal{Y} = F \cap \mathcal{Y}$, then $F \sqsubseteq F'$.*

Then the following map $\phi\colon \widehat{\mathcal{C}}(F \cap \mathcal{Y}) \to \widehat{\mathcal{C}}F$ is a $(3,0)$–quasiisometric embedding: ϕ is the inclusion on contact graphs; for each $F' \cap \mathcal{Y} \in \mathfrak{F}'$ properly nested in $F \cap \mathcal{Y}$ (with F' minimal with this intersection with $\mathcal{Y}$), the cone-point in $\widehat{\mathcal{C}}(F \cap \mathcal{Y})$ corresponding to $F' \cap \mathcal{Y}$ is sent to the cone-point of $\widehat{\mathcal{C}}\mathcal{X}$ corresponding to F'.

Remark 10.12 Recall from the discussion in the proof of Theorem 10.1 of the hieromorphism $(\mathcal{Y}_m, \overline{\mathfrak{F}}_m) \to (\mathcal{X}, \overline{\mathfrak{F}})$ that if $F' \cap \mathcal{Y} \sqsubseteq F \cap \mathcal{Y}$ and F and F' are each $\sqsubseteq$–minimal with the given intersections with $\mathcal{Y}$, then $F \sqsubseteq F'$.

Proof of Lemma 10.11 Let v and v' be vertices of $\widehat{\mathcal{C}}(F \cap \mathcal{Y}_m)$. Let $v = v_0, v_1, \dots,$ $v_n = v'$ be a geodesic sequence in $\widehat{\mathcal{C}}F$ from v to v'. If v_i is a hyperplane vertex, let H_i be the corresponding hyperplane of F (so H crosses $F \cap \mathcal{Y}$). If v_i is a cone-vertex, let H_i be a subcomplex in $\mathfrak{F}$, properly contained in F, that represents the parallelism class corresponding to the cone-vertex v_i. (For $i \in \{0, n\}$, if H_i is a hyperplane, then it crosses $\mathcal{Y}$. Otherwise, $H_i \in \mathfrak{F}$ is $\sqsubseteq$–minimal among all $U \in \mathfrak{F}_F$ with $U \cap \mathcal{Y} = H_i \cap \mathcal{Y}$.)

If H_i is a cone-vertex, then $H_{i\pm1}$ are hyperplanes crossing H_i. This gives a sequence $H_0, H_1, \dots, H_n$ of hyperplanes or factor-system elements in F such that $\mathcal{N}(H_i) \cap \mathcal{N}(H_{i+1}) \neq \varnothing$ when H_i and H_{i+1} are hyperplanes, and $H_i \cap H_{i+1} \neq \varnothing$ when H_{i+1} is a subcomplex in $\mathfrak{F}$.

For each i such that $H_i \in \mathfrak{F}$, we have $H_i \subsetneq F$. In particular, our minimality assumption on F ensures that if $H_i \cap \mathcal{Y} \neq \varnothing$, then $H_i \cap \mathcal{Y} \subsetneq F \cap \mathcal{Y}$. Otherwise, we would have $H_i \cap \mathcal{Y} = F \cap \mathcal{Y}$ while $H_i \subsetneq F$, contradicting minimality of F.

For each i with H_i a hyperplane, choose a combinatorial geodesic $\gamma_i \to \mathcal{N}(H_i)$ joining the terminal point of $\gamma_{i=1}$ to a closest point on H_{i+1} (or $\mathcal{N}(H_{i+1})$ if v_{i+1} is a hyperplane vertex). Similarly, choose $\gamma_i \to H_i$ when v_i is a cone-vertex. The geodesic $\gamma_1 \to H_1$ joins $H_1 \cap \mathcal{Y}$ (or $\mathcal{N}(H_1) \cap \mathcal{Y}$ to $H_1 \cap H_2$, or $\mathcal{N}(H_1) \cap H_2$ etc), and $\gamma_n \to H_n$ (or $\mathcal{N}(H_n)$) is similarly chosen to end in $\mathcal{Y}$. Let $D \to F$ be a minimal-area disc diagram bounded by $\gamma_1 \cdot \gamma_2 \cdots \gamma_n$ and a geodesic of $\mathcal{Y}$ joining its endpoints. Moreover, suppose that each of the geodesics, and indeed the sequence $v_0, \dots, v_n$ and the representative subspaces, are chosen so as to minimize the area of D among all possible such choices.

Then, arguing exactly as in the proof of Proposition 3.1 of [5], we see that $\gamma_1 \cdots \gamma_n$ can be chosen to be a geodesic since a minimal D cannot contain a dual curve traveling from γ_i to γ_j for any i and j. It follows that $\gamma_1 \cdots \gamma_n$ lies in $\mathcal{Y}$, so each H_i that is a hyperplane either crosses $\mathcal{Y}$ or contributes a combinatorial hyperplane to $\mathfrak{F}'$, while each H_i that is a subcomplex contributes an element to $\mathfrak{F}'$; as explained above, for each such H_i, we have $H_i \cap \mathcal{Y} \subsetneq F \cap \mathcal{Y}$, so $H_i \cap \mathcal{Y}$ corresponds to a cone-point in $\widehat{\mathcal{C}}(F \cap \mathcal{Y})$. We thus have a sequence $H_1, \dots, H_n$ of (non-$\sqsubseteq$–maximal) elements of $\mathfrak{F}'$ and hyperplanes crossing $\mathcal{Y}$, which determines a path of length between $n-1$ and $3(n-1)$ in $\widehat{\mathcal{C}}(F \cap \mathcal{Y})$. □

Lemma 10.13 *Let $\mathcal{X}$ be a* CAT(0) *cube complex with a factor system $\mathfrak{F}$. Suppose that A and B are unbounded convex subcomplexes of $\mathcal{X}$ such that there is a cubical* *isometric embedding $A \times B \to \mathcal{X}$ extending $A, B \hookrightarrow \mathcal{X}$. Then there exist $P_A, P_B \in \mathfrak{F}$ with $P_A \perp P_B$ and $A \subseteq P_A$ and $B \subseteq P_B$.*

Proof Let $x = A \cap B$. Then A and B are contained in combinatorial hyperplanes H_A and H_B, respectively. Indeed, every hyperplane crossing A (including the one whose carrier contains H_B) crosses every hyperplane crossing B (including the one whose carrier contains H_A). For each hyperplane V' crossing H_B, let V be one of the two associated combinatorial hyperplanes and consider $\mathfrak{g}_{H_A}(V)$. Observe that $\mathfrak{g}_{H_A}(V) \in \mathfrak{F}$ since it contains A and is thus unbounded. Since $\mathfrak{F}$ has finite multiplicity, there are only finitely many distinct subcomplexes $\mathfrak{g}_{H_A}(V)$, as V varies over all hyperplanes whose projection to H_A contains A; let $P_A \in \mathfrak{F}$ be their intersection. Define P_B analogously. Then P_A and P_B have the desired properties. (Indeed, a hyperplane H crosses P_A if and only if H crosses every hyperplane V whose projection to H_A contains A; the projection of H to H_B thus contains B, so every hyperplane crossing P_B crosses H, whence $P_A \times P_B \subset \mathcal{X}$.) $\square$

References

[1] **T Aougab**, *Uniform hyperbolicity of the graphs of curves*, Geom. Topol. 17 (2013) 2855–2875 MR

[2] **W Ballmann**, *On the Dirichlet problem at infinity for manifolds of nonpositive curvature*, Forum Math. 1 (1989) 201–213 MR

[3] **J A Behrstock**, *Asymptotic geometry of the mapping class group and Teichmüller space*, Geom. Topol. 10 (2006) 1523–1578 MR

[4] **J Behrstock**, **M F Hagen**, *Cubulated groups: thickness, relative hyperbolicity, and simplicial boundaries*, Groups Geom. Dyn. 10 (2016) 649–707 MR

[5] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, I: Curve complexes for cubical groups*, preprint (2014) arXiv

[6] **J Behrstock**, **M F Hagen**, **A Sisto**, *Hierarchically hyperbolic spaces, II: Combination theorems and the distance formula*, preprint (2015) arXiv

[7] **J Behrstock**, **M F Hagen**, **A Sisto**, *Asymptotic dimension and small-cancellation for hierarchically hyperbolic spaces and groups*, preprint (2015) arXiv

[8] **J Behrstock**, **B Kleiner**, **Y Minsky**, **L Mosher**, *Geometry and rigidity of mapping class groups*, Geom. Topol. 16 (2012) 781–888 MR

[9] **N Bergeron**, **D T Wise**, *A boundary criterion for cubulation*, Amer. J. Math. 134 (2012) 843–859 MR

[10] **B H Bowditch**, *Cut points and canonical splittings of hyperbolic groups*, Acta Math. 180 (1998) 145–186 MR

[11] **B H Bowditch**, *Tight geodesics in the curve complex*, Invent. Math. 171 (2008) 281–300 MR

[12] **B H Bowditch**, *Uniform hyperbolicity of the curve graphs*, Pacific J. Math. 269 (2014) 269–280 MR

[13] **B H Bowditch**, *Large-scale rank and rigidity of the Weil–Petersson metric*, preprint (2015) Available at `http://tinyurl.com/bowditchWP`

[14] **B H Bowditch**, *Large-scale rank and rigidity of the Teichmüller metric*, J. Topol. 9 (2016) 985–1020 MR

[15] **J F Brock**, *The Weil–Petersson metric and volumes of 3–dimensional hyperbolic convex cores*, J. Amer. Math. Soc. 16 (2003) 495–535 MR

[16] **J Brock**, **C Leininger**, **B Modami**, **K Rafi**, *Limit sets of Teichmüller geodesics with minimal nonuniquely ergodic vertical foliation, II*, preprint (2016) arXiv

[17] **P-E Caprace**, **Y Cornulier**, **N Monod**, **R Tessera**, *Amenable hyperbolic groups*, J. Eur. Math. Soc. 17 (2015) 2903–2947 MR

[18] **P-E Caprace**, **M Sageev**, *Rank rigidity for* CAT(0) *cube complexes*, Geom. Funct. Anal. 21 (2011) 851–891 MR

[19] **J Chaika**, **H Masur**, **M Wolf**, *Limits in $\mathcal{PMF}$ of Teichmüller geodesics*, preprint (2014) arXiv

[20] **M T Clay**, **C J Leininger**, **J Mangahas**, *The geometry of right-angled Artin subgroups of mapping class groups*, Groups Geom. Dyn. 6 (2012) 249–278 MR

[21] **M Clay**, **K Rafi**, **S Schleimer**, *Uniform hyperbolicity of the curve graph via surgery sequences*, Algebr. Geom. Topol. 14 (2014) 3325–3344 MR

[22] **C B Croke**, **B Kleiner**, *Spaces with nonpositive curvature and their ideal boundaries*, Topology 39 (2000) 549–556 MR

[23] **F Dahmani**, **V Guirardel**, **D Osin**, *Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces*, Mem. Amer. Math. Soc. 1156, Amer. Math. Soc., Providence, RI (2017) MR

[24] **R J Daverman**, *Decompositions of manifolds*, Pure and Applied Mathematics 124, Academic Press, Orlando, FL (1986) MR Reprinted as AMS-Chelsea 362, Amer. Math. Soc., Providence, RI (2007)

[25] **M J Dunwoody**, **M E Sageev**, *JSJ-splittings for finitely presented groups over slender groups*, Invent. Math. 135 (1999) 25–44 MR

[26] **M G Durham**, *The augmented marking complex of a surface*, J. Lond. Math. Soc. 94 (2016) 933–969 MR

[27] **M G Durham**, **S J Taylor**, *Convex cocompactness and stability in mapping class groups*, Algebr. Geom. Topol. 15 (2015) 2839–2859 MR

[28] **R D Edwards**, *Approximating certain cell-like maps by homeomorphisms*, Notices Amer. Math. Soc 24 (1977) art. id. A649

[29] **A Eskin**, **H Masur**, **K Rafi**, *Large-scale rank of Teichmüller space*, Duke Math. J. 166 (2017) 1517–1572 MR

[30] **B Farb**, **A Lubotzky**, **Y Minsky**, *Rank-1 phenomena for mapping class groups*, Duke Math. J. 106 (2001) 581–597 MR

[31] **B Farb**, **L Mosher**, *Convex cocompact subgroups of mapping class groups*, Geom. Topol. 6 (2002) 91–152 MR

[32] **K Fujiwara**, *Subgroups generated by two pseudo-Anosov elements in a mapping class group, I: Uniform exponential growth*, from "Groups of diffeomorphisms" (R Penner, D Kotschick, T Tsuboi, N Kawazumi, T Kitano, Y Mitsumatsu, editors), Adv. Stud. Pure Math. 52, Math. Soc. Japan, Tokyo (2008) 283–296 MR

[33] **K Fujiwara**, **P Papasoglu**, *JSJ-decompositions of finitely presented groups and complexes of groups*, Geom. Funct. Anal. 16 (2006) 70–125 MR

[34] **H Furstenberg**, *Noncommuting random products*, Trans. Amer. Math. Soc. 108 (1963) 377–428 MR

[35] **M Gromov**, *Hyperbolic groups*, from "Essays in group theory" (S M Gersten, editor), Math. Sci. Res. Inst. Publ. 8, Springer (1987) 75–263 MR

[36] **M F Hagen**, *The simplicial boundary of a* CAT(0) *cube complex*, Algebr. Geom. Topol. 13 (2013) 1299–1367 MR

[37] **M F Hagen**, *Weak hyperbolicity of cube complexes and quasi-arboreal groups*, J. Topol. 7 (2014) 385–418 MR

[38] **M F Hagen**, **T Susse**, *Hierarchical hyperbolicity of all cubical groups*, preprint (2016) arXiv

[39] **F Haglund**, *Isometries of* CAT(0) *cube complexes are semi-simple*, preprint (2007) arXiv

[40] **U Hamenstädt**, *Word hyperbolic extensions of surface groups*, preprint (2005) arXiv

[41] **U Hamenstädt**, *Geometric properties of the mapping class group*, from "Problems on mapping class groups and related topics" (B Farb, editor), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 215–232 MR

[42] **M Handel**, **L Mosher**, *Subgroups of mapping class groups, revisited*, preprint (2010)

[43] **S Hensel**, **P Przytycki**, **R C H Webb**, *1–slim triangles and uniform hyperbolicity for arc graphs and curve graphs*, J. Eur. Math. Soc. 17 (2015) 755–762 MR

[44] **C Horbez**, *A short proof of Handel and Mosher's alternative for subgroups of* $\mathrm{Out}(F_N)$, Groups Geom. Dyn. 10 (2016) 709–721 MR

[45] **V A Kaimanovich**, **H Masur**, *The Poisson boundary of the mapping class group*, Invent. Math. 125 (1996) 221–264 MR

[46] **M Kalantar**, **M Kennedy**, *Boundaries of reduced* C^**–algebras of discrete groups*, J. Reine Angew. Math. 727 (2017) 247–267 MR

[47] **R P Kent, IV**, **C J Leininger**, *Subgroups of mapping class groups from the geometrical viewpoint*, from "In the tradition of Ahlfors–Bers, IV" (D Canary, J Gilman, J Heinonen, H Masur, editors), Contemp. Math. 432, Amer. Math. Soc., Providence, RI (2007) 119–141 MR

[48] **R P Kent, IV**, **C J Leininger**, *Shadows of mapping class groups: capturing convex cocompactness*, Geom. Funct. Anal. 18 (2008) 1270–1325 MR

[49] **S P Kerckhoff**, *The asymptotic geometry of Teichmüller space*, Topology 19 (1980) 23–41 MR

[50] **T Koberda**, *Right-angled Artin groups and a generalized isomorphism problem for finitely generated subgroups of mapping class groups*, Geom. Funct. Anal. 22 (2012) 1541–1590 MR

[51] **C Leininger**, **A Lenzhen**, **K Rafi**, *Limit sets of Teichmüller geodesics with minimal non-uniquely ergodic vertical foliation*, J. Reine Angew. Math. (online publication August 2015)

[52] **C J Leininger**, **A W Reid**, *A combination theorem for Veech subgroups of the mapping class group*, Geom. Funct. Anal. 16 (2006) 403–436 MR

[53] **A Lenzhen**, *Teichmüller geodesics that do not have a limit in $\mathcal{PMF}$*, Geom. Topol. 12 (2008) 177–197 MR

[54] **J Maher**, **G Tiozzo**, *Random walks on weakly hyperbolic groups*, preprint (2014) arXiv

[55] **J Mangahas**, *Uniform uniform exponential growth of subgroups of the mapping class group*, Geom. Funct. Anal. 19 (2010) 1468–1480 MR

[56] **J Mangahas**, *A recipe for shortword pseudo-Anosovs*, Amer. J. Math. 135 (2013) 1087–1116 MR

[57] **A Marden**, *On finitely generated Fuchsian groups*, Comment. Math. Helv. 42 (1967) 81–85 MR

[58] **H Masur**, *Two boundaries of Teichmüller space*, Duke Math. J. 49 (1982) 183–190 MR

[59] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, I: Hyperbolicity*, Invent. Math. 138 (1999) 103–149 MR

[60] **H A Masur**, **Y N Minsky**, *Geometry of the complex of curves, II: Hierarchical structure*, Geom. Funct. Anal. 10 (2000) 902–974 MR

[61] **Y N Minsky**, *Extremal length estimates and product regions in Teichmüller space*, Duke Math. J. 83 (1996) 249–286 MR

[62] **L Mosher**, *Problems in the geometry of surface group extensions*, from "Problems on mapping class groups and related topics" (B Farb, editor), Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI (2006) 245–256 MR

[63] **S C Mousley**, *Non-existence of boundary maps for some hierarchically hyperbolic spaces*, preprint (2016) arXiv

[64] **A Nevo**, **M Sageev**, *The Poisson boundary of* CAT(0) *cube complex groups*, Groups Geom. Dyn. 7 (2013) 653–695 MR

[65] **D Osin**, *Acylindrically hyperbolic groups*, Trans. Amer. Math. Soc. 368 (2016) 851–888 MR

[66] **P Przytycki**, **A Sisto**, *A note on acylindrical hyperbolicity of mapping class groups*, preprint (2015) arXiv

[67] **K Rafi**, *Hyperbolicity in Teichmüller space*, Geom. Topol. 18 (2014) 3025–3053 MR

[68] **A W Reid**, *Surface subgroups of 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) 257–268 MR

[69] **E Rips**, **Z Sela**, *Cyclic splittings of finitely presented groups and the canonical JSJ decomposition*, Ann. of Math. 146 (1997) 53–109 MR

[70] **A Sisto**, *Contracting elements and random walks*, preprint (2011) arXiv

[71] **J Stallings**, *Group theory and three-dimensional manifolds*, Yale Mathematical Monographs 4, Yale University Press, New Haven, CT (1971) MR

[72] **W P Thurston**, *On the geometry and dynamics of diffeomorphisms of surfaces*, Bull. Amer. Math. Soc. 19 (1988) 417–431 MR

[73] **W A Veech**, *Teichmüller curves in moduli space, Eisenstein series and an application to triangular billiards*, Invent. Math. 97 (1989) 553–583 MR

[74] **R C H Webb**, *A short proof of the bounded geodesic image theorem*, preprint (2013) arXiv

[75] **W Woess**, *Boundaries of random walks on graphs and groups with infinitely many ends*, Israel J. Math. 68 (1989) 271–301 MR

Department of Mathematics, University of Michigan
Ann Arbor, MI, United States

Department of Pure Maths and Math. Stat., University of Cambridge
Cambridge, United Kingdom

Departement Mathematik, ETH Zürich
Zürich, Switzerland

durhamma@umich.edu, markfhagen@gmail.com, sisto@math.ethz.ch

http://www-personal.umich.edu/~durhamma/, http://www.wescac.net/, http://people.math.ethz.ch/~alsisto/

Proposed: Benson Farb Received: 23 May 2016
Seconded: Anna Wienhard, Bruce Kleiner Revised: 30 December 2016

Geometry & Topology 21 (2017) 3759–3784

Thurston norm via Fox calculus

STEFAN FRIEDL
KEVIN SCHREVE
STEPHAN TILLMANN

In 1976 Thurston associated to a 3–manifold N a marked polytope in $H_1(N;\mathbb{R})$, which measures the minimal complexity of surfaces representing homology classes and determines all fibered classes in $H^1(N;\mathbb{R})$. Recently the first and third authors associated to a presentation π with two generators and one relator a marked polytope in $H_1(\pi;\mathbb{R})$ and showed that it determines the Bieri–Neumann–Strebel invariant of π. We show that if the fundamental group of a 3–manifold N admits such a presentation π, then the corresponding marked polytopes in $H_1(N;\mathbb{R}) = H_1(\pi;\mathbb{R})$ agree.

20J05, 57M05, 57M27, 57R19

1 Summary of results

Throughout this paper all 3–manifolds are compact, connected and orientable. Suppose N is a 3–manifold. In 1976 Thurston [49] introduced a seminorm x_N on $H^1(N;\mathbb{R})$, henceforth referred to as the *Thurston norm*, which is a natural measure of the complexity of surfaces dual to integral classes. A class $\phi \in H^1(N;\mathbb{R})$ is *fibered* if ϕ can be represented by a nondegenerate closed 1–form. If ϕ is integral, then ϕ is fibered if and only if it is induced by a surface bundle $N \to S^1$. We refer to Section 2.4 for details.

Thurston [49] showed that the information on the Thurston seminorm and the fibered classes can be encapsulated in terms of a *marked polytope*.

A marked polytope is a polytope in a vector space together with a (possibly empty) set of marked vertices. In order to state Thurston's result precisely we need one more definition. Given a polytope in a vector space V we say that a homomorphism $\phi \in \operatorname{Hom}(V,\mathbb{R})$ *pairs maximally with the vertex* v if $\phi(v) > \phi(w)$ for all other vertices $w \neq v$. In this language, the main result of [49] can be stated as follows:

Theorem 1.1 *Let N be a 3–manifold. There exists a unique symmetric marked polytope $\mathcal{M}_N$ in $H_1(N;\mathbb{R})$ such that for any $\phi \in H^1(N;\mathbb{R}) = \operatorname{Hom}(\pi_1(N),\mathbb{R})$ we have*

$$x_N(\phi) = \max\{\phi(p) - \phi(q) \mid p, q \in \mathcal{M}_N\},$$

and ϕ is fibered if and only if it pairs maximally with a marked vertex of $\mathcal{M}_N$.

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3759

Subsequently, by a $(2,1)$–presentation we mean a group presentation with precisely two generators and one nonempty relator. A $(2,1)$–presentation is *cyclically reduced* if the relator is a cyclically reduced word. Recently, the first and third authors [24] associated to a cyclically reduced $(2,1)$–presentation $\pi = \langle x, y \mid r \rangle$ a marked polytope $\mathcal{M}_\pi$ in $H_1(\pi;\mathbb{R})$.

Now we outline the definition of $\mathcal{M}_\pi$ in the case that $b_1(\pi) = 2$. A different (but equivalent) definition is given in Section 2.6, as well as a definition for cyclically reduced $(2,1)$–presentations π with $b_1(\pi) = 1$.

Identify $H_1(G_\pi;\mathbb{Z})$ with $\mathbb{Z}^2$ such that x corresponds to $(1,0)$ and y corresponds to $(0,1)$. Then the relator r determines a discrete walk on the integer lattice in $H_1(G_\pi;\mathbb{R})$, and the marked polytope $\mathcal{M}_\pi$ is obtained from the convex hull of the trace of this walk as follows:

(1) Start at the origin and walk across $\mathbb{Z}^2$ reading the word r from the left.

(2) Take the convex hull $\mathcal{C}$ of the set of all lattice points reached by the walk.

(3) Mark precisely those vertices of $\mathcal{C}$ which the walk passes through exactly once.

(4) Now consider the unit squares that are completely contained in $\mathcal{C}$ and touch a vertex of $\mathcal{C}$. Mark a midpoint of a square precisely when one (and hence all) vertices of $\mathcal{C}$ incident with the square are marked.

(5) The set of vertices of $\mathcal{M}_\pi$ is the set of midpoints of all of these squares, and a vertex of $\mathcal{M}_\pi$ is marked precisely when it is a marked midpoint of a square.

In Figure 1 we sketch the construction of $\mathcal{M}_\pi$ for the presentation $\pi = \langle x, y \mid r \rangle$, where

$$r = x^2yx^{-1}yx^2yx^{-1}y^{-3}x^{-1}yx^2yx^{-1}yxy^{-1}x^{-2}y^{-1}xy^{-1}x^{-2}y^{-1}xy^3xy^{-1} \cdot x^{-2}y^{-1}xy^{-1}x^{-1}y.$$

This example is due to Dunfield [12] and presents the fundamental group of the exterior of the 2–component link in S^3 shown in Figure 2 (see Section 6.3).

Given two polytopes $\mathcal{P}$ and $\mathcal{Q}$ in a vector space V, we write $\mathcal{P} \doteq \mathcal{Q}$ if the polytopes $\mathcal{P}$ and $\mathcal{Q}$ differ by a translation, ie if there exists $v \in V$ with $\mathcal{P} = v + \mathcal{Q}$. The following is the main theorem of this paper:

Theorem 1.2 *Let N be an irreducible 3–manifold that admits a cyclically reduced $(2,1)$–presentation $\pi = \langle x, y \mid r \rangle$. Then*

$$\mathcal{M}_N \doteq \mathcal{M}_\pi.$$

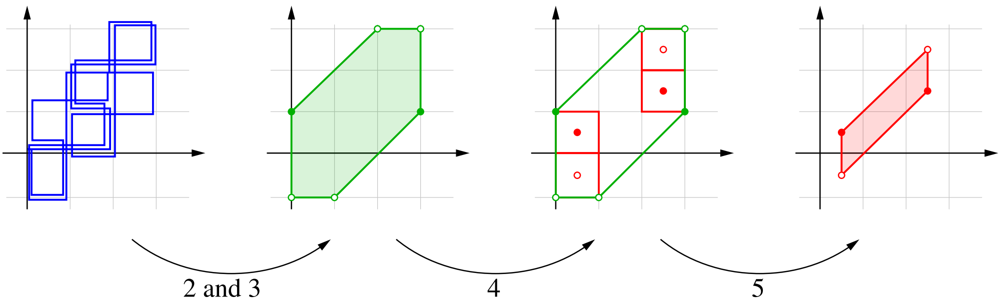


Figure 1: The marked polytope $\mathcal{M}_\pi$ for Dunfield's example, starting from the path determined by the relation r. Marked vertices are filled and unmarked vertices are empty; Labels on arrows correspond to steps in the above algorithm.

This theorem answers in particular a question of Sikorav [48] in the affirmative for 3–manifolds that admit a $(2,1)$–presentation.

The proof of Theorem 1.2 relies on the virtually special theorem of Agol [2], Liu [36], Przytycki and Wise [42; 43] and Wise [56; 57; 58], which we recall in Section 3.1. It also hinges on the following general result, which is of independent interest.

Theorem 1.3 *Let N be an irreducible 3–manifold with empty or toroidal boundary. If N is not a closed graph manifold, then $\pi_1(N)$ is residually a torsion-free and elementary amenable group.*

The proof of Theorem 1.3 uses the virtually special theorem and builds on work of Linnell and Schick [35]. It is proved in Section 3, where we also give several consequences.

We give a brief outline of the proof of Theorem 1.2. The starting point is an alternative definition of the marked polytope $\mathcal{M}_\pi$ using Fox derivatives [17] (see Section 2.6). This definition is less pictorial, but it allows us to relate the polytope $\mathcal{M}_\pi$ to the chain complex of the universal cover of the 2–complex X associated to the presentation π. This makes it possible to study the "size" of $\mathcal{M}_\pi$ using twisted Reidemeister torsions corresponding to finite-dimensional complex representations and corresponding to skew fields of X; see Cochran [8], Friedl [18], Friedl and Vidussi [25], Harvey [27] and Wada [54]. Since X is simple homotopy equivalent to N, these twisted Reidemeister torsions agree with the twisted Reidemeister torsions of N.

In the following we denote by $\mathcal{P}_N$ and $\mathcal{P}_\pi$ the polytopes $\mathcal{M}_N$ and $\mathcal{M}_\pi$ without the markings. Given two polytopes $\mathcal{P}$ and $\mathcal{Q}$ in a vector space V we write $\mathcal{P} \le \mathcal{Q}$ if there exists $v \in V$ with $v + \mathcal{P} \subset \mathcal{Q}$. The proof of Theorem 1.2 now breaks up into three parts:

(1) We first show that $\mathcal{P}_N \le \mathcal{P}_\pi$. Put differently, we show that $\mathcal{P}_\pi$ is "big enough" to contain $\mathcal{P}_N$. This is achieved with the main theorem of Friedl and Vidussi [26], which states that twisted Reidemeister torsions corresponding to finite-dimensional complex representations detect the Thurston norm of N. This relies on the virtually special theorem. See Section 4.

(2) Next we show the reverse inclusion $\mathcal{P}_\pi \le \mathcal{P}_N$. This means that $\mathcal{P}_\pi$ is "not bigger than necessary". At this stage it is crucial that r is cyclically reduced. Using Theorem 1.3 and the noncommutative Reidemeister torsions of Cochran [8], Friedl [18] and Harvey [27] we show that indeed $\mathcal{P}_\pi \subset \mathcal{P}_N$. See Section 5.

(3) Finally we need to show that the markings of $\mathcal{M}_N$ and $\mathcal{M}_\pi$ agree. This follows immediately from Friedl and Tillmann [24, Theorem 1.1] and Bieri, Neumann and Strebel [4, Theorem E]. See Section 5.3.

The paper is concluded with a conjecture and a question in Section 7.

Convention Throughout this paper, all groups are finitely generated, all vector spaces are finite-dimensional, and all 3–manifolds are compact, connected and orientable.

Acknowledgments Friedl is very grateful for being hosted by the University of Sydney in March 2013 and again in March 2014. Friedl and Schreve also wish to thank the Mathematisches Forschungsinstitut Oberwolfach for hosting them in January 2015. Friedl was supported by the SFB 1085 *Higher invariants*, funded by the Deutsche Forschungsgemeinschaft (DFG). Schreve is supported by NSF grant DMS-1045119. Tillmann is partially supported under the Australian Research Council's Discovery funding scheme (project number DP140100158). The authors thank Nathan Dunfield for very insightful conversations. The authors also thank the referee for a careful reading of this paper and a very helpful report.

2 Polytopes associated to 3–manifolds and groups

2.1 Polytopes

Let V be a real vector space and let $Q = \{Q_1, \dots, Q_k\} \subset V$ be a finite (possibly empty) subset. Denote by

$$\mathcal{P}(Q) = \operatorname{conv}(Q) = \left\{ \sum_{i=1}^{k} t_i Q_i \;\middle|\; \sum_{i=1}^{k} t_i = 1,\ t_i \ge 0 \right\}$$

the *polytope spanned by* Q. A *polytope* in V is a subset of the form $\mathcal{P}(Q)$ for some finite subset Q of V. For any polytope $\mathcal{P}$ there exists a unique smallest subset $\mathcal{V}(\mathcal{P}) \subset \mathcal{P}$ such that $\mathcal{P}$ is the polytope spanned by $\mathcal{V}(\mathcal{P})$. The elements of $\mathcal{V}(\mathcal{P})$ are the *vertices* of $\mathcal{P}$. Note $v \in \mathcal{P}$ is a vertex if and only if there exists a homomorphism $\phi\colon V \to \mathbb{R}$ such that $\phi(v) > \phi(p)$ for every $p \in \mathcal{P}$ with $p \neq v$.

Let V be a real vector space and let $\mathcal{P}$ and $\mathcal{Q}$ be two polytopes in V. The *Minkowski sum of $\mathcal{P}$ and $\mathcal{Q}$* is

$$\mathcal{P} + \mathcal{Q} := \{p + q \mid p \in \mathcal{P} \text{ and } q \in \mathcal{Q}\}.$$

It is straightforward to see that $\mathcal{P} + \mathcal{Q}$ is again a polytope. Furthermore, for each vertex u of $\mathcal{P} + \mathcal{Q}$ there exists a unique vertex v of $\mathcal{P}$ and a unique vertex w of $\mathcal{Q}$ such that $u = v + w$. Conversely, for each vertex v of $\mathcal{P}$ there exists a (not necessarily unique) vertex w of $\mathcal{Q}$ such that $v + w$ is a vertex of $\mathcal{P} + \mathcal{Q}$.

If $\mathcal{P}$, $\mathcal{Q}$ and $\mathcal{R}$ are polytopes with $\mathcal{P} + \mathcal{Q} = \mathcal{R}$, then we write $\mathcal{P} = \mathcal{R} - \mathcal{Q}$. We have

$$\mathcal{P} = \{p \in V \mid p + \mathcal{Q} \subset \mathcal{R}\};$$

in particular, $\mathcal{R} - \mathcal{Q}$ is well-defined.

There is a natural scaling operation on polytopes

$$\lambda \cdot \mathcal{P} := \{\lambda p \mid p \in \mathcal{P}\},$$

where $\mathcal{P} \subset V$ is a polytope and $\lambda \in \mathbb{R}^+$. If $k \in \mathbb{N}$, then the Minkowski sum of k copies of $\mathcal{P}$ equals $k\mathcal{P}$.

2.2 Convex sets and seminorms

Let $\mathcal{C}$ be a nonempty convex set in the real vector space V. Given $\phi \in \operatorname{Hom}(V, \mathbb{R})$ we define the *thickness of $\mathcal{C}$ in the ϕ–direction* by

$$\operatorname{th}_{\mathcal{C}}(\phi) := \max\{\phi(c) - \phi(d) \mid c, d \in \mathcal{C}\}.$$

It is straightforward to see that the function

$$\lambda_{\mathcal{C}}\colon \operatorname{Hom}(V, \mathbb{R}) \to \mathbb{R}_{\geq 0}, \quad \phi \mapsto \operatorname{th}_{\mathcal{C}}(\phi),$$

is a seminorm. Conversely, a seminorm $\lambda\colon \operatorname{Hom}(V, \mathbb{R}) \to \mathbb{R}_{\geq 0}$ defines the convex set

$$\mathcal{C}(\lambda) := \{v \in V \mid \phi(v) \leq 1 \text{ for all } \phi \in \operatorname{Hom}(V, \mathbb{R}) \text{ with } \lambda(\phi) \leq 1\}.$$

Note that $\mathcal{C}(\lambda)$ is *symmetric* since $v \in \mathcal{C}(\lambda)$ implies $-v \in \mathcal{C}(\lambda)$. For any seminorm λ on $\operatorname{Hom}(V, \mathbb{R})$ we have $\lambda_{\mathcal{C}(\lambda)} = \lambda$. On the other hand, if $\mathcal{C}$ is a nonempty convex set of V, then $\mathcal{C}(\lambda_{\mathcal{C}})$ equals the symmetrization of $\mathcal{C}$,

$$\mathcal{C}^{\mathrm{sym}} := \left\{\tfrac{1}{2}(c - d) \mid c, d \in \mathcal{C}\right\}.$$

Finally, given a convex set $\mathcal{C}$ in V the *dual* of $\mathcal{C}$ is

$$\mathcal{C}^* := \{\phi \in \operatorname{Hom}(V, \mathbb{R}) \mid \phi(v) \leq 1 \text{ for all } v \in \mathcal{C}\}.$$

2.3 Marked polytopes

Let V be a real vector space. A *marked polytope* $\mathcal{M}$ in V is a polytope $\mathcal{P}$ and a (possibly empty) subset $\mathcal{V}^+$ of $\mathcal{V}(\mathcal{P})$. The elements of $\mathcal{V}^+$ are the *marked vertices*; the elements of $\mathcal{V}(\mathcal{P}) \setminus \mathcal{V}^+$ are the *unmarked vertices* and $\mathcal{P}$ is the *underlying polytope* of $\mathcal{M}$.

If $\mathcal{M} = (\mathcal{P}, \mathcal{V}^+)$ and $\mathcal{N} = (\mathcal{Q}, \mathcal{W}^+)$ are two marked polytopes, then the *Minkowski sum of* $\mathcal{M}$ *and* $\mathcal{N}$ has underlying polytope the Minkowski sum of the underlying polytopes and set of marked vertices precisely those that are sums of marked vertices:

$$\mathcal{M} + \mathcal{N} = (\,\mathcal{P} + \mathcal{Q},\ \mathcal{V}(\mathcal{P} + \mathcal{Q}) \cap (\mathcal{V}^+ + \mathcal{W}^+)\,).$$

The marked polytope $\mathcal{M} = (\mathcal{P}, \mathcal{V}^+)$ is *symmetric* if the underlying polytope $\mathcal{P}$ is symmetric and $\mathcal{V}^+ = -\mathcal{V}^+$.

2.4 The Thurston norm and fibered classes

Let N be a 3–manifold. For each $\phi \in H^1(N;\mathbb{Z})$ there is a properly embedded oriented surface Σ such that $[\Sigma] \in H_2(N, \partial N;\mathbb{Z})$ is the Poincaré dual to ϕ. Letting $\chi_-(\Sigma) = \sum_{i=1}^{k} \max\{-\chi(\Sigma_i), 0\}$, where $\Sigma_1, \dots, \Sigma_k$ are the connected components of Σ, the *Thurston norm* of $\phi \in H^1(N;\mathbb{Z})$ is

$$x_N(\phi) = \min\{\chi_-(\Sigma) \mid [\Sigma] = \phi\}.$$

The class $\phi \in H^1(N;\mathbb{R})$ is called *fibered* if it can be represented by a nondegenerate closed 1–form. By [50] an integral class $\phi \in H^1(N;\mathbb{Z}) = \mathrm{Hom}(\pi_1(N), \mathbb{Z})$ is fibered if and only if there exists a fibration $p\colon N \to S^1$ such that $p_* = \phi\colon \pi_1(N) \to \pi_1(S^1) = \mathbb{Z}$.

Thurston [49] showed that x_N extends to a seminorm x_N on $H^1(N;\mathbb{R})$ and that the dual $\mathcal{C}(x_N)^*$ to the unit norm ball $\mathcal{C}(x_N)$ of the seminorm x_N is a polytope $\mathcal{P}_N$ with vertices in $\mathrm{Im}\{H_1(N;\mathbb{Z})/\text{torsion} \to H_1(N;\mathbb{R})\}$. Furthermore, Thurston showed that we can turn $\mathcal{P}_N$ into a marked polytope $\mathcal{M}_N$, which has the property that $\phi \in H^1(N;\mathbb{R}) = \mathrm{Hom}(H_1(N;\mathbb{R}), \mathbb{R})$ is fibered if and only if it pairs maximally with a marked vertex.

2.5 The marked polytope for elements of group rings

Let G be a group. Throughout this paper, given $f \in \mathbb{C}[G]$ and $g \in G$ we let f_g denote the g–coefficient of f. Let $\psi\colon G \to H_1(G;\mathbb{Z})/\text{torsion}$ be the canonical map.

We write $V = H_1(G;\mathbb{R})$ and we view $H_1(G;\mathbb{Z})/\text{torsion}$ as a subset of V. With this convention the above map ψ gives rise to a map $\psi\colon G \to V$. Given $f \in \mathbb{C}[G]$ we refer to

$$\mathcal{P}(f) := \mathcal{P}\big(\{\psi(g) \mid g \in G \text{ with } f_g \neq 0\}\big) \subset V$$

as the *polytope of* f. We will now associate to $\mathcal{P}(f)$ a marking. In order to do this we need a few more definitions:

(1) For $v \in V$ we refer to $f^v := \sum_{g \in \psi^{-1}(v)} f_g g$ as the *v–component of f*.

(2) We say that an element $r \in \mathbb{C}[G]$ is a monomial if it is of the form $r = \pm g$ for some $g \in G$.

A vertex v of $\mathcal{P}(f)$ is *marked* precisely when the v–component of f is a monomial. We then refer to the polytope $\mathcal{P}(f)$ together with the set of all marked vertices as the *marked polytope $\mathcal{M}(f)$ of f*.

The proof of [24, Lemma 3.2] applies with the above definitions, to give:

Lemma 2.1 *Let G be a group and let $f, g \in \mathbb{C}[G]$. Then the following hold:*

(1) *If for every vertex v of $\mathcal{P}(f)$ the v–component $f^v \in \mathbb{C}[G]$ is not a zero divisor, then $\mathcal{P}(f \cdot g) = \mathcal{P}(f) + \mathcal{P}(g)$.*

(2) *If each vertex of $\mathcal{M}(f)$ is marked, then $\mathcal{M}(f \cdot g) = \mathcal{M}(f) + \mathcal{M}(g)$.*

2.6 The marked polytope for a $(2,1)$–presentation

Let F be the free group with generators x and y. Following [17] we denote by $\partial/\partial x \colon \mathbb{Z}[F] \to \mathbb{Z}[F]$ the *Fox derivative with respect to x*, ie the unique $\mathbb{Z}$–linear map such that

$$\frac{\partial 1}{\partial x} = 0, \qquad \frac{\partial x}{\partial x} = 1, \qquad \frac{\partial y}{\partial x} = 0 \quad \text{and} \quad \frac{\partial uv}{\partial x} = \frac{\partial u}{\partial x} + u\frac{\partial v}{\partial x}$$

for all $u, v \in F$. We similarly define the Fox derivative with respect to y, and often write

$$u_x := \frac{\partial u}{\partial x} \quad \text{and} \quad u_y := \frac{\partial u}{\partial y}.$$

In [24] we proved the following proposition:

Proposition 2.2 *Let $\pi = \langle x, y \mid r\rangle$ be a $(2,1)$–presentation with $b_1(\pi) = 2$. Then there exists a marked polytope $\mathcal{M}$, unique up to translation, such that*

$$\mathcal{M} + \mathcal{M}(x-1) \doteq \mathcal{M}(r_y) \quad \text{and} \quad \mathcal{M} + \mathcal{M}(y-1) \doteq \mathcal{M}(r_x).$$

Denote by $\mathcal{M}_\pi$ the marked polytope of Proposition 2.2. Up to translation it is a well-defined invariant of the presentation, and it is shown in [24] that this definition is equivalent to the one sketched in the introduction.

A $(2,1)$–presentation $\pi = \langle x, y \mid r\rangle$ is *simple* if $b_1(G_\pi) = 1$, x defines a generator of $H_1(\pi;\mathbb{Z})/\text{torsion}$ and y represents the trivial element in $H_1(\pi;\mathbb{Z})/\text{torsion}$. In [24] we showed that given a simple $(2,1)$–presentation $\pi = \langle x, y \mid r\rangle$ there exists a marked polytope $\mathcal{M}_\pi$, unique up to translation, such that

$$\mathcal{M}_\pi + \mathcal{M}(x-1) \doteq \mathcal{M}(r_y).$$

It was shown in [24] that there is a canonical way to associate to any $(2,1)$–presentation $\pi = \langle x, y \mid r\rangle$ with $b_1(G_\pi) = 1$ a simple presentation $\pi' = \langle x', y' \mid r'\rangle$ representing the same group. We then define $\mathcal{M}_\pi := \mathcal{M}_{\pi'}$.

2.7 3–manifold groups which admit $(2,1)$–presentations

Manifolds having fundamental group with a $(2,1)$–presentation are described in Section 6. The only specific result needed to develop our theory is the following, which follows from work of Epstein [16].

Theorem 2.3 *Let N be an irreducible (compact, connected and orientable) 3–manifold such that $\pi := \pi_1(N)$ admits a $(2,1)$–presentation. Then the boundary of N consists of one or two tori.*

Proof Groups that admit a $(2,1)$–presentation have deficiency 1, while the fundamental group of a closed irreducible 3–manifold has deficiency zero [16, Section 3]. Whence N has nonempty boundary, and [16, Lemma 2.2] implies that $\frac{1}{2}\chi(\partial N) = \chi(N) \geq 0$. No boundary component of N is a sphere since we assume N is irreducible and $\pi_1(N) \neq \{1\}$. Since N (and hence each of its boundary components) is orientable, we now have $\chi(\partial N) = 0$ and every boundary component is a torus.

A standard half-lives, half-dies argument shows $b_1(\partial N) \leq 2b_1(N)$. Since $b_1(N) \leq 2$ we deduce that ∂N consists of either one or two tori. □

3 Properties of 3–manifold groups

3.1 The virtually special theorem

As usual, given a property of groups or spaces we say this property is satisfied *virtually* if a finite-index subgroup (not necessarily normal) or a finite-index cover (not necessarily regular) has the property.

In the following, given a 3–manifold N we say that $\phi \in H^1(N;\mathbb{R})$ is *quasifibered* if it is a limit of fibered classes in $H^1(N;\mathbb{R})$. The following theorem is now a variation of the virtually special theorem combined with Agol's virtual fibering theorem [1, Theorem 5.1] (see also [22, Theorem 5.1] for an exposition).

Theorem 3.1 *Let N be an irreducible 3–manifold with empty or toroidal boundary. If N is not a closed graph manifold, then for every $\phi \in H^1(N;\mathbb{R})$ there exists a finite-index cover $p\colon N' \to N$ such that $p^*(\phi)$ is quasifibered.*

The theorem was proved by Agol [2] for all closed hyperbolic 3–manifolds, by Wise [56; 57; 58] for all hyperbolic 3–manifolds with boundary, by Liu [36] and Przytycki and Wise [43] for all graph manifolds with boundary and by Przytycki and Wise [42] for all 3–manifolds with a nontrivial JSJ–decomposition that has at least one hyperbolic JSJ–component. We refer to [3] for precise references.

If we apply the theorem to the zero class we get in particular the following corollary:

Corollary 3.2 *An irreducible 3–manifold with empty or toroidal boundary is virtually fibered unless it is a closed graph manifold.*

3.2 Residual properties of 3–manifold groups

We start with several definitions, most of which are standard. Let $\mathcal{P}$ be a class of groups.

(1) The group π is *residually* $\mathcal{P}$ if for every nontrivial $g \in \pi$, there exists a homomorphism $\alpha\colon \pi \to \Gamma$ to a group in Γ in $\mathcal{P}$ such that $\alpha(g) \neq 1$.

(2) The group π is *fully residually* $\mathcal{P}$ if for every finite subset $\{g_1, \dots, g_n\} \subset \pi \setminus \{1\}$, there exists a epimorphism $\alpha\colon \pi \to G$ to a group in Γ in $\mathcal{P}$ such that $\alpha(g_i) \neq 1$ for all $i = 1, \dots, n$.

(3) The group π has the *$\mathcal{P}$–factorization property* if for every epimorphism $\alpha\colon \pi \to G$ onto a finite group G there exists an epimorphism $\beta\colon \pi \to \Gamma$ to a group Γ in $\mathcal{P}$ such that α factors through β.

We are mostly interested in the following classes of groups.

(1) The class $\mathcal{EA}$ of *elementary amenable groups* is the smallest class of groups that contains all abelian and all finite groups and that is closed under extensions and directed unions.

(2) We denote by $\mathcal{TEA}$ the class of all groups that are torsion-free and elementary amenable. It is clear that $\mathcal{TEA}$ is closed under taking finite direct products.

Using Corollary 3.2 and work of Linnell and Schick [35] we will prove the following theorem:

Theorem 3.3 *Let N be an irreducible 3–manifold with empty or toroidal boundary. If N is not a closed graph manifold, then $\pi_1(N)$ has the $\mathcal{TEA}$–factorization property.*

The question of to what degree this statement holds for closed graph manifolds is discussed in Section 3.4. We postpone the proof of the theorem to Section 3.3, and point out several corollaries.

Theorem 1.3 *Let N be an irreducible 3–manifold with empty or toroidal boundary. If N is not a closed graph manifold, then $\pi_1(N)$ is residually $\mathcal{TEA}$.*

Proof Let $\mathcal{P}$ be any class of groups. If a group π is residually finite and has the $\mathcal{P}$–factorization property, then G is also residually $\mathcal{P}$. The statement of the theorem now follows from Theorem 3.3 and the fact that 3–manifold groups are residually finite [29]. □

Corollary 3.4 *Let π be the fundamental group of an irreducible 3–manifold that has empty or toroidal boundary and is not a closed graph manifold. For every nonzero element $p \in \mathbb{Z}[\pi]$, there exists a homomorphism $\alpha\colon \pi \to \Gamma \in \mathcal{TEA}$ such that $0 \neq \alpha(p) \in \mathbb{Z}[\Gamma]$.*

Proof We write $p = \sum_{i=1}^{k} a_i g_i$, where $a_1, \dots, a_k \neq 0$ and $g_1, \dots, g_n \in \pi$ are pairwise distinct. By Theorem 1.3 the group π is residually $\mathcal{TEA}$. Since $\mathcal{TEA}$ is closed under taking finite direct products, π is also fully residually $\mathcal{TEA}$. We can thus find a homomorphism $\alpha\colon \pi \to \Gamma$ to a group $\Gamma \in \mathcal{TEA}$ such that all $\alpha(g_i)$ and all products $\alpha(g_i g_j^{-1})$ with $i \neq j$ are nontrivial. Whence $\alpha(p) \in \mathbb{Z}[\Gamma]$ is nonzero. □

3.3 Proof of Theorem 3.3

The following lemma is probably well-known to the experts.

Lemma 3.5 *Let E be a surface group (ie the fundamental group of a compact orientable surface, possibly with boundary) and let $R \subset E$ be a normal subgroup. Then $E/[R,R]$ is torsion-free.*

Proof Let $g \in E/[R,R]$ be a nontrivial element. We pick a representative for g in E, which by slight abuse of notation we also denote by g. We denote by S the subgroup of E generated by g and R. It suffices to prove the following claim:

Claim *The group $S/[R,R]$ is torsion-free.*

We consider the short exact sequence

$$1 \to [S,S]/[R,R] \to S/[R,R] \to S/[S,S] \to 0.$$

Since R and S are either surface groups or infinitely generated free groups we deduce that $S/[S,S] = H_1(S;\mathbb{Z})$ and $R/[R,R] = H_1(R;\mathbb{Z})$ are torsion-free. The group S/R is generated by one element, which implies that S/R is cyclic, in particular abelian. It follows that $[S,S]\subset R$. We thus see that $[S,S]/[R,R]$ is a subgroup of $R/[R,R]$. So the groups on the left and on the right of the above short exact sequence are torsion-free. It follows that $S/[R,R]$ is torsion-free. □

Proposition 3.6 *If $1\to E\to\pi\to M\to 1$ is an exact sequence with E a surface group and $M\in\mathcal{TEA}$, then π has the $\mathcal{TEA}$–factorization property.*

Proof Let $\alpha\colon \pi\to P$ be a map to a finite group. Let $R=E\cap\operatorname{Ker}\alpha$. By Lemma 3.5 the group $E/[R,R]$ is torsion-free. Furthermore it is elementary amenable by the exact sequence

$$1\to R/[R,R]\to E/[R,R]\to E/R\to 1.$$

Now α factors through $\pi/[R,R]$, and this is in $\mathcal{TEA}$ due to the sequence

$$1\to E/[R,R]\to \pi/[R,R]\to M\to 1.$$ □

The *profinite completion* of the group π is denoted by $\hat{\pi}$; see [44, Section 3.2] for a definition and its main properties. Following Serre [47, I.2.6, Exercise 2] we say that a group π is *good* if the natural morphism $H^*(\hat{\pi};A)\to H^*(\pi;A)$ is an isomorphism for any finite abelian group A with a π–action.

In the proof of the following theorem we will on several occasions use the following standard notation: if Γ is a subgroup of π, then $\Gamma^\pi:=\bigcap_{g\in\pi} g\Gamma g^{-1}$. Note that Γ^π is always a normal subgroup of π, and if Γ is of finite index, then Γ^π is of finite index. We also note that the methods of the proof build heavily on the work of Linnell and Schick [35].

Theorem 3.7 *Let π be a finitely generated torsion-free group that has a finite-dimensional classifying space and which is good. If π admits a finite-index subgroup Γ which has the $\mathcal{TEA}$–factorization property, then π also has the $\mathcal{TEA}$–factorization property.*

Proof Let $\alpha\colon\pi\to G$ be a homomorphism to a finite group. We denote by $K\subset\pi$ the intersection of $\operatorname{Ker}(\alpha)$ and Γ^π. The subgroup K is of finite index in π and is clearly contained in Γ. It follows from Lemma 2.1 of [46] that K also has the $\mathcal{TEA}$–factorization property. We write $Q:=\pi/K$. First suppose that Q is a p–group. It suffices to show there is a subgroup $U\trianglelefteq\pi$ such that the map $\pi\to Q$ factors through π/U and π/U is in $\mathcal{TEA}$.

If no such U exists, then since K has the $\mathcal{TEA}$–factorization property, there is a nontrivial subgroup Q' of Q that splits in the induced sequence of profinite completions

$$1\to\hat{K}\to\hat{\pi}\to Q\to 1;$$

see [46, Lemmas 3.4–3.6]. However, putting the following two observations together shows that this is not possible:

(1) The cohomology $H^*(Q', \mathbb{F}_p)$ is nonzero in infinitely many dimensions.

(2) By [47, I.2.6, Exercise 1, page 15] any finite-index subgroup L (such as K or the preimage of Q' under $\pi \to Q$) of π is also good and has a finite-dimensional classifying space. This implies that $H^*(\widehat{L}, \mathbb{F}_p) \cong H^*(L, \mathbb{F}_p)$ is nonzero in only finitely many dimensions.

For the general case, we use a trick from [35]. For each Sylow p–subgroup S of Q, consider the exact sequence $1 \to K \to \pi_S \to S \to 1$, where π_S is the preimage of S. By the above, we get for each S a subgroup U_S such that the quotient π_S/U_S is torsion-free elementary amenable. Let $U = \cap_S U_S$. Since π/U^π is a finite extension of Γ/U^π, elementary amenability follows from [35, Lemma 4.11]. It remains to show that π/U^π is torsion-free.

There is an exact sequence

$$1 \to U_S^\pi/U^\pi \to \pi_S/U^\pi \to \pi_S/U_S^\pi \to 1$$

with U_S^π/U^π and π_S/U_S^π torsion-free [35, Lemma 4.11]. Therefore, π_S/U^π is torsion-free.

Suppose that π/U^π has a nontrivial torsion element γ. By raising γ to some power we get an element γ' that is p–torsion for some prime p. Since K/U^π is torsion-free, γ' would map to some Sylow p–subgroup, in which case $\gamma' \in \pi_S/U^\pi$, which is torsion-free by the above. Therefore, π/U^π is torsion-free. □

Now we are finally in a position to prove Theorem 3.3.

Proof of Theorem 3.3 Let N be an irreducible 3–manifold that has empty or toroidal boundary and that is not a closed graph manifold. According to Corollary 3.2, N has a finite cover M that is fibered. The fundamental group of M is a semidirect product of $\mathbb{Z}$ with a surface group, and hence Lemma 3.5 and Proposition 3.6 imply $\pi_1(M)$ has the $\mathcal{TEA}$–factorization property.

It follows from [47, Exercise 2(b), page 16] that $\pi_1(M)$ is good. By [47, Exercise 1, page 15] the group $\pi_1(N)$ is also good. It is well-known (see eg [3, (A.1), page 44]) that N is aspherical and that in particular $\pi_1(N)$ is torsion-free. Thus we can apply Theorem 3.7 to $\pi_1(N)$ and the finite-index subgroup $\pi_1(M)$, giving the desired result that $\pi_1(N)$ has the $\mathcal{TEA}$–factorization property. □

Remark The same proof also shows that torsion-free virtually cocompact special groups have the $\mathcal{TEA}$–factorization property. Indeed, these groups are virtual retracts of right-angled Artin groups, and therefore contain finite index subgroups that are good and have the $\mathcal{TEA}$–factorization property [46]. The 3–manifold groups that we consider in Theorem 1.3 are not generally known to be virtually cocompact special. However, this observation implies that Theorem 1.3 holds for many other 3–manifold groups, eg for fundamental groups of hyperbolic 3–manifolds with infinite volume. We refer to [3, Theorem 4.3.6] for details and references.

3.4 The case of closed graph manifolds

It is natural to ask for which closed graph manifolds the conclusions of Theorem 3.3 and its corollaries hold. It follows from the work of Liu [36] that the conclusion of the theorem also holds for closed nonpositively curved graph manifolds. The question of which closed graph manifolds are nonpositively curved was treated in detail by Buyalo and Svetlov [7]. In the following we give a short list of examples of graph manifolds that are not nonpositively curved:

(1) spherical 3–manifolds;

(2) Sol– and Nil–manifolds;

(3) Seifert fibered 3–manifolds that are finitely covered by a nontrivial S^1–bundles over a closed surface.

It is clear that the statements do not hold for spherical 3–manifolds with nontrivial fundamental group. The following lemma takes care of the second case:

Lemma 3.8 *The fundamental groups of Sol– and Nil–manifolds are $\mathcal{TEA}$; in particular they have the $\mathcal{TEA}$–factorization property.*

Proof Sol– and Nil–manifolds are finitely covered by torus-bundles over S^1. Hence their fundamental groups are elementary amenable, but the fundamental groups are also torsion-free, so they are $\mathcal{TEA}$. □

Lemma 3.9 *Let N be a Seifert fibered space with infinite fundamental group. Then $\pi_1(N)$ has the $\mathcal{TEA}$–factorization property.*

Proof Since we will not make use of this lemma we only sketch the proof. The manifold N is finitely covered by an S^1–bundle over a surface. By Theorem 3.7 we can thus without loss of generality assume that N is an S^1–bundle over a surface F. Since $\pi_1(N)$ is infinite there exists a short exact sequence

$$1 \to \langle t\rangle \to \pi_1(N) \to \pi_1(F) \to 1,$$

where the subgroup $\langle t\rangle$ is generated by the S^1–fiber. By Proposition 3.6 the group $\pi_1(F)$ has the $\mathcal{TEA}$–factorization property. Let e denote the Euler number of the

S^1–bundle over F and denote by M the total space of the S^1–bundle over the torus with Euler number e. Then there exists a fiber-preserving map from N to M. Since $\pi_1(M)$ is $\mathcal{TEA}$ we have found a homomorphism from $\pi_1(N)$ to a $\mathcal{TEA}$ group which is injective on $\langle t\rangle$. Now it is straightforward to see that $\pi_1(N)$ has the $\mathcal{TEA}$–factorization property. □

The above discussion shows that the fundamental groups of many closed graph manifolds have the $\mathcal{TEA}$–factorization property. Nonetheless we expect that there are many closed graph manifolds whose fundamental groups do not have the $\mathcal{TEA}$–factorization property.

4 Proof of Theorem 1.2, I

The goal of this section is to prove the following proposition.

Proposition 4.1 *Let $\pi = \langle x, y \mid r\rangle$ be a cyclically reduced $(2,1)$–presentation for the fundamental group of an irreducible 3–manifold N. Then*

$$\mathcal{P}_N \leq \mathcal{P}_\pi.$$

The main ingredient in the proof will be the fact that twisted Reidemeister torsions corresponding to finite-dimensional complex representations detect the Thurston norm of 3–manifolds.

4.1 Tensor representations

Let π be a group, let $\alpha\colon \pi \to \mathrm{GL}(k,\mathbb{C})$ be a representation and let $\psi\colon \pi \to H$ be a homomorphism to a free abelian group. We denote by $\mathbb{C}(H)$ the quotient field of the group ring $\mathbb{C}[H]$. The homomorphisms α and ψ give rise to the representation

$$\alpha\otimes\psi\colon \pi \to \mathrm{GL}(k,\mathbb{C}(H)), \quad g \mapsto \alpha(g)\cdot\psi(k),$$

which we refer to as the tensor product of α and ψ. This representation extends to a ring homomorphism $\mathbb{Z}[\pi] \to M(k\times k,\mathbb{C}(H))$, which we also denote by $\alpha\otimes\psi$.

4.2 The definition of the twisted Reidemeister torsion

Let X be a finite CW–complex, $\pi := \pi_1(X)$, and denote by $\widetilde{X}$ the universal cover of X. The action of π via deck transformations on $\widetilde{X}$ equips the chain complex $C_*(\widetilde{X};\mathbb{Z})$ with the structure of a chain complex of $\mathbb{Z}[\pi]$–left modules.

Let $\alpha\colon \pi \to \mathrm{GL}(k,\mathbb{C})$ be a representation. We let $\psi\colon \pi \to H := H_1(X;\mathbb{Z})/\text{torsion}$ be the obvious projection map. Using the representation $\alpha \otimes \psi$ we can now view $\mathbb{C}(H)^k$ as a right $\mathbb{Z}[\pi]$–module, where the action is given by right multiplication on row vectors.

We consider the chain complex

$$C_*(X;\mathbb{C}(H)^k) := \mathbb{C}(H)^k \otimes_{\mathbb{Z}[\pi]} C_*(\widetilde{X};\mathbb{Z})$$

of $\mathbb{C}(H)$–modules. For each cell in X pick a lift to a cell in $\widetilde{X}$. We denote by $e_1,\dots,e_k$ the standard basis for $\mathbb{C}(H)^k$. The tensor products of the lifts of the cells and the vectors e_i turn $C_*(X;\mathbb{C}(H)^k)$ into a chain complex of based $\mathbb{C}(H)$–vector spaces.

If the chain complex $C_*(X;\mathbb{C}(H)^k)$ is not acyclic, then we define the corresponding twisted Reidemeister torsion $\tau(X,\alpha)$ to be zero. Otherwise we let $\tau(X,\alpha)\in\mathbb{C}(H)\setminus\{0\}$ be the torsion of the based chain complex $C_*(X;\mathbb{C}(H)^k)$. We refer to [52] for the definition of the torsion of a based chain complex. Standard arguments show that $\tau(X,\alpha)\in\mathbb{C}(H)\setminus\{0\}$ is well-defined up to multiplication by an element of the form zh, where $z\in\pm\det(\alpha(\pi))$ and $h\in H$. The indeterminacy arises from the fact that we had to choose lifts and an ordering of the cells.

Suppose N is a 3–manifold and let $\alpha\colon \pi_1(N)\to\mathrm{GL}(k,\mathbb{C})$ be a representation. Choose a CW–structure X for N and define $\tau(N,\alpha) := \tau(X,\alpha)$. It is well-known (see eg [52; 25]) that this definition does not depend on the choice of the CW–structure.

4.3 The polytopes corresponding to twisted Reidemeister torsion

As above, suppose N is a 3–manifold and $\alpha\colon \pi_1(N)\to\mathrm{GL}(k,\mathbb{C})$ a representation. If $\tau(N,\alpha)$ is zero, then we define $\mathcal{T}(N,\alpha)=\varnothing$.

Otherwise we write $\tau(N,\alpha)=p\cdot q^{-1}$ with $p,q\in\mathbb{C}[H]$. If the Minkowski difference $\mathcal{P}(q)-\mathcal{P}(p)$ exists (and by [25, page 53] this is the case if $b_1(N)\ge 2$), then we define

$$\mathcal{T}(N,\alpha) := \frac{1}{k}\cdot(\mathcal{P}(p)-\mathcal{P}(q)),$$

and otherwise define $\mathcal{T}(N,\alpha) := \{0\}$.

Proposition 4.2 *Let $\pi=\langle x,y\,|\,r\rangle$ be a $(2,1)$–presentation for the fundamental group of an irreducible 3–manifold N. Then for any representation we have*

$$\mathcal{T}(N,\alpha)\subset\mathcal{P}_\pi.$$

In the proof of the proposition we will need one more definition and one more lemma. Let π be a group, $f \in \mathbb{Z}[\pi]$, $\alpha\colon \pi \to \mathrm{GL}(k,\mathbb{C})$ be a representation, and $\psi_\pi\colon \pi \to H := H_1(\pi;\mathbb{Z})/\text{torsion}$ be the canonical epimorphism. Then $\det((\alpha\otimes\psi_\pi)(f)) \in \mathbb{C}[H]$ and we write

$$\mathcal{P}(f,\alpha) := \frac{1}{k}\mathcal{P}\big(\det((\alpha\otimes\psi_\pi)(f))\big) \subset H_1(\pi;\mathbb{R}).$$

Lemma 4.3 *Let π be a group, $f \in \mathbb{Z}[\pi]$ and $\alpha\colon \pi \to \mathrm{GL}(k,\mathbb{C})$ be a representation. Then*

$$\mathcal{P}(f,\alpha) \subset \mathcal{P}(f).$$

Proof We write $f = c_1h_1 + \dots + c_lh_l$ with $h_1,\dots,h_l \in \pi$ and $c_1,\dots,c_l \neq 0$. We consider

$$S := \{s_1\psi(g_1) + \dots + s_l\psi(h_l) \mid s_1,\dots,s_l \in \mathbb{C}\}.$$

Put differently, S is the set of all elements in $\mathbb{C}[H]$ with support some subset of $\{\psi(g_1),\dots,\psi(g_l)\}$. For every $p \in S$ we have $\mathcal{P}(p) \subset \mathcal{P}(\psi(g_1),\dots,\psi(g_l)) = \mathcal{P}(f)$. This implies that if $p_1,\dots,p_k$ are elements in S, then

$$\mathcal{P}(p_1\cdots p_k) = \mathcal{P}(p_1) + \dots + \mathcal{P}(p_k) \subset \mathcal{P}(f) + \dots + \mathcal{P}(f) = k\mathcal{P}(f).$$

We write $M := (\alpha\otimes\psi)(f) = \sum_{i=1}^{l} c_i\alpha(h_i)\cdot\psi(h_i)$. Each entry of $\det(M)$ lies in S. It follows from the Laplace formula that $\det(M)$ is a sum of products of the form $p_1\cdots p_k$, where each p_i lies in S. By the above we have $\mathcal{P}(p_1\cdots p_k) \subset k\mathcal{P}(f)$. The definitions imply that if $a,b \in \mathbb{C}[\pi]$ are such that $\mathcal{P}(a)$ and $\mathcal{P}(b)$ are contained in a polytope $\mathcal{Q}$, then we have also have $\mathcal{P}(a+b) \subset \mathcal{Q}$. Hence $\mathcal{P}(\det(M)) \subset k\mathcal{P}$. □

Proof of Proposition 4.2 We again denote by $\psi\colon \pi_1(N) \to H_1(N;\mathbb{Z})/\text{torsion}$ the canonical epimorphism. Note that $\psi(x) \neq 0$ or $\psi(y) \neq 0$. Without loss of generality we may assume $\psi(y) \neq 0$.

Theorem 2.3 shows that N has nontrivial toroidal boundary. It thus follows from [32, Theorem A] (see also [25, page 50]) that

$$\tau(N,\alpha) = \det((\alpha\otimes\psi)(r_y))\cdot\det((\alpha\otimes\psi)(y-1))^{-1}.$$

By Lemma 4.3 we have $\mathcal{P}(r_y,\alpha) \subset \mathcal{P}(r_y)$. Since $\psi(y) \neq 0$ we know that $\psi(y)$ and 1 are the two distinct vertices of $\mathcal{P}(y-1)$. Also, we have $\mathcal{P}(y-1,\alpha) = \frac{1}{k}\mathcal{P}\big(\det(\alpha(y)\psi(y) - \mathrm{id}_k)\big)$ and it is straightforward to see that this polytope equals $\mathcal{P}(y-1)$.

Combining these results we obtain

$$\mathcal{T}(N,\alpha) = \mathcal{P}(r_y,\alpha) - \mathcal{P}(y-1,\alpha) = \mathcal{P}(r_y,\alpha) - \mathcal{P}(y-1) \subset \mathcal{P}(r_y) - \mathcal{P}(y-1) = \mathcal{P}_\pi. \quad \square$$

4.4 The proof of Proposition 4.1

Proposition 4.1 is an immediate consequence of Theorem 2.3, Proposition 4.2 and the second statement of the following proposition:

Proposition 4.4 *Let N be a 3–manifold with empty or toroidal boundary and let $\alpha\colon \pi_1(N) \to U(k,\mathbb{C})$ be a unitary representation. Then*

$$\mathcal{T}(N,\alpha) \leq \mathcal{P}_N.$$

Furthermore, if N is irreducible, then there exists a unitary representation

$$\alpha\colon \pi_1(N) \to U(k,\mathbb{C})$$

such that

$$\mathcal{T}(N,\alpha) \doteq \mathcal{P}_N.$$

Proof Let N be a 3–manifold with empty or toroidal boundary. We write $\pi = \pi_1(N)$. Let $\alpha\colon \pi \to U(k,\mathbb{C})$ be a unitary representation. If $\tau(N,\alpha) = 0$, then there is nothing to show. So suppose that $\tau(N,\alpha) \neq 0$. In [19, Theorem 1.1; 20, Theorem 3.1] it was shown that for any $\phi \in H^1(N;\mathbb{R}) = \mathrm{Hom}(\pi,\mathbb{R})$ we have

$$\max\{\phi(p) - \phi(q) \mid p, q \in \mathcal{T}(N,\alpha)\} \leq x_N(\phi).$$

It follows from the definitions and the discussion in Section 2.2 that $\mathcal{T}(N,\alpha)^{\mathrm{sym}} \leq \mathcal{P}_N$. Since α is a unitary representation, it follows from [21, Theorem 1.2] that $\mathcal{T}(N,\alpha)^{\mathrm{sym}} \doteq \mathcal{T}(N,\alpha)$. It thus follows that indeed $\mathcal{T}(N,\alpha) \leq \mathcal{P}_N$.

If N is not a closed graph manifold, then, building on Theorem 3.1, it was shown in [26, Corollary 5.10] that there exists a unitary representation $\alpha\colon \pi \to U(k,\mathbb{C})$ such that

$$\max\{\phi(p) - \phi(q) \mid p, q \in \mathcal{T}(N,\alpha)\} = x_N(\phi)$$

for every $\phi \in H^1(N;\mathbb{R})$. The same argument as above then implies that $\mathcal{T}(N,\alpha) \doteq \mathcal{P}_N$. If N is a closed graph manifold, then the same statement holds by [23]. □

5 Proof of Theorem 1.2, II

The goal of this section is to prove the following proposition, and to complete the proof of the main theorem.

Proposition 5.1 *Let $\pi = \langle x, y \mid r\rangle$ be a cyclically reduced $(2,1)$–presentation for the fundamental group of an irreducible 3–manifold N. Then*

$$\mathcal{P}_\pi^{\mathrm{sym}} \leq \mathcal{P}_N.$$

In the proof of Proposition 4.1 we used twisted Reidemeister torsions corresponding to finite-dimensional complex representations. In the proof of Proposition 5.1 we use a different but related object, namely noncommutative Reidemeister torsions. In this context they were first studied in [9; 8; 27; 18].

5.1 The Ore localization of group rings and degrees

Let $\Gamma \in \mathcal{TEA}$. It follows from [33, Theorem 1.4] that the group ring $\mathbb{Z}[\Gamma]$ is a domain. Since Γ is amenable it follows from [11, Corollary 6.3] that $\mathbb{Z}[\Gamma]$ satisfies the Ore condition. This means that for any two nonzero elements $x, y \in \mathbb{Z}[\Gamma]$ there exist nonzero elements $p, q \in \mathbb{Z}[\Gamma]$ such that $xp = yq$. By [41, Section 4.4] this implies that $\mathbb{Z}[\Gamma]$ has a classical fraction field, referred to as the Ore localization of $\mathbb{Z}[\Gamma]$, which we denote by $\mathbb{K}(\Gamma)$.

Let $\phi\colon \Gamma \to \mathbb{Z}$ be a homomorphism. For every nonzero $p = \sum_{g\in\Gamma} p_g g \in \mathbb{Z}[\Gamma]$ we define

$$\deg_\phi(p) = \max\{\phi(g) - \phi(h) \mid p_g \neq 0 \text{ and } p_h \neq 0\}.$$

We extend this to all of $\mathbb{Z}[\Gamma]$ by letting $\deg_\phi(0) = -\infty$. Since $\mathbb{Z}[\Gamma]$ has no nontrivial zero divisors it follows that for $p, q \in \mathbb{Z}[\Gamma]$ we have $\deg_\phi(pq) = \deg_\phi(p) + \deg_\phi(q)$. Given $pq^{-1} \in \mathbb{K}(\Gamma)$ we also define

$$\deg_\phi(pq^{-1}) := \deg_\phi(p) - \deg_\phi(q).$$

It is straightforward to see that this is indeed well-defined.

5.2 Noncommutative Reidemeister torsion of presentations

Let X be a finite CW–complex with $G = \pi_1(X)$, and let $\widetilde{X}$ denote the universal cover of X. As in Section 4.2 we view $C_*(\widetilde{X})$ as a chain complex of left $\mathbb{Z}[G]$–modules. Now let $\varphi\colon G \to \Gamma \in \mathcal{TEA}$ be a homomorphism, and consider the chain complex of left $\mathbb{K}(G)$–modules

$$C_*(X; \mathbb{K}(\Gamma)) = \mathbb{K}(\Gamma) \otimes_{\mathbb{Z}[G]} C_*(\widetilde{X}),$$

where G acts on $\mathbb{K}(\Gamma)$ on the right via the homomorphism φ. If $C_*(X; \mathbb{K}(\Gamma))$ is not acyclic, define the corresponding Reidemeister torsion $\tau(X, \varphi)$ to be zero. Otherwise choose an ordering of the cells of X and for each cell in X pick a lift to $\widetilde{X}$. This turns $C_*(X; \mathbb{K}(\Gamma))$ into a chain complex of based $\mathbb{K}(\Gamma)$ left-modules and we define

$$\tau(X, \varphi) \in K_1(\mathbb{K}(\Gamma))$$

to be the Reidemeister torsion of the based chain complex $C_*(X; \mathbb{K}(\Gamma))$. Here $K_1(\mathbb{K}(\Gamma))$ is the abelianization of the direct limit $\lim_{n\to\infty} \mathrm{GL}(n, \mathbb{K}(\Gamma))$ of the general linear groups over $\mathbb{K}(\Gamma)$ (see [37; 45] for details). We write $\mathbb{K}(\Gamma)^\times = \mathbb{K}(\Gamma) \setminus \{0\}$ and denote by $\mathbb{K}(\Gamma)^\times_{\mathrm{ab}}$ the abelianization of the multiplicative group $\mathbb{K}(\Gamma)^\times$. The Dieudonné determinant (see [45]) gives rise to an isomorphism $K_1(\mathbb{K}(\Gamma)) \to \mathbb{K}(\Gamma)^\times_{\mathrm{ab}}$, which we will use to identify these two groups. The invariant $\tau(X,\varphi) \in \mathbb{K}(\Gamma)^\times$ is well-defined up to multiplication by an element of the form $\pm g$, where $g \in \Gamma$. Furthermore, it does not depend on the homeomorphism type of X. We refer to [52; 18; 28] for details.

It follows from $\deg_\phi(p \cdot q) = \deg(p) + \deg(q)$ for $p, q \in \mathbb{K}(\Gamma)^\times$ that $\deg_\phi$ descends to a homomorphism $\deg_\phi\colon \mathbb{K}(\Gamma)^\times_{\mathrm{ab}} \to \mathbb{Z}$. In particular $\deg_\phi(\tau(X,\varphi))$ is defined.

Proof of Proposition 5.1 Let N be an irreducible 3–manifold and suppose $\pi = \langle x, y \mid r\rangle$ is a cyclically reduced $(2,1)$–presentation of its fundamental group. Without loss of generality we may assume that x represents a nonzero element in $H := H_1(N;\mathbb{Z})/\text{torsion}$. We need to show that $\mathcal{P}_\pi \le \mathcal{P}_N$.

We call $\phi \in \operatorname{Hom}(\pi,\mathbb{R})$ *generic* if there are vertices v and w of $\mathcal{P}(r_y)$ such that ϕ pairs maximally with v and ϕ pairs minimally with w.

Claim *For any generic epimorphism $\phi\colon \pi \to \mathbb{Z}$, we have $\operatorname{th}_{\mathcal{P}_\pi}(\phi) \le x_N(\phi)$.*

We denote by v and w the (necessarily unique) vertices of $\mathcal{P}(r_y)$ such that ϕ pairs maximally with v and minimally with w. By Corollary 3.4 and Theorem 2.3 there exists a homomorphism $\alpha\colon \pi_1(N) \to \Gamma \in \mathcal{TEA}$ such that $\alpha(r_y^v \cdot r_y^w) \ne 0$. In particular, $\alpha(r_y^v) \ne 0$ and $\alpha(r_y^w) \ne 0$. Let $\psi\colon \pi \to H$ denote the canonical epimorphism. After possibly replacing α by $\alpha \times \psi$ we can and will assume that ψ factors through α. In particular ϕ factors through α and $\alpha(x)$ is a nontrivial element in Γ.

We denote by X the CW–complex corresponding to the presentation π with one 0–cell, two 1–cells corresponding to the generators x and y and one 2–cell corresponding to the relator r. As in [24] we have $\tau(N,\alpha) = \tau(X,\alpha)$. We then have

$$\begin{aligned}
\operatorname{th}_{\mathcal{P}_\pi}(\phi) &= \operatorname{th}_{\mathcal{P}(r_y)}(\phi) - \operatorname{th}_{\mathcal{P}(x-1)}(\phi)\\
&= (\phi(v) - \phi(w)) - |\phi(x)|\\
&= \deg_\phi(\alpha(r_y)) - \deg_\phi(\alpha(x) - 1)\\
&= \deg_\phi(\alpha(r_y) \cdot \alpha(x-1)^{-1})\\
&= \deg_\phi(\tau(X,\alpha)) = \deg_\phi(\tau(N,\alpha)) \le x_N(\phi).
\end{aligned}$$

Here the first two equalities follows from the definitions and the choice of v and w. The fifth equality is [18, Theorem 2.1] and the last inequality is given by [18, Theorem 1.2] (see also [8; 27; 53]). This concludes the proof of the claim.

It is straightforward to see that the nongeneric elements in $\operatorname{Hom}(\pi,\mathbb{R})$ correspond to a union of proper subspaces of $\operatorname{Hom}(\pi,\mathbb{R})$. By continuity and linearity of seminorms we see that the inequality $\mathrm{th}_{\mathcal{P}_\pi}(\phi) \le x_N(\phi)$ holds in fact for all $\phi \in \operatorname{Hom}(\pi,\mathbb{R})$. It follows from the definitions and the discussion in Section 2.2 that $\mathcal{P}_N^{\mathrm{sym}} \le \mathcal{P}_N$. □

5.3 Proof of the main theorem

For the reader's convenience we recall the statement of Theorem 1.2.

Theorem 1.2 *Let N be an irreducible 3–manifold that admits a cyclically reduced $(2,1)$–presentation $\pi = \langle x, y \mid r\rangle$. Then*

$$\mathcal{M}_N \doteq \mathcal{M}_\pi.$$

Proof It follows from Propositions 4.1 and 5.1 that $\mathcal{P}_N \le \mathcal{P}_\pi$ and $\mathcal{P}_\pi^{\mathrm{sym}} \le \mathcal{P}_N$. By the symmetry of the Thurston norm we also have $\mathcal{P}_N = \mathcal{P}_N^{\mathrm{sym}}$, and this implies $\mathcal{P}_N \doteq \mathcal{P}_\pi$. The fact that the markings agree is an immediate consequence of [24, Theorem 1.1; 4, Theorem E]. □

6 Examples

Currently there is no geometric characterization of those 3–manifolds whose fundamental group may be presented using only two generators and one relator. Waldhausen's question [55] of whether the rank of the fundamental group equals the Heegaard genus gives the conjectural picture that all of these manifolds have tunnel-number one. Li [34] gives examples of 3–manifolds whose rank is strictly smaller than the genus, including closed manifolds, manifolds with boundary, hyperbolic manifolds, and manifolds with nontrivial JSJ decomposition. See also related work of Boileau, Weidmann and Zieschang [6; 5]. However, Waldhausen's question remains open for hyperbolic 3–manifolds of rank 2 and for knot complements in S^3.

6.1 Tunnel-number one manifolds

A *tunnel-number one* 3–manifold is a 3–manifold obtained by attaching a 2–handle to a 3–dimensional 1–handlebody of genus two. The fundamental group has a presentation with two generators from the handlebody and one relator corresponding to the attaching circle of the 2–handle. Theorem 1.2 allows us to compute the unit ball of the Thurston norm with ease, whilst other methods, such as normal surface theory [51; 10] have limited scope (see [13]). Moreover, with Theorem 1.2 one can easily construct examples with prescribed combinatorics or geometry of the unit ball.

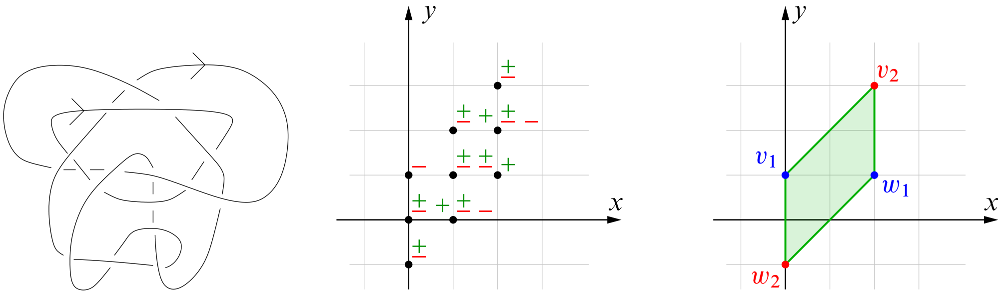


Figure 2: The calculation of $\partial r/\partial x$ (right) for Dunfield's link (left). Center shows the terms appearing in $\partial r/\partial x$ sorted according to their abelianization, with signs as indicated.

Brown's algorithm is an essential ingredient in Dunfield and D Thurston's proof [14] that the probability of a tunnel-number one manifold fibering over the circle is zero. This can be paraphrased as: the probability that the unit ball has a nonempty set of marked vertices is zero. Interesting applications of Theorem 1.2 combined with the methods of [14] would be further predictions about the unit ball of a random tunnel-number one manifold.

6.2 Knots or links in S^3

Norwood [40] showed that if the complement of a given knot in S^3 has fundamental group generated by two elements, then the knot is prime. The complements of *tunnel-number one* knots or links in S^3 are tunnel-number one manifolds. This includes the 2–bridge knots and links, but Johnson [31] showed that there are hyperbolic tunnel-number one knots with arbitrarily high bridge number. There is a complete classification of all tunnel-number one satellite knots by Morimoto and Sakuma [39], and Morimoto [38] also showed that a composite link has tunnel-number one if and only if it is a connected sum of a 2–bridge knot and the Hopf link.

6.3 Dunfield's link

We conclude this section with an explicit calculation for the link L shown in Figure 2, left, which was studied by Dunfield [12]. Write $X_L := S^3 \setminus \nu L$ and write $\pi := \pi_1(X_L)$ for the link group. Then π has the presentation

$$\langle x, y \mid x^2yx^{-1}yx^2yx^{-1}y^{-3}x^{-1}yx^2yx^{-1}yxy^{-1}x^{-2}y^{-1}xy^{-1}x^{-2}y^{-1}xy^3xy^{-1} \cdot x^{-2}y^{-1}xy^{-1}x^{-1}y\rangle,$$

where a meridian for the unknotted component is $y^{-1}x^{-1}yx^2yx^{-1}yx^2yx^{-1}y^{-3}$ and a meridian for the other component is $x^{-1}y^{-1}$. Theorem 1.2 implies that $\mathcal{P}_\pi \doteq \mathcal{P}_N$.

We use the map induced by $x \mapsto (1,0)$ and $y \mapsto (0,1)$ to identify $H_1(X_L;\mathbb{Z}) = H_1(\pi;\mathbb{Z})$ with $\mathbb{Z}^2$.

A straightforward calculation shows that $\mathcal{P}(r_x)$ is the polytope with vertices $v_1 = (0,1)$, $v_2 = (2,3)$, $w_1 = (2,1)$ and $w_2 = (0,-1)$ shown in Figure 2, right. Here v_1 and w_1 are opposite vertices of $\mathcal{P}(r_x)$ and v_2 and w_2 are opposite vertices of $\mathcal{P}(r_x)$. Subtracting the underlying polytope of $\mathcal{M}(y-1)$ from $\mathcal{P}(r_x)$ gives $\mathcal{P}_\pi$, and this agrees (up to translation) with Figure 1. The following computation shows that the markings are the same:

$$\begin{aligned}
(r_x)^{v_1} &= x^2yx^{-1}yx^2yx^{-1}y^{-3}x^{-1}yx^2yx^{-1}yxy^{-1}x^{-2}y^{-1}xy^{-1}x^{-2}y^{-1}xy^3 \\
&\qquad \cdot xy^{-1}x^{-2}, \\
(r_x)^{w_1} &= x^2yx^{-1}yx^2yx^{-1}y^{-3}x^{-1}yx, \\
(r_x)^{v_2} &= x^2yx^{-1}yx^2yx^{-1}(-1+y^{-3}x^{-1}yx^2yx^{-1}y), \\
(r_x)^{w_2} &= x^2yx^{-1}yx^2yx^{-1}y^{-3}x^{-1}yx^2yx^{-1}yxy^{-1}x^{-2}y^{-1}xy^{-1}x^{-2}y^{-1} \\
&\qquad \cdot (1-xy^3xy^{-1}x^{-2}y^{-1}xy^{-1}x^{-1}).
\end{aligned}$$

7 A conjecture and a question

7.1 A conjecture

We conjecture that Poincaré duality for the 3–manifold can be seen on the level of group presentations as follows:

Conjecture 7.1 *Let $\pi = \langle x, y \mid r \rangle$ be a $(2,1)$–presentation for the fundamental group of a 3–manifold. Then there exists $u \in \left(\frac{1}{2}\mathbb{Z}\right)^2$ such that for any vertex v of $\mathcal{P}(r_x)$ the reflection of v in u, ie the point $w = u-(v-u) = 2u - v$, is also a vertex of $\mathcal{P}(r_x)$. Furthermore we have*

$$(r_x)^v \equiv (-1)^{b_0(\partial N)-1}\,\overline{(r_x)^w}.$$

The twisted Reidemeister torsions of [25] can be computed in terms of Fox derivatives, and the symmetry results for twisted Reidemeister torsions proved in [32; 30; 21] give strong evidence towards this conjecture. Also note that if π is a geometric presentation, ie if it comes the presentation given by a genus-2 handlebody with a 1–handle attached, then r is palindromic, ie reads the same forward and backward (see eg [15, Section 5.2]), and then it is elementary to verify that the conjecture holds.

To give an explicit example, let us return to Dunfield's link. Given the group G and $p, q \in \mathbb{Z}[G]$, write $p \equiv q$ if there exist $g, h \in G$ such that $p = gqh$. Furthermore, denote by $p \mapsto \overline{p}$ the involution of $\mathbb{Z}[G]$ defined by the inversion map $g \mapsto g^{-1}$ for each $g \in G$. We denote by $\pi = \langle x, y \mid r \rangle$ the presentation from Section 6.3. We then note that

$$(r_x)^{v_2} \equiv -1 + y^{-3}x^{-1}yx^2yx^{-1}y(r_x)^{w_2} \equiv 1 - xy^3xy^{-1}x^{-2}y^{-1}xy^{-1}x^{-1}.$$

The relator r is conjugate to

$$yx^2yx^{-1}yx^2yx^{-1}(y^{-3}x^{-1}yx^2yx^{-1}y)xy^{-1}x^{-2}y^{-1}xy^{-1}x^{-2}y^{-1} \cdot (xy^3xy^{-1}x^{-2}y^{-1}xy^{-1}x^{-1}).$$

In particular writing $s = yx^2yx^{-1}yx^2yx^{-1}$ we have the following equality in $\mathbb{Z}[\pi]$:

$$\begin{aligned}(r_x)^{v_2} &\equiv s(r_y)^{v_2}s^{-1} = s(-1 + y^{-3}x^{-1}yx^2yx^{-1}y)s^{-1} \\ &= -1 + (xy^3xy^{-1}x^{-2}y^{-1}xy^{-1}x^{-1})^{-1} \\ &= -\overline{(r_x)^{w_2}}.\end{aligned}$$

7.2 A question

We initially attempted to prove Theorem 1.2 just using twisted Reidemeister torsions corresponding to finite-dimensional representations, noting that Theorem 1.2 follows from the first part of Proposition 4.4 together with an affirmative answer to the following question, which is interesting in its own right.

Question 7.2 *Let N be an aspherical 3–manifold and write $\pi = \pi_1(N)$. Let p be a nonzero element in $\mathbb{Z}[\pi]$. Does there exist a representation $\alpha\colon \pi \to \mathrm{GL}(k, \mathbb{C})$ such that* $\det(\alpha(f)) \neq 0$*?*

References

[1] **I Agol**, *Criteria for virtual fibering*, J. Topol. 1 (2008) 269–284 MR

[2] **I Agol**, *The virtual Haken conjecture*, Doc. Math. 18 (2013) 1045–1087 MR

[3] **M Aschenbrenner**, **S Friedl**, **H Wilton**, *3–manifold groups*, Eur. Math. Soc., Zürich (2015) MR

[4] **R Bieri**, **W D Neumann**, **R Strebel**, *A geometric invariant of discrete groups*, Invent. Math. 90 (1987) 451–477 MR

[5] **M Boileau**, **R Weidmann**, *The structure of 3–manifolds with two-generated fundamental group*, Topology 44 (2005) 283–320 MR

[6] **M Boileau**, **H Zieschang**, *Heegaard genus of closed orientable Seifert 3–manifolds*, Invent. Math. 76 (1984) 455–468 MR

[7] **S V Buyalo**, **P V Svetlov**, *Topological and geometric properties of graph manifolds*, Algebra i Analiz 16 (2004) 3–68 MR In Russian; translated in St. Petersburg Math. J. 16 (2005) 297–340

[8] **T D Cochran**, *Noncommutative knot theory*, Algebr. Geom. Topol. 4 (2004) 347–398 MR

[9] **T D Cochran**, **K E Orr**, **P Teichner**, *Knot concordance, Whitney towers and L^2–signatures*, Ann. of Math. 157 (2003) 433–519 MR

[10] **D Cooper**, **S Tillmann**, *The Thurston norm via normal surfaces*, Pacific J. Math. 239 (2009) 1–15 MR

[11] **J Dodziuk**, **P Linnell**, **V Mathai**, **T Schick**, **S Yates**, *Approximating L^2–invariants and the Atiyah conjecture*, Comm. Pure Appl. Math. 56 (2003) 839–873 MR

[12] **N M Dunfield**, *Alexander and Thurston norms of fibered 3–manifolds*, Pacific J. Math. 200 (2001) 43–58 MR

[13] **N M Dunfield**, **D Ramakrishnan**, *Increasing the number of fibered faces of arithmetic hyperbolic 3–manifolds*, Amer. J. Math. 132 (2010) 53–97 MR

[14] **N M Dunfield**, **D P Thurston**, *A random tunnel number one 3–manifold does not fiber over the circle*, Geom. Topol. 10 (2006) 2431–2499 MR

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

[16] **D B A Epstein**, *Finite presentations of groups and 3–manifolds*, Quart. J. Math. Oxford Ser. 12 (1961) 205–212 MR

[17] **R H Fox**, *Free differential calculus, I: Derivation in the free group ring*, Ann. of Math. 57 (1953) 547–560 MR

[18] **S Friedl**, *Reidemeister torsion, the Thurston norm and Harvey's invariants*, Pacific J. Math. 230 (2007) 271–296 MR

[19] **S Friedl**, **T Kim**, *The Thurston norm, fibered manifolds and twisted Alexander polynomials*, Topology 45 (2006) 929–953 MR

[20] **S Friedl**, **T Kim**, *Twisted Alexander norms give lower bounds on the Thurston norm*, Trans. Amer. Math. Soc. 360 (2008) 4597–4618 MR

[21] **S Friedl**, **T Kim**, **T Kitayama**, *Poincaré duality and degrees of twisted Alexander polynomials*, Indiana Univ. Math. J. 61 (2012) 147–192 MR

[22] **S Friedl**, **T Kitayama**, *The virtual fibering theorem for 3–manifolds*, Enseign. Math. 60 (2014) 79–107 MR

[23] **S Friedl**, **M Nagel**, *Twisted Reidemeister torsion and the Thurston norm: graph manifolds and finite representations*, Illinois J. Math. 59 (2015) 691–705 MR

[24] **S Friedl**, **S Tillmann**, *Two-generator one-relator groups and marked polytopes*, preprint (2015) arXiv

[25] **S Friedl**, **S Vidussi**, *A survey of twisted Alexander polynomials*, from "The mathematics of knots: theory and application" (M Banagl, D Vogel, editors), Contrib. Math. Comput. Sci. 1, Springer (2011) 45–94 MR

[26] **S Friedl**, **S Vidussi**, *The Thurston norm and twisted Alexander polynomials*, J. Reine Angew. Math. 707 (2015) 87–102 MR

[27] **S L Harvey**, *Higher-order polynomial invariants of 3–manifolds giving lower bounds for the Thurston norm*, Topology 44 (2005) 895–945 MR

[28] **S Harvey**, **S Friedl**, *Non-commutative multivariable Reidemester torsion and the Thurston norm*, Algebr. Geom. Topol. 7 (2007) 755–777 MR

[29] **J Hempel**, *Residual finiteness for 3–manifolds*, from "Combinatorial group theory and topology" (S M Gersten, J R Stallings, editors), Ann. of Math. Stud. 111, Princeton Univ. Press (1987) 379–396 MR

[30] **J A Hillman**, **D S Silver**, **S G Williams**, *On reciprocality of twisted Alexander invariants*, Algebr. Geom. Topol. 10 (2010) 1017–1026 MR

[31] **J Johnson**, *Bridge number and the curve complex*, preprint (2006) arXiv

[32] **T Kitano**, *Twisted Alexander polynomial and Reidemeister torsion*, Pacific J. Math. 174 (1996) 431–442 MR

[33] **P H Kropholler**, **P A Linnell**, **J A Moody**, *Applications of a new K–theoretic theorem to soluble group rings*, Proc. Amer. Math. Soc. 104 (1988) 675–684 MR

[34] **T Li**, *Rank and genus of 3–manifolds*, J. Amer. Math. Soc. 26 (2013) 777–829 MR

[35] **P Linnell**, **T Schick**, *Finite group extensions and the Atiyah conjecture*, J. Amer. Math. Soc. 20 (2007) 1003–1051 MR

[36] **Y Liu**, *Virtual cubulation of nonpositively curved graph manifolds*, J. Topol. 6 (2013) 793–822 MR

[37] **J Milnor**, *Whitehead torsion*, Bull. Amer. Math. Soc. 72 (1966) 358–426 MR

[38] **K Morimoto**, *On composite tunnel number one links*, Topology Appl. 59 (1994) 59–71 MR

[39] **K Morimoto**, **M Sakuma**, *On unknotting tunnels for knots*, Math. Ann. 289 (1991) 143–167 MR

[40] **F H Norwood**, *Every two-generator knot is prime*, Proc. Amer. Math. Soc. 86 (1982) 143–147 MR

[41] **D S Passman**, *The algebraic structure of group rings*, Wiley, New York (1977) MR

[42] **P Przytycki**, **D Wise**, *Mixed 3–manifolds are virtually special*, preprint (2012) arXiv

[43] **P Przytycki**, **D T Wise**, *Graph manifolds with boundary are virtually special*, J. Topol. 7 (2014) 419–435 MR

[44] **L Ribes**, **P Zalesskii**, *Profinite groups*, 2nd edition, Ergeb. Math. Grenzgeb. 40, Springer (2010) MR

[45] **J Rosenberg**, *Algebraic K–theory and its applications*, Graduate Texts in Mathematics 147, Springer (1994) MR

[46] **K Schreve**, *The strong Atiyah conjecture for virtually cocompact special groups*, Math. Ann. 359 (2014) 629–636 MR

[47] **J-P Serre**, *Galois cohomology*, Springer (1997) MR

[48] **J-C Sikorav**, *Novikov homology and three-manifolds*, Oberwolfach Reports 4 (2007) 2334–2336

[49] **W P Thurston**, *A norm for the homology of 3–manifolds*, Mem. Amer. Math. Soc. 339, Amer. Math. Soc., Providence, RI (1986) 99–130 MR

[50] **D Tischler**, *On fibering certain foliated manifolds over S^1*, Topology 9 (1970) 153–154 MR

[51] **J L Tollefson**, **N Wang**, *Taut normal surfaces*, Topology 35 (1996) 55–75 MR

[52] **V Turaev**, *Introduction to combinatorial torsions*, Birkhäuser, Basel (2001) MR

[53] **V Turaev**, *A homological estimate for the Thurston norm*, preprint (2002) arXiv

[54] **M Wada**, *Twisted Alexander polynomial for finitely presentable groups*, Topology 33 (1994) 241–256 MR

[55] **F Waldhausen**, *Some problems on 3–manifolds*, from "Algebraic and geometric topology" (R J Milgram, editor), Proc. Sympos. Pure Math. 32, Amer. Math. Soc., Providence, RI (1978) 313–322 MR

[56] **D T Wise**, *Research announcement: the structure of groups with a quasiconvex hierarchy*, Electron. Res. Announc. Math. Sci. 16 (2009) 44–55 MR

[57] **D T Wise**, *From riches to raags: 3–manifolds, right-angled Artin groups, and cubical geometry*, CBMS Regional Conference Series in Mathematics 117, Amer. Math. Soc., Providence, RI (2012) MR

[58] **D Wise**, *The structure of groups with a quasi-convex hierarchy*, preprint (2012) Available at `http://www.math.mcgill.ca/wise/papers.html`

Fakultät für Mathematik, Universität Regensburg
Regensburg, Germany

Department of Mathematics, University of Michigan
Ann Arbor, MI, United States

School of Mathematics and Statistics, The University of Sydney
Sydney, NSW, Australia

`sfriedl@gmail.com`, `schreve@umich.edu`, `stephan.tillmann@sydney.edu.au`

Proposed: Ian Agol
Seconded: Peter Teichner, András I. Stipsicz

Received: 5 June 2016
Revised: 25 October 2016

Geometry & Topology 21 (2017) 3785–3810

The L^p–diameter of the group of area-preserving diffeomorphisms of S^2

MICHAEL BRANDENBURSKY
EGOR SHELUKHIN

We show that for each $p \geq 1$, the L^p–metric on the group of area-preserving diffeomorphisms of the two-sphere has infinite diameter. This solves the last open case of a conjecture of Shnirelman from 1985. Our methods extend to yield stronger results on the large-scale geometry of the corresponding metric space, completing an answer to a question of Kapovich from 2012. Our proof uses configuration spaces of points on the two-sphere, quasimorphisms, optimally chosen braid diagrams, and, as a key element, the cross-ratio map $X_4(\mathbb{C}P^1) \to \mathcal{M}_{0,4} \cong \mathbb{C}P^1 \setminus \{\infty, 0, 1\}$ from the configuration space of 4 points on $\mathbb{C}P^1$ to the moduli space of complex rational curves with 4 marked points.

20F65, 37E30, 53D99; 20F36, 57M07, 57R50, 57S05

1 Introduction and main results

1.1 Introduction

The L^2–length of a path of volume-preserving diffeomorphisms, which describes a time-dependent flow of an ideal incompressible fluid, corresponds to the hydrodynamic action of the flow in the same way as the length of a path in a Riemannian manifold corresponds to its energy; see Shnirelman [41]. Indeed, it is the length of this path with respect to the formal right-invariant Riemannian metric on the group $\mathcal{G}$ of volume-preserving diffeomorphisms introduced by Arnold in [1]. The L^1–length of the same path has a dynamical interpretation as the average length of a trajectory of a point under the flow.

Following the principle of least action, it therefore makes sense to consider the infimum of the lengths of paths connecting two fixed volume-preserving diffeomorphisms. This gives rise to a right-invariant distance function (metric) on $\mathcal{G}$. Taking the identity transformation as the initial point, Arnold observes that a path whose L^2–length is minimal (and equal to the distance) necessarily solves the Euler equation of an ideal incompressible fluid.

Published: 31 August 2017 DOI: 10.2140/gt.2017.21.3785

It follows from works of Ebin and Marsden [18] that for diffeomorphisms in $\mathcal{G}$ that are C^2–close to the identity, the infimum is indeed achieved. Further, more global results on the corresponding Riemannian exponential map were obtained in Ebin, Misiołek and Preston [19] and Shnirelman [42]; see Ebin [17] for additional references. Shnirelman [40; 41] showed, among a number of surprising facts related to this subject, that in the case of the ball of dimension 3, the diameter of the L^2–metric is bounded. This result is known[1] to hold for all compact simply connected manifolds of dimension 3 or larger (see Eliashberg and Ratiu [20], Khesin and Wendt [29] and Arnold and Khesin [3]), while its analogue in the non–simply connected case is false (see Eliashberg and Ratiu [20] and Brandenbursky [9]). Furthermore, Shnirelman has conjectured that for compact manifolds of dimension 2, the L^2–diameter is infinite.

In this paper we consider Shnirelman's conjecture, and its analogues for L^p–metrics, with $p \geq 1$. It follows from results of [20] that on compact surfaces (possibly with boundary) other than T^2 and S^2, Shnirelman's conjecture holds for all $p \geq 1$. Their arguments rely on the Calabi homomorphism Cal (see Calabi [14]) from the compactly supported Hamiltonian group $\mathrm{Ham}_c(\Sigma, \sigma)$ to the real numbers in the case of a surface Σ with nonempty boundary (σ is the area form), and on nontrivial first cohomology combined with trivial center of the fundamental group in the closed case. For the two-torus T^2 this conjecture of Shnirelman also holds for all $p \geq 1$, as can be quickly seen by the following steps. First, the methods of [20] together with the fact that the Hamiltonian group $\mathrm{Ham}(T^2, dx \wedge dy)$ is simply connected as a topological space (see eg Polterovich [35, Chapter 7.2.B]) imply that $\mathrm{Ham}(T^2, dx \wedge dy)$ with the L^p–metric has infinite diameter (compare with Brandenbursky and Kędra [11, Theorem 1.2]). Second, the inclusion $\mathrm{Ham}(T^2, dx \wedge dy) \hookrightarrow \mathrm{Diff}_0(T^2, dx \wedge dy)$, the two groups being equipped with their respective L^p–metrics, is a quasi-isometry (see Proposition A.1). The case of the two-sphere S^2, to which previous methods do not apply, remained open.

The case $p > 2$ (but not that of Shnirelman's original conjecture!) is well-known, as it follows from a result of Polterovich [34] regarding Hofer's metric on $\mathrm{Ham}(S^2)$ by an application of the Sobolev inequality. The authors gave a different proof of this case by elementary methods in the preprint [12].

The main result of this paper is the unboundedness of the L^p–metric on $\mathrm{Ham}(S^2)$ for all $p \geq 1$. This completes a full answer to Shnirelman's question. Our methods extend to yield stronger results on the large-scale geometry on the L^p–metric on $\mathrm{Ham}(S^2)$. In particular, we provide bi-Lipschitz group monomorphisms of $\mathbb{R}^m$ endowed with the standard (say Euclidean) metric into $(\mathrm{Ham}(S^2), d_{L^p})$ for each positive integer m

[1] The authors have not found a detailed proof of this generalization in the literature.

and each $p \geq 1$. Moreover, our key technical estimate implies by an argument of Kim and Koberda [30] (cf Crisp and Wiest [16] and Benaim and Gambaudo [5]) the existence of quasi-isometric group monomorphisms from each right-angled Artin group to $(\mathrm{Ham}(S^2), d_{L^p})$ for each $p \geq 1$, completing the resolution of a question of Kapovich [27] in the case of S^2 (the case $p > 2$ shown in Kim and Koberda uses [12]).

Our methods are two-dimensional in nature, and have to do with braiding and relative rotation numbers of trajectories of time-dependent two-dimensional Hamiltonian flows (in extended phase space). We note that Shnirelman has proposed to use relative rotation numbers to bound from below the L^2–lengths of two-dimensional Hamiltonian paths in [41]. This direction is related to the method of Eliashberg and Ratiu by a theorem of Gambaudo and Ghys [22] and Fathi [21] (compare with Shelukhin [39]), stating that the Calabi homomorphism is proportional to the relative rotation number of the trajectories of two distinct points in the two-disc $\mathbb{D}$ under a Hamiltonian flow, averaged over the configuration space of ordered pairs of distinct points (x_1, x_2) in the two-disc.

This line of research was pursued in Gambaudo and Lagrange [24], Benaim and Gambaudo [5], Crisp and Wiest [16], Brandenbursky [9], Brandenbursky and Kędra [10] and Kim and Koberda [30], obtaining quasi-isometric and bi-Lipschitz embeddings of various groups (right-angled Artin groups and additive groups of finite-dimensional real vector spaces) into $\mathrm{Ham}_c(\mathbb{D}^2, dx \wedge dy)$ and into $\ker(\mathrm{Cal}) \subset \mathrm{Ham}_c(\mathbb{D}^2, dx \wedge dy)$ endowed with their respective L^p–metrics (see Brandenbursky and Kędra [11] for similar embedding results on manifolds with a sufficiently complicated fundamental group). In all cases, the key technical estimate is an upper bound, via the L^p–length of an isotopy of volume-preserving diffeomorphisms, of the average, over all points in a configuration space of the manifold, of the word length in the fundamental group of the configuration space of the trace of the point under the induced isotopy (closed up to a loop by a system of short paths on the configuration space). In this paper we produce similar estimates for the case of the two-sphere. Our case of $\mathrm{Diff}_0(S^2, \sigma)$, with $p \leq 2$, is more difficult than that of $\ker(\mathrm{Cal}) \subset \mathrm{Ham}_c(\mathbb{D}, dx \wedge dy)$ because the necessary analytical and topological bounds require a more global approach and have to take into account the geometry and topology of the sphere.

In turn, lower bounds on the average word length can often be provided by quasimorphisms — functions that are additive with respect to the group multiplication — up to an error which is uniformly bounded (as a function of two variables). The quasimorphisms we use were introduced and studied by Gambaudo and Ghys in the beautiful paper [23]; see also Polterovich [36], Py [37; 38] and Brandenbursky [8]. These quasimorphisms essentially appear from invariants of braids traced out by the action of a Hamiltonian path on an ordered n–tuple of distinct points in the surface (suitably closed up), averaged over the configuration space $X_n(\Sigma)$ of n–tuples of distinct points on the surface Σ.

Comparison with [12] The first step in our study of Shnirelman's conjecture is found in the unpublished preprint [12], where we saw which elements of the approach of [9] extend to the case of S^2. There we found that without a key new idea one could only obtain the necessary estimates for $p > 2$. The main novelty of this paper consists indeed of a new geometric idea, which is of independent interest. To wit, we introduce certain canonical "logarithmic" differential forms on $X_n(\mathbb{C}P^1)$, which play a key role in our arguments. These forms can be considered as analogues for the case of $\mathbb{C}P^1$ of the differential forms of Arnold [2] on $X_n(\mathbb{C})$. One curious aspect of these forms is that while in Arnold's case they appeared from pairs of points, that is from the natural projections $X_n(\mathbb{C}) \to X_2(\mathbb{C})$ on pairs of coordinates, in the case of $\mathbb{C}P^1$ they are constructed from quadruples of points, that is from projections $X_n(\mathbb{C}P^1) \to X_4(\mathbb{C}P^1)$ on quadruples of coordinates. This fits with $P_2(\mathbb{C}) = \pi_1(X_2(\mathbb{C})) \cong \pi_1(\mathbb{C} \setminus \{0\}) = \mathbb{Z}$ and $P_4(\mathbb{C}P^1) = \pi_1(X_4(\mathbb{C}P^1)) \cong \mathbb{Z}/2\mathbb{Z} \times \pi_1(\mathbb{C} \setminus \{0,1\}) = \mathbb{Z}/2\mathbb{Z} \times (\mathbb{Z} * \mathbb{Z})$ being the first infinite pure braid groups in the two cases.

1.2 Preliminaries

1.2.1 The L^p–metric Let M denote a smooth oriented manifold without boundary that is either closed, or has $M = X \setminus \partial X$ for a compact manifold X. Let M be endowed with a Riemannian metric g and smooth measure μ (given by a volume form, which in our case that M is a surface is an area form σ, and orientation on M). We require g and μ to extend continuously to X in the second case. Finally denote by

$$\mathcal{G} = \mathrm{Diff}_{c,0}(M, \mu)$$

the identity component of the group of compactly supported diffeomorphisms of M preserving the smooth measure μ.

Fix $p \geq 1$. For a smooth isotopy $\{\phi_t\}_{t \in [0,1]}$ from $\phi_0 = 1$ to $\phi_1 = \phi$, we define the L^p–length by

$$l_p(\{\phi_t\}) = \int_0^1 \left(\frac{1}{\mathrm{vol}(M,\mu)} \cdot \int_M |X_t|^p \, d\mu \right)^{1/p} dt,$$

where $X_t = \frac{d}{dt'}\big|_{t'=t} \phi_{t'} \circ \phi_t^{-1}$ is the time-dependent vector field generating the isotopy $\{\phi_t\}$, and $|X_t|$ is its length with respect to the Riemannian structure on M. As is easily seen by a displacement argument, the L^p–length functional determines a nondegenerate norm on $\mathcal{G}$ by the formula

$$d_p(\mathbf{1}, \phi) = \inf l_p(\{\phi_t\}).$$

This in turn defines a right-invariant metric on $\mathcal{G}$ by the formula

$$d_p(\phi_0,\phi_1)=d_p(\mathbf{1},\phi_1{\phi_0}^{-1}).$$

Remark 1.1 Consider the case $p=1$. It is easy to see that the L^1–length of an isotopy is equal to the average Riemannian length of the trajectory $\{\phi_t(x)\}_{t\in[0,1]}$ (over $x\in M$, with respect to μ). Moreover for each $p\geq 1$, by Jensen's (or Hölder's) inequality, we have

$$l_p(\{\phi_t\})\geq l_1(\{\phi_t\}).$$

Denote by $\widetilde{\mathbf{1}}$ the identity element of the universal cover $\widetilde{\mathcal{G}}$ of $\mathcal{G}$. Similarly one has the L^p–pseudonorm (that induces the right-invariant L^p–pseudometric) on $\widetilde{\mathcal{G}}$, defined for $\widetilde{\phi}\in\widetilde{\mathcal{G}}$ as

$$d_p(\widetilde{\mathbf{1}},\widetilde{\phi})=\inf l_p(\{\phi_t\}),$$

where the infimum is taken over all paths $\{\phi_t\}$ in the class of $\widetilde{\phi}$. Clearly $d_p(\mathbf{1},\phi)=\inf d_p(\widetilde{\mathbf{1}},\widetilde{\phi})$, where the infimum runs over all $\widetilde{\phi}\in\widetilde{\mathcal{G}}$ that map to ϕ under the natural epimorphism $\widetilde{\mathcal{G}}\to\mathcal{G}$.

Up to bi-Lipschitz equivalence of metrics (d and d' are equivalent if $d/C\leq d'\leq Cd$ for a certain constant $C>0$) the L^p–metric on $\mathcal{G}$ (and its pseudometric analogue on $\widetilde{\mathcal{G}}$) is independent of the choice Riemannian structure and of the volume form μ on M. In particular, the question of boundedness or unboundedness of the L^p–metric enjoys the same invariance property.

Terminology For a positive integer n, we use $A,B,C>0$ as generic notation for positive constants that depend only on M, μ, g and n.

1.2.2 Quasimorphisms For some of our results, we require the notion of a quasimorphism. Quasimorphisms are a helpful tool for the study of nonabelian groups, especially those that admit few homomorphisms to $\mathbb{R}$. A quasimorphism $r\colon G\to\mathbb{R}$ on a group G is a real-valued function that satisfies

$$r(xy)=r(x)+r(y)+b_r(x,y),$$

for a function $b_r\colon G\times G\to\mathbb{R}$ that is uniformly bounded:

$$\delta(r):=\sup_{G\times G}|b_r|<\infty.$$

A quasimorphism $\overline{r}\colon G\to\mathbb{R}$ is called *homogeneous* if $\overline{r}(x^k)=k\overline{r}(x)$ for all $x\in G$ and $k\in\mathbb{Z}$. In this case, it is additive on each pair $x,y\in G$ of commuting elements: $r(xy)=r(x)+r(y)$ if $xy=yx$.

For each quasimorphism $r\colon G \to \mathbb{R}$ there exists a unique homogeneous quasimorphism $\overline{r}$ that differs from r by a bounded function:

$$\sup_G |\overline{r} - r| < \infty.$$

It is called the *homogenization* of r and satisfies

$$\overline{r}(x) = \lim_{n\to\infty} \frac{r(x^n)}{n}.$$

Denote by $Q(G)$ the real vector space of homogeneous quasimorphisms on G.

For a finitely generated group G, with finite symmetric generating set S, define the word norm $|\cdot|_S\colon G \to \mathbb{Z}_{\geq 0}$ by

$$|g|_S = \min\{k \mid g = s_1 \cdots s_k \text{ for some } s_1, \dots, s_k \in S\}$$

for $g \in G$. This is a norm on G, and as such it induces a right-invariant metric $d_S\colon G \times G \to \mathbb{Z}_{\geq 0}$ by $d_S(f, g) = |gf^{-1}|_S$. This metric is called the word metric. In this setting, any quasimorphism $r\colon G \to \mathbb{R}$ is controlled by the word norm. Indeed, for all $g \in G$,

$$|r(g)| \leq \Big(\delta(r) + \max_{s\in S}|r(s)|\Big) \cdot |g|_S.$$

When a specific symmetric generating set S for G can be fixed, we will usually denote $|\cdot|_S$ by $|\cdot|_G$.

We refer to [15] for more information about quasimorphisms.

1.2.3 Configuration spaces and braid groups For a manifold M, which shall in this paper be usually of dimension 2 and without boundary, the configuration space $X_n(M) \subset M^n$ of n–tuples of points on M is defined as

$$X_n(M) = \{(x_1, \dots, x_n) \mid x_i \neq x_j,\ 1 \leq i < j \leq n\}.$$

That is,

$$X_n(M) = M^n \setminus \bigcup_{1\leq i<j\leq n} D_{ij},$$

where the partial diagonal $D_{ij} \subset M^n$ is defined as $D_{ij} = \{(x_1, \dots, x_n) \mid x_i = x_j\}$ for $1 \leq i < j \leq n$. Note that D_{ij} is a submanifold of M^n of codimension $\dim M$. When $\dim M = 2$ and M is endowed with a complex structure, D_{ij} is a complex hypersurface. Therefore we shall sometimes refer to D_{ij} and $D = \bigcup_{1\leq i<j\leq n} D_{ij}$ as divisors. Indeed, complex coordinates serve an important role in our arguments.

Finally we define the pure braid group of M as

$$P_n(M) = \pi_1(X_n(M)).$$

Noting that the symmetric group S_n on n elements acts on $X_n(M)$, we form the quotient $C_n(M) = X_n(M)/S_n$ and define the full braid group of M as

$$B_n(M) = \pi_1(C_n(M)).$$

For smooth surfaces M endowed with a complex structure (hence smooth complex manifolds of complex dimension 1), $C_n(M)$ turns out to inherit the structure of a smooth complex manifold of complex dimension n.

We note that $P_n(M)$ and $B_n(M)$ enter the exact sequence

$$1 \to P_n(M) \to B_n(M) \to S_n \to 1.$$

In particular $P_n(M)$ is a normal subgroup of $B_n(M)$ of finite index. We refer to Kassel and Turaev [28] for further information about braid groups.

1.2.4 Short paths and the Gambaudo–Ghys construction Given a real-valued quasimorphism r on $P_n(M) = \pi_1(X_n(M), q)$, for a fixed basepoint $q \in X_n(M)$ there is a natural way to construct a real-valued quasimorphism on the universal cover $\widetilde{\mathcal{G}}$ of the group $\mathcal{G} = \mathrm{Diff}_0(S^2, \sigma)$ of area-preserving diffeomorphisms of $M = S^2$. We shall see that in our case of $M = S^2$ this induces a quasimorphism on $\mathcal{G}$ itself, because the fundamental group of $\mathcal{G}$ is finite. The construction is carried out by the following steps; see Gambaudo and Ghys [23], Polterovich [36] and Brandenbursky [8].

Step 1 For all $x \in X_n(S^2) \setminus Z$ with Z a closed negligible subset (eg a union of submanifolds of positive codimension), choose a smooth path $\gamma(x)\colon [0,1] \to X_n(S^2)$ between the basepoint $q \in X_n(S^2)$ and x. Make this choice continuous in $X_n(S^2) \setminus Z$. We first choose a system of paths on $M = S^2$ itself, in our case the minimal geodesics with respect to the round metric, and then consider the induced coordinate-wise paths in M^n, and pick Z to ensure that these induced paths actually lie in $X_n(S^2)$. After choosing the system of paths $\{\gamma(x)\}_{x \in X_n(S^2)\setminus Z}$ we extend it measurably to all $x \in X_n(S^2)$ (obviously, no numerical values computed in the paper will depend on this extension). We call the resulting choice a "system of short paths".

Step 2 Given a path $\{\phi_t\}_{t\in[0,1]}$ in $\mathcal{G}$ starting at Id, and a point $x \in X_n(S^2)$, consider the path $\{\phi_t \cdot x\}$, to which we then concatenate the corresponding short paths. That is, consider the loop

$$\lambda(x, \{\phi_t\}) := \gamma(x) \,\#\, \{\phi_t \cdot x\} \,\#\, \gamma(y)^{-1}$$

in $X_n(S^2)$ based at q, where $^{-1}$ denotes time reversal, and $y = \phi_1 \cdot x$. Hence we obtain for each $x \in X_n(S^2) \setminus Z \cap (\phi_1)^{-1}(Z)$ an element $[\lambda(x, \{\phi_t\})] \in \pi_1(X_n(S^2), q)$ (or rather for each $x \in X_n(S^2)$ after the measurable extension in Step 1).

Step 3 Consequently applying the quasimorphism $r\colon \pi_1(X_n(S^2), q) \to \mathbb{R}$ we obtain a measurable function $f\colon X_n(S^2) \to \mathbb{R}$. Namely, $f(x) = r([\lambda(x, \{\phi_t\})])$. The quasimorphism Φ on $\widetilde{\mathcal{G}}$ is defined by

$$\Phi([\{\phi_t\}]) = \int_{X_n(S^2)} f \, d\mu^{\otimes n}.$$

It is immediate to see that this function is well-defined for topological reasons. The quasimorphism property follows by the quasimorphism property of r combined with finiteness of volume. The fact that the function f is absolutely integrable can be shown to hold a priori by a reduction to the case of the disc. We note, however, that by Tonelli's theorem this fact follows as a byproduct of the proof of our main theorem, and therefore requires no additional proof.

Step 4 Of course our quasimorphism can be homogenized, to obtain a homogeneous quasimorphism $\overline{\Phi}$.

Remark 1.2 In our case, by the result of Smale [43], $\pi_1(\mathcal{G}) = \mathbb{Z}/2\mathbb{Z}$, and hence the quasimorphisms descend to quasimorphisms on $\mathcal{G}$, eg by minimizing over the two-element fibers of the projection $\widetilde{\mathcal{G}} \to \mathcal{G}$. For $\overline{\Phi}$, the situation is easier since by homogeneity it vanishes on $\pi_1(\mathcal{G}) \subset Z(\widetilde{\mathcal{G}})$, and therefore depends only on the image in $\mathcal{G}$ of an element in $\widetilde{\mathcal{G}}$. We keep the same notation for the induced quasimorphisms.

1.2.5 The cross-ratio map Recall that S^2 can be identified with $\mathbb{C}P^1$, and the latter has an affine chart $u_0\colon \mathbb{C} \to \mathbb{C}P^1$ with $u_0(z) = [z, 1]$ in homogeneous coordinates, whose image is the complement of the point $\infty := [1, 0]$.

The cross-ratio map is given by the natural[2] projection

$$X_4(\mathbb{C}P^1) \to \mathcal{M}_{0,4} = X_4(\mathbb{C}P^1)/\mathrm{PSL}(2, \mathbb{C}).$$

Composing it with the isomorphism $\mathcal{M}_{0,4} \cong \mathbb{C}P^1 \setminus \{\infty, 0, 1\} \cong \mathbb{C} \setminus \{0, 1\}$ given by the inverse of the map $u \mapsto [(\infty, 0, 1, u)]$, we obtain a map

$$\mathrm{cr}\colon X_4(\mathbb{C}P^1) \to \mathbb{C} \setminus \{0, 1\}.$$

In other words $\mathrm{cr}(x_1, x_2, x_3, x_4) = A(x_4)$ for the unique map $A \in \mathrm{PSL}(2, \mathbb{C})$ with $A(x_1) = \infty$, $A(x_2) = 0$, $A(x_3) = 1$.

[2] Recall that the holomorphic automorphism group of $\mathbb{C}P^1$ is isomorphic to $\mathrm{PSL}(2, \mathbb{C})$ acting by fractional-linear transformations.

In homogeneous coordinates, for $(x_1, x_2, x_3, x_4) \in X_4(\mathbb{C}P^1)$ where $x_j = [z_j, w_j]$ for $1 \le j \le 4$, the map cr is given by

$$\mathrm{cr}(x_1, x_2, x_3, x_4) = \frac{(z_1 w_3 - z_3 w_1)(z_2 w_4 - z_4 w_2)}{(z_2 w_3 - z_3 w_2)(z_1 w_4 - z_4 w_1)}.$$

In the affine chart $u_0 \times u_0 \times u_0 \times u_0$ it looks like

$$\mathrm{cr}(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_2 - z_3)(z_1 - z_4)}.$$

The cross-ratio map allows us to write down a diffeomorphism (in fact isomorphism of quasiprojective varieties)

$$\begin{gathered} c\colon X_n(\mathbb{C}P^1) \xrightarrow{\sim} X_3(\mathbb{C}P^1) \times X_{n-3}(\mathbb{C} \setminus \{0,1\}), \\ (\vec{x}, y_1, \dots, y_{n-3}) \mapsto (\vec{x}, \mathrm{cr}(\vec{x}, y_1), \dots, \mathrm{cr}(\vec{x}, y_{n-3})), \end{gathered}$$

where $\vec{x} = (x_1, x_2, x_3)$ denotes a point in $X_3(\mathbb{C}P^1)$ and $\mathrm{cr}(\vec{x}, y) = \mathrm{cr}(x_1, x_2, x_3, y)$ is the cross-ratio map. Later we shall see that this diffeomorphism is precisely what makes the proofs work, as it allows one to use the affine structure on $\mathbb{C} \supset \mathbb{C} \setminus \{0,1\}$.

Note that

$$\pi_c := \mathrm{pr}_2 \circ c\colon X_n(\mathbb{C}P^1) \to X_{n-3}(\mathbb{C} \setminus \{0,1\}),$$

where $\mathrm{pr}_2\colon X_3(\mathbb{C}P^1) \times X_{n-3}(\mathbb{C} \setminus \{0,1\}) \to X_{n-3}(\mathbb{C} \setminus \{0,1\})$ is the projection to the second factor, is simply a coordinate description of the natural projection

$$X_n(\mathbb{C}P^1) \to \mathcal{M}_{0,n} \cong X_n(\mathbb{C}P^1)/\mathrm{PSL}(2,\mathbb{C}).$$

Finally, note that c induces an isomorphism $c_\#\colon P_n(\mathbb{C}P^1) \to \mathbb{Z}/2\mathbb{Z} \times P_{n-3}(\mathbb{C} \setminus \{0,1\})$ on fundamental groups (recall that $P_3(\mathbb{C}P^1) \cong \pi_1(\mathrm{PSL}(2,\mathbb{C})) \cong \mathbb{Z}/2\mathbb{Z}$).

1.2.6 Differential 1–forms on configuration spaces Using the isomorphism c, we introduce special differential 1–forms on $X_n(\mathbb{C}P^1)$ for $n \ge 4$, that we consequently use as an intermediate step in our results. Denote by $u_1, \dots, u_{n-3}$ the affine coordinates on $\mathbb{C}^{n-3} \supset (\mathbb{C} \setminus \{0,1\})^{n-3} \supset X_{n-3}(\mathbb{C} \setminus \{0,1\})$. For each element $\nu \in I$ of an index set

$$I = \{(i;0)\}_{1 \le i \le n-3} \cup \{(i;1)\}_{1 \le i \le n-3} \cup \{(ij)\}_{1 \le i \ne j \le n-3},$$

define an $\mathbb{R}$–valued differential 1–form on $X_{n-3}(\mathbb{C} \setminus \{0,1\})$ by

$$\theta_\nu = \frac{1}{2\pi} \mathrm{Im}(\alpha_\nu),$$

with

$$\alpha_{i;0} = \frac{du_i}{u_i}, \quad \alpha_{i;1} = \frac{d(u_i - 1)}{u_i - 1}, \quad \alpha_{ij} = \frac{d(u_i - u_j)}{u_i - u_j}.$$

Finally, define
$$\widetilde{\theta}_\nu = (\pi_c)^*\theta_\nu \in \Omega^1(X_n(\mathbb{C}P^1), \mathbb{R}),$$
for each $\nu \in I$.

For a 1–form θ on a manifold Y and a smooth parametrized path $\gamma\colon [0,1] \to Y$, set
$$\int_\gamma |\theta| := \int_0^1 |\theta_{\gamma(t)}(\dot{\gamma}(t))|\, dt.$$
Clearly, for a smooth loop γ we have $|\int_\gamma \theta| \leq \int_\gamma |\theta|$. Moreover, $\int_\gamma |\theta| = \int_{\gamma^{-1}} |\theta|$, where γ^{-1} is the time-reversal of γ.

1.3 Main results

Our main technical result is:

Theorem 1.3 *For an isotopy $\overline{\phi} = \{\phi_t\}$ in $\mathcal{G}$, the average word norm of a trajectory $\lambda(x, \overline{\phi})$ is controlled by the L^1–length of $\overline{\phi}$:*
$$W(\overline{\phi}) = \int_{X_n(\mathbb{C}P^1)\setminus Z \cap (\phi_1)^{-1}(Z)} |[\lambda(x, \overline{\phi})]|_{P_n(S^2)}\, d\mu^{\otimes n}(x) \leq A \cdot l_1(\overline{\phi}) + B$$
for certain constants $A, B > 0$.

Remark 1.4 Note that $W(\overline{\phi})$ depends only on the class $\widetilde{\phi} = [\overline{\phi}] \in \widetilde{\mathcal{G}}$ of $\overline{\phi}$ in the universal cover $\widetilde{\mathcal{G}}$ of $\mathcal{G}$.

Theorem 1.3 has a number of consequences concerning the large-scale geometry of the L^1–metric on $\mathcal{G}$. Firstly, as any quasimorphism on a finitely generated group is controlled by the word norm, we immediately obtain the following statement.

Corollary 1.5 *The homogenization $\overline{\Phi}$ of each Gambaudo–Ghys quasimorphism Φ satisfies*
$$|\overline{\Phi}(\phi)| \leq C \cdot d_1(\phi, 1).$$

By a theorem of Ishida [26], the composition $Q(B_n(S^2)) \to Q(P_n(S^2)) \xrightarrow{\mathrm{GG}} Q(\mathcal{G})$, where the first arrow is the natural restriction map and the second is the Gambaudo–Ghys map, is an embedding. Hence for $n \geq 4$, by results of Bestvina and Fujiwara [6], $Q(\mathcal{G})$ is an infinite-dimensional vector space. Thus by Corollary 1.5 the diameter of $\mathcal{G}$ with the L^1–distance is infinite.

Corollary 1.6 *The L^1–diameter of $\mathcal{G}$ is infinite.*

Considering certain special examples of Gambaudo–Ghys quasimorphisms, and their calculations for certain autonomous flows, as in [12], we find for each integer $k \geq 1$ a k–

tuple of homogeneous Gambaudo–Ghys quasimorphisms $\{\overline{\Phi}_i\}_{1\leq i\leq k}$ and a k–tuple of commuting autonomous Hamiltonian flows (one-parameter subgroups) $\{\{\phi_i^t\}_{t\in\mathbb{R}}\}_{1\leq i\leq k}$ such that $\overline{\Phi}_i(\phi_j^t)=t\delta_{ij}$. This implies the following stronger statement.

Corollary 1.7 *The metric group $(\mathcal{G},d_1)$ admits a bi-Lipschitz group monomorphism from $(\mathbb{R}^k,d)$, where d is any metric on $\mathbb{R}^k$ induced by a vector-space norm.*

Moreover, by an argument of Kim and Koberda [30] (cf Benaim and Gambaudo [5] and Crisp and Wiest [16]), Theorem 1.3 implies the following statement, finishing an answer to a question of Kapovich [27] in the case of S^2.

Corollary 1.8 *The metric group $(\mathcal{G},d_1)$ admits a quasi-isometric group embedding from each right-angled Artin group endowed with the word metric.*

Remark 1.9 We note that Corollary 1.8 implies Corollary 1.6, providing the latter with a proof that does not use quasimorphisms.

Finally, Proposition 2.7 in [13] combined with Corollary 1.5 implies the following.

Corollary 1.10 *For each positive integer k, the complement in $\mathcal{G}$ of the set Aut^k of products of at most k autonomous diffeomorphisms contains a ball of any arbitrarily large radius in the L^1–metric.*

Remark 1.11 Let $p\geq 1$. By Jensen's (or Hölder's) inequality we have

$$d_1\leq d_p,$$

hence all the results above for d_1 continue to hold for d_p.

1.4 Outline of the proof

Theorem 1.3 is an immediate consequence of the following lemma and two propositions. The lemma states that for our purposes two different choices of short paths are equivalent.

Lemma 1.12 *Choosing as short paths the component-wise affine segments $\gamma'(x)$ in the chart $u_0\times\cdots\times u_0\colon\mathbb{C}^n\to\mathbb{C}P^n$ to the basepoint, obtain from the isotopy $\overline{\phi}=\{\phi_t\}$ another family of loops $\lambda'(x,\overline{\phi})$ for $x\in X_n(\mathbb{C}P^1)\setminus Z'\cap(\phi_1)^{-1}(Z')$, for a different negligible subset Z', and hence another average word norm function*

$$W'(\overline{\phi})=\int_{X_n(\mathbb{C}P^1)\setminus Z'\cap(\phi_1)^{-1}(Z')}|[\lambda'(x,\overline{\phi})]|_{P_n(S^2)}\,d\mu^{\otimes n}(x).$$

Then $|W(\overline{\phi})-W'(\overline{\phi})|\leq C$, for a constant C depending only on the systems of paths.

The first proposition is a purely topological fact about the word norm of the classes of loops in the fundamental group of the configuration space.

Proposition 1.13 *Let λ be a piecewise C^1 loop in $X_n(S^2)$ based at q. Let S be a finite generating set of $P_n(S^2)$. The word norm of the class $[\lambda] \in \pi_1(X_n(S^2), q) \cong P_n(S^2)$ with respect to S satisfies*

$$\|[\lambda]\|_S \leq A_0 \cdot \sum_{\nu \in I} \int_\lambda |\widetilde{\theta}_\nu| + B_0$$

for constants $A_0, B_0 > 0$ depending only on S and on n.

The second lemma is purely analytical and relies on the fact that we work with area-preserving diffeomorphisms, as well as on the fact that the differential forms we consider have integrable singularities near the divisors of $(\mathbb{C}P^1)^n$ that we excise to obtain $X_n(\mathbb{C}P^1)$.

Proposition 1.14 *There exist constants $A_1, B_1 > 0$, depending only on n, such that for each $\nu \in I$,*

$$\int_{X_n(\mathbb{C}P^1)\setminus Z} \left(\int_{\lambda(x,\bar{\phi})} |\widetilde{\theta}_\nu| \right) d\mu^{\otimes n}(x) \;\leq\; A_1 \cdot l_1(\bar{\phi}) + B_1.$$

Acknowledgements Part of this work was carried out during Brandenbursky's stay at IHES and CRM Montreal. He wishes to express his gratitude to both institutes. Brandenbursky was supported by a CRM-ISM fellowship and NSF grant no. 1002477.

This work was carried out during Shelukhin's stay in CRM Montreal, ICJ Lyon 1, Institut Mittag-Leffler and IAS. He thanks these institutions for their warm hospitality. He was supported by a CRM-ISM fellowship, FP7-IDEAS-ERC grant no. 258204 RealUman, a Mittag-Leffler fellowship and NSF grant no. DMS-1128155.

We thank Leonid Polterovich and Pierre Py for useful comments on the manuscript.

We thank Mikhail Belolipetsky, Octav Cornea, Steven Hurder, Jake Solomon and Leonid Polterovich for the opportunity to speak on a preliminary version of these results in seminars at IMPA, CRM, UIC, Hebrew University and the University of Chicago.

Finally, we thank the referees for their careful reading and useful comments on the exposition.

2 Proofs

Proof of Proposition 1.13 Let $\bar{\lambda}$ be a loop in $X_{n-3}(\mathbb{C}\setminus\{0,1\})$ based at $\bar{q}=\pi_c(q)$. We first claim that the word norm of $[\bar{\lambda}]$ in $P_{n-3}(\mathbb{C}\setminus\{0,1\})\cong\pi_1(X_{n-3}(\mathbb{C}\setminus\{0,1\}))$ satisfies

$$\|[\bar{\lambda}]\|_{P_{n-3}(\mathbb{C}\setminus\{0,1\})}\le A_2\cdot\sum_{\nu\in I}\int_{\bar{\lambda}}|\theta_\nu|+B_2 \tag{1}$$

for $A_2, B_2>0$. Proposition 1.13 follows immediately from this statement by setting $\bar{\lambda}=\pi_c\circ\lambda$, since the map $P_n(\mathbb{C}P^1)\to P_{n-3}(\mathbb{C}\setminus\{0,1\})$ induced by π_c is a quasi-isometry (note that it is identified with the projection $\mathbb{Z}/2\mathbb{Z}\times P_{n-3}(\mathbb{C}\setminus\{0,1\})\to P_{n-3}(\mathbb{C}\setminus\{0,1\})$ to the second factor, under the isomorphism given by c).

We require the following two lemmas from geometric group theory.

Lemma 2.1 *The natural map $e\colon P_{n-3}(\mathbb{C}\setminus\{0,1\})\to P_{n-1}(\mathbb{C})$ induced by adding constant strands at the punctures $\{0,1\}$ is a quasi-isometric embedding of groups.*

Lemma 2.2 *The inclusion $P_{n-1}(\mathbb{C})\to B_{n-1}(\mathbb{C})$ is a quasi-isometric embedding of groups.*

Lemma 2.2 is a consequence of a general fact about cocompact group actions [25, Corollary 24], as P_{n-1} is a subgroup of finite index in B_{n-1}. Lemma 2.1 is rather special to our case, and hence we provide a proof.

Proof of Lemma 2.1 The map $e\colon P_{n-3}(\mathbb{C}\setminus\{0,1\})\to P_{n-1}(\mathbb{C})$ fits into the following exact sequence [28]:

$$1\to P_{n-3}(\mathbb{C}\setminus\{0,1\})\to P_{n-1}(\mathbb{C})\to P_2(\mathbb{C})\to 1.$$

Note that $P_2(\mathbb{C})\cong\mathbb{Z}$. Moreover the generator z of the center $Z(P_{n-1}(\mathbb{C}))$ of $P_{n-1}(\mathbb{C})$ maps to a generator $1\in\mathbb{Z}$ of $P_2(\mathbb{C})$. Hence mapping 1 to z determines a section for $P_{n-1}(\mathbb{C})\to P_2(\mathbb{C})$ that yields an isomorphism between the above exact sequence and

$$1\to P_{n-3}(\mathbb{C}\setminus\{0,1\})\to P_{n-3}(\mathbb{C}\setminus\{0,1\})\times\mathbb{Z}\to\mathbb{Z}\to 1,$$

the first map taking the form $x\mapsto(x,0)$, and the second map being the projection to the second coordinate. The statement follows. □

Consider the geometric braid $\bar{\lambda}$ and add two constant strands at 0 and 1. Call the new geometric braid $\bar{\lambda}'$. It is now a loop in $X_{n-1}(\mathbb{C})$ based at $\bar{q}\cup\{0,1\}$. We show that for any geometric braid β in $X_{n-1}(\mathbb{C})$ based at $\overline{q} \cup \{0,1\}$, its word norm in $B_{n-1}(\mathbb{C})$ satisfies

$$\|[\beta]\|_{B_{n-1}(\mathbb{C})} \leq A_3 \cdot \sum_{1 \leq i<j \leq n-1} \int_\beta |\theta'_{ij}| + B_3 \tag{2}$$

for some $A_3, B_3 > 0$ and

$$\theta'_{ij} = \frac{1}{2\pi} \operatorname{Im}\left(\frac{d(u_i - u_j)}{u_i - u_j} \right)$$

for $1 \leq i \neq j \leq n-1$. Note that the forms $\{\theta'_{\nu'}\}_{\nu' \in I'}$ with

$$I' = \{(ij) \mid 1 \leq i < j \leq n,\ (i,j) \neq (n-2, n-1)\}$$

pull back to $\{\theta_\nu\}_{\nu \in I}$ under the natural embedding $X_{n-3}(\mathbb{C} \setminus \{0,1\}) \to X_{n-1}(\mathbb{C})$ given by $(u_1, \ldots, u_{n-3}) \mapsto (u_1, \ldots, u_{n-3}, 0, 1)$, and the form $\theta_{n-2,n-1}$ pulls back to the zero form. Hence by Lemmas 2.1 and 2.2 estimate (2) implies estimate (1).

For $1 \leq i \neq j \leq n-1$, let

$$p_{ij} \colon X_{n-1}(\mathbb{C}) \to X_2(\mathbb{C}), \quad (u_1, \ldots, u_{n-1}) \mapsto (u_i, u_j)$$

be the natural projection on the respective pair of coordinates. Note that $\theta'_{ij} = p^*_{ij}\theta$, with $\theta = \frac{1}{2\pi}\operatorname{Im}(d(u-v)/(u-v))$. Hence we have $\int_\beta |\theta'_{ij}| = \int_{\beta_{ij}} |\theta|$, where $\beta_{ij} = p_{ij} \circ \beta$, and moreover the following equality holds by the co-area formula; see [32, Theorem 5.1.12] or [24].

For almost all $\omega \in S^1$, the quantity

$$n_{ij}(\omega) = \#\left\{ t \in [0,1) \;\middle|\; \frac{p_i \circ \beta(t) - p_j \circ \beta(t)}{|p_i \circ \beta(t) - p_j \circ \beta(t)|} = \omega \in S^1 \right\}$$

is finite, and defines an L^1–function with norm

$$\int_{S^1} n_{ij}(\omega)\, dm(\omega) = \int_\beta |\theta'_{ij}|$$

for m the Haar (Lebesgue) measure on S^1. Note that (see [9; 10]) $n_{ij}(\omega)$ is the number of times that the i^{th} strand crosses over the j^{th} strand in the diagram of the braid β obtained by projection in the direction ω.

We claim that there exist a constant C (which depends only on n) and $\omega \in S^1$ such that for all $1 \leq i, j \leq n-1$,

$$n_{ij}(\omega) \leq C \int_\beta |\theta'_{ij}|$$

and all crossings are transverse. Indeed, if n_{ij} does not have vanishing L^1–norm, by Markov's (or Chebyshev's) inequality (see [31, Section 29.2, Theorem 5]) we estimate

$$m\left(\left\{\omega \in S^1 \;\middle|\; n_{ij}(\omega) \geq C \int_\beta |\theta'_{ij}|\right\}\right) \leq \frac{1}{C}.$$

Hence any $C > (n-1)(n-2)$ would be sufficient to ensure that the intersection

$$\bigcap_{1 \leq i \neq j \leq n-1} \left\{\omega \in S^1 \;\middle|\; n_{ij}(\omega) \leq C \int_\beta |\theta'_{ij}|\right\}$$

has positive measure and hence is nonempty. Moreover, clearly the set of all ω for which all crossings are transverse has full measure.

Hence, from the ω–projection diagram of the braid β we get a presentation of β as a word in the full braid group $B_n(\mathbb{C})$, generated by say the half-twists, that has exactly one generator for each overcrossing. Hence

$$\|[\beta]\|_{B_{n-1}(\mathbb{C})} \leq \sum_{i \neq j} n_{ij}(\omega) \leq 2C \sum_{i<j} \int_\beta |\theta_{ij}|.$$

This finishes the proof. □

Proof of Lemma 1.12 For a subset W of $X_n(\mathbb{C}P^1)$, set $W_{\phi_1} := W \cap (\phi_1)^{-1}(W)$. We note that for any negligible subset Z'' of $X_n(\mathbb{C}P^1)$,

$$W(\bar{\phi}) = \int_{X_n(\mathbb{C}P^1) \setminus (Z_{\phi_1} \cup Z'_{\phi_1} \cup Z''_{\phi_1})} \|[\lambda(x,\bar{\phi})]\|_{P_n(\mathbb{C}P^1)} \, d\mu^{\otimes n}(x),$$

$$W'(\bar{\phi}) = \int_{X_n(\mathbb{C}P^1) \setminus (Z_{\phi_1} \cup Z'_{\phi_1} \cup Z''_{\phi_1})} \|[\lambda'(x,\bar{\phi})]\|_{P_n(\mathbb{C}P^1)} \, d\mu^{\otimes n}(x),$$

whether these integrals are finite or not (simply by the definition of the Lebesgue integral).

Hence it is sufficient to show that there exists a constant C, depending only on the systems of paths and a negligible subset Z'' of $X_n(\mathbb{C}P^1)$, such that for each x in

$$X_n(\mathbb{C}P^1) \setminus (Z_{\phi_1} \cup Z'_{\phi_1} \cup Z''_{\phi_1}) = \mathbb{C}^n \setminus ((\mathbb{C}^n \cap Z_{\phi_1}) \cup (\mathbb{C}^n \cap Z'_{\phi_1}) \cup (\mathbb{C}^n \cap Z''_{\phi_1}))$$

we have $\left| \|[\lambda'(x,\bar{\phi})]\|_{P_n(\mathbb{C}P^1)} - \|[\lambda(x,\bar{\phi})]\|_{P_n(\mathbb{C}P^1)} \right| \leq C$.

And indeed we see that

$$[\lambda(x,\bar{\phi})] = [\delta(\phi_1 \cdot x)]^{-1} [\lambda'(x,\bar{\phi})][\delta(x)]$$

for $\delta(x)=\gamma(x)\#\overline{\gamma}'(x)$ and $[\delta(x)]_{P_n(\mathbb{C}P^1)}\leq C$, as can be seen by direct calculation on braid diagrams in $\mathbb{C}$. Indeed, as spherical geodesics map to circular arcs or affine rays under stereographic projection, and the latter happens for x in a negligible subset Z'' of $\mathbb{C}^n\setminus((\mathbb{C}^n\cap Z)\cup(\mathbb{C}^n\cap Z'))$, considering for $x\in X_n(\mathbb{C}P^1)\setminus(Z\cup Z'\cup Z'')$ the diagram of the geometric braid $\delta(x)$ in a generic direction $\omega\in S^1$, we see that it has at most $4\binom{n}{2}+\binom{n}{2}$ crossings, corresponding to the $\gamma(x)$ and $\overline{\gamma}'(x)$ parts of the geometric braid. Therefore $[\delta(x)]_{B_n(\mathbb{C})}\leq 5\binom{n}{2}$. However, $[\delta(x)]_{P_n(\mathbb{C})}\leq A\cdot[\delta(x)]_{B_n(\mathbb{C})}+B$ for constants $A,B>0$ (see Lemma 2.2 below), and obviously $[\delta(x)]_{P_n(\mathbb{C}P^1)}\leq[\delta(x)]_{P_n(\mathbb{C})}$. This finishes the proof. □

Proof of Proposition 1.14 We proceed to prove the analytic estimate on averages. First we show that for each $\nu\in I$, the integral of $|\widetilde{\theta}_\nu|$ on each of the short paths is universally bounded.

Lemma 2.3 *For each $\nu\in I$ and $x\in X_n(\mathbb{C}P^1)\setminus Z'$, we have $\int_{\gamma'(x)}|\widetilde{\theta}_\nu|\leq C$.*

Proof of Lemma 2.3 Recall that by definition of $\gamma'(x)$, we work in the chart $\mathbb{C}^n$. Since $\gamma'(x)$ is a component-wise affine segment, any linear function $h=z_i-z_j$ composed with $\gamma'(x)$ is an affine segment in $\mathbb{C}\setminus\{0\}$. Therefore $\int_{\gamma'(x)}|dh/h|\leq\pi$. By the definition of $\widetilde{\theta}_\nu$ (see (3) below) we obtain $\int_{\gamma'(x)}|\widetilde{\theta}_\nu|\leq\frac{1}{2\pi}\cdot 6\cdot\pi=3$ for all $\nu\in I$. □

By Lemma 2.3 it is sufficient to give a bound on

$$\int_{X_n(\mathbb{C}P^1)\setminus Z'}\left(\int_{\bar{\phi}\cdot x}|\widetilde{\theta}_\nu|\right)d\mu^{\otimes n}(x),$$

which by preservation of area and continuity can be rewritten as

$$\int_0^1\left(\int_{X_n(\mathbb{C}P^1)}|\widetilde{\theta}_\nu(X_t^{\oplus n})|(x)\,d\mu^{\otimes n}(x)\right)dt.$$

As the integrands are nonnegative, Tonelli's theorem ensures that one can change the order of integration without knowing in advance that the integrals converge.

We note that the above calculation is the only place in the proof that uses area-preservation.

Now note that under the standard stereographic projection, the lower hemisphere in S^2 is identified with the standard unit disk $\mathbb{D}=\{|z|\leq 1\}$ in $\mathbb{C}$. This embeds as

$$H_0=\{[z,1]\mid |z|\leq 1\}$$

in $\mathbb{C}P^1$ under the standard affine chart $u_0\colon \mathbb{C} \xrightarrow{\sim} U_0$ containing the point $[0,1]$. Similarly, the upper hemisphere is identified with the subset

$$H_\infty = \{[1,w] \mid |w| \le 1\}$$

of the image of the affine chart $u_\infty\colon \mathbb{C} \xrightarrow{\sim} U_\infty$ in $\mathbb{C}P^1$. Moreover, $\mathbb{C}P^1$ is the measure-disjoint union of H_0 and H_∞.

Let us write $(\mathbb{C}P^1)^n = \bigcup_{\epsilon\in\{0,\infty\}^n} H_\epsilon$ as a measure-disjoint union of products $H_\epsilon = H_{\epsilon_1} \times \cdots \times H_{\epsilon_n}$ of hemispheres. Let e_ϵ denote the isomorphism

$$e_\epsilon = e_{\epsilon_1} \times \cdots \times e_{\epsilon_n}\colon \mathbb{D}^n = \mathbb{D} \times \cdots \times \mathbb{D} \to H_\epsilon,$$

where $e_{\epsilon_j} = u_{\epsilon_j}|_{\mathbb{D}}\colon \mathbb{D} \to H_{\epsilon_j}$ for $1 \le j \le n$ are given by the embeddings above. Write $H'_\epsilon = H_\epsilon \cap X_n(\mathbb{C}P^1)$ and $\Delta_\epsilon = e_\epsilon^{-1}(H_\epsilon \setminus H'_\epsilon)$. Put $\overline{e}_\epsilon = e_\epsilon|_{H'_\epsilon}\colon \mathbb{D}^n \setminus \Delta_\epsilon \to H'_\epsilon$. Then

$$\int_{X_n(\mathbb{C}P^1)} |\widetilde{\theta}_\nu(X_t^{\oplus n})|(x)\, d\mu^{\otimes n}(x) = \sum_{\epsilon\in\{0,\infty\}^n} \int_{H'_\epsilon} |\widetilde{\theta}_\nu(X_t^{\oplus n})|(x)\, d\mu^{\otimes n}(x).$$

Write

$$\int_{H'_\epsilon} |\widetilde{\theta}_\nu(X_t^{\oplus n})|(x)\, d\mu^{\otimes n}(x) = \int_{\mathbb{D}^n\setminus\Delta_\epsilon} |(\overline{e}_\epsilon)^*\widetilde{\theta}_\nu(e_\epsilon^* X_t^{\oplus n})|(x)\, de_\epsilon^*(\mu^{\otimes n}).$$

Note that $e_\epsilon^* X_t^{\oplus n} = e_{\epsilon_1}^* X_t \oplus \cdots \oplus e_{\epsilon_n}^* X_t$, and that for each $\nu \in I$ we have $(\overline{e}_\epsilon)^*\widetilde{\theta}_\nu = \frac{1}{2\pi}\operatorname{Im}(df_{\epsilon,\nu}/f_{\epsilon,\nu})$ with $f_{\epsilon,\nu} = l_1 l_2/(l_3 l_4)$ or $f_{\epsilon,\nu} = l_1 l_2 l_3/(l_4 l_5 l_6)$, with each l_k of the form $a-b$ or $ab-1$, where a and b are natural coordinates on two of the factors in the product $\mathbb{D}^n$. From now on, we focus on the second case, since the first case is simpler and is treated analogously. We state it more precisely: if $(a_1,\dots,a_n)$ are natural coordinates on $\mathbb{D}^n$, then $\Delta_\epsilon = \bigcup_{1\le i<j\le n}\{h_{ij} = 0\}$, with $h_{ij} = a_i - a_j$ if $\epsilon_i = \epsilon_j$ and $h_{ij} = a_i a_j - 1$ if $\epsilon_i \ne \epsilon_j$, and for each $1 \le k \le 6$ we have $l_k = h_{ij}$ for some $1 \le i < j \le n$. Indeed, this follows immediately from the identities

$$\operatorname{cr}(x_1,x_2,x_3,x_4) - 1 = -\operatorname{cr}(x_1,x_3,x_2,x_4) = \frac{(z_1w_2 - z_2w_1)(z_3w_4 - z_4w_3)}{(z_2w_3 - z_3w_2)(z_1w_4 - z_4w_1)}$$

and

$$\operatorname{cr}(x_1,x_2,x_3,x_4) - \operatorname{cr}(x_1,x_2,x_3,x_5) = \frac{(z_1w_3 - z_3w_1)(z_1w_2 - z_2w_1)(z_5w_4 - z_4w_5)}{(z_2w_3 - z_3w_2)(z_1w_4 - z_4w_1)(z_1w_5 - z_5w_1)}$$

for $x_j = [z_j, w_j]$ in homogeneous coordinates on $\mathbb{C}P^1$ for $1 \le j \le 5$.

We record the formula

$$(\overline{e}_\epsilon)^*\widetilde{\theta}_\nu = \frac{1}{2\pi}\operatorname{Im}\left(\frac{dl_1}{l_1} + \frac{dl_2}{l_2} + \frac{dl_3}{l_3} - \frac{dl_4}{l_4} - \frac{dl_5}{l_5} - \frac{dl_6}{l_6}\right), \tag{3}$$

which follows immediately from the above discussion. From (3) it follows that it is sufficient to estimate

$$\int_{\mathbb{D}\times\mathbb{D}\setminus\{h_{ij}=0\}} \left|\frac{dh_{ij}}{h_{ij}}\right| (e^*_{\epsilon_i} X_t \oplus e^*_{\epsilon_j} X_t)(e_{\epsilon_i} \times e_{\epsilon_j})^* \, d\mu^{\otimes 2}$$

for each $1 \le i < j \le n$.

The pullback $|\cdot|_{\mathrm{Sph}} = e^*_\epsilon |\cdot|_{S^2}$ for $\epsilon \in \{0, \infty\}$ of the metric on the sphere in either of the coordinate charts is equal to

$$(|\cdot|_{\mathrm{Sph}})_\zeta = (1+|\zeta|^2)^{-1}\,|\cdot|_{\mathrm{Eucl}}.$$

Abbreviating $|\cdot|_{\mathrm{Eucl}} = |\cdot|$ we therefore have, for all $\zeta \in \mathbb{D}$,

$$\tfrac{1}{2}|\cdot| \le |\cdot|_{\mathrm{Sph}} \le |\cdot|.$$

Hence in order to obtain an estimate via $\int_{\mathbb{D}} |(e^*_{\epsilon_k} X_t)|_{\mathrm{Sph}} \, d\mu$ for $1 \le k \le n$, it is sufficient to estimate via $\int_{\mathbb{D}} |(e^*_{\epsilon_k} X_t)| \, d\mu$.

Hence it is sufficient to estimate

$$\int_{\mathbb{D}^2\setminus\{a-b=0\}} \left|\frac{d(a-b)}{a-b}\right| (A_t \oplus B_t)\, d\mu^{\otimes 2}$$

or

$$\int_{\mathbb{D}^2\setminus\{ab-1=0\}} \left|\frac{d(ab-1)}{ab-1}\right| (A_t \oplus B_t)\, d\mu^{\otimes 2}.$$

Here

$$d\mu(\zeta) = 2(1+|\zeta|^2)^{-2}\, dm(\zeta)$$

is the pullback of the spherical measure to $\mathbb{D}$ by any of the maps e_{ϵ_j} (note that the map $\mathbb{C}P^1 \to \mathbb{C}P^1$ given by $[z,w] \mapsto [w,z]$ is an isometry of the spherical metric, and hence preserves the volume form), and A_t, B_t are $e^*_{\epsilon_i} X_t, e^*_{\epsilon_j} X_t$ for appropriate i, j.

Start with

$$\int_{\mathbb{D}^2\setminus\{a-b=0\}} \frac{|A_t(a) - B_t(b)|}{|a-b|}\, d\mu(a)\, d\mu(b).$$

We apply the triangle inequality $|A_t(a) - B_t(b)| \le |A_t(a)| + |B_t(b)|$, and estimate the two resulting terms separately. Since they are estimated analogously, we show the estimate for the first term only. We have

$$\begin{aligned}\int_{\mathbb{D}}\int_{\mathbb{D}} \frac{|A_t(a)|}{|a-b|}\, d\mu(b)\, d\mu(a) &= \int_{\mathbb{D}} |A_t(a)| \left(\int_{\mathbb{D}} \frac{1}{|a-b|}\, d\mu(b)\right) d\mu(a)\\ &\le C \cdot \int_{\mathbb{D}} |A_t(a)|\, d\mu(a),\end{aligned}$$

since
$$\int_{\mathbb{D}} \frac{1}{|a-b|} d\mu(b) \le 2 \int_{\mathbb{D}} \frac{1}{|a-b|} dm(b) \le 8\pi = C.$$
We continue with
$$\int_{\mathbb{D}^2 \setminus \{ab-1=0\}} \frac{|aB_t(b) - A_t(a)b|}{|ab-1|} \, d\mu(a) \, d\mu(b).$$
Using the triangle inequality in the numerator, we estimate the two terms separately. Consider for example the first term. The estimate proceeds analogously to the previous case, the only difference being the following calculation. Writing $a^* = 1/a$, we compute
$$\int_{\mathbb{D}} \frac{|a|}{|ab-1|} \, d\mu(a) = \int_{\{|a^*| \ge 1\}} \frac{1}{|b-a^*|} \, d\mu(a^*) \le C.$$
Indeed, write the last integral as the sum of the integrals over the measure-disjoint subsets $\{1 \le |a^*| \le 2\}$ and $\{|a^*| \ge 2\}$ of $\mathbb{C}$. Then we estimate
$$\begin{aligned} \int_{\{|a^*| \le 2\}} \frac{1}{|b-a^*|} \frac{1}{(1+|a^*|^2)^2} \, dm(z) &\le \int_{\{|a^*| \le 2\}} \frac{1}{|b-a^*|} \, dm(a^*) \\ &\le \int_{\{|b-a^*| \le 3\}} \frac{1}{|b-a^*|} \, dm(b-a^*) \\ &= 6\pi, \end{aligned}$$
and, recalling that $|b| \le 1$,
$$\begin{aligned} \int_{\{|a^*| \ge 2\}} \frac{1}{|a^*-b|} \frac{1}{(1+|a^*|^2)^2} \, dm(a^*) &\le \int_{\{|a^*| \ge 2\}} \frac{1}{|a^*|-|b|} \frac{1}{(1+|a^*|^2)^2} \, dm(a^*) \\ &\le \int_{\{|a^*| \ge 2\}} \frac{2}{|a^*|} \frac{1}{(1+|a^*|^2)^2} \, dm(a^*) \\ &= C_1 < \infty. \end{aligned}$$
This gives us an estimate as required, with $C = 12\pi + 2C_1$. □

3 Examples of quasimorphisms and bi-Lipschitz embeddings of vector spaces

For $\alpha \in P_n = P_n(\mathbb{C})$ we denote by $\widehat{\alpha}$ the n–component link which is a closure of α; see Figure 1.

Let $\operatorname{sign}_n \colon P_n \to \mathbb{Z}$ be a map such that $\operatorname{sign}_n(\alpha) = \operatorname{sign}(\widehat{\alpha})$, where sign is the signature invariant of links in $\mathbb{R}^3$. Gambaudo and Ghys [23] showed that sign_n defines a

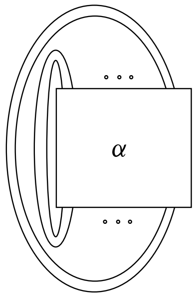


Figure 1: Closure $\widehat{\alpha}$ of a braid α

quasimorphism on P_n; see [8] for a different proof. We denote by $\overline{\mathrm{sign}}_n\colon P_n \to \mathbb{R}$ the induced homogeneous quasimorphism. Recall that the center of P_n is isomorphic to $\mathbb{Z}$. Let Δ_n be a generator of the center of P_n. It is a well-known fact that $P_n(S^2)$ is isomorphic to the quotient of P_{n-1} by the cyclic group $\langle \Delta_{n-1}^2 \rangle$; see [7]. Let $\mathrm{lk}_n\colon P_n \to \mathbb{Z}$ be a restriction to P_n of a canonical homomorphism from $B_n = B_n(\mathbb{C})$ to $\mathbb{Z}$ which takes value 1 on each Artin generator of B_n. Let $s_{n-1}\colon P_{n-1} \to \mathbb{R}$ be a homogeneous quasimorphism defined by

$$s_{n-1}(\alpha) := \overline{\mathrm{sign}}_{n-1}(\alpha) - \frac{\overline{\mathrm{sign}}_{n-1}(\Delta_{n-1})}{\mathrm{lk}_{n-1}(\Delta_{n-1})}\,\mathrm{lk}_{n-1}(\alpha).$$

Since $s_{n-1}(\Delta_{n-1}) = 0$, the homogeneous quasimorphism s_{n-1} descends to a homogeneous quasimorphism $\overline{s}_n\colon P_n(S^2) \to \mathbb{R}$. Note that $\overline{s}_2$ and $\overline{s}_3$ are trivial because $P_2(S^2)$ and $P_3(S^2)$ are finite groups.

For each $n \geq 4$, let

$$\overline{\mathrm{Sign}}_n\colon \mathrm{Diff}_0(S^2, \sigma) \to \mathbb{R}$$

be the induced homogeneous quasimorphism. In [23, Section 5.3] Gambaudo and Ghys evaluated quasimorphisms $\overline{\mathrm{Sign}}_{2n}$ on a family of diffeomorphisms

$$f_\omega\colon S^2 \to S^2$$

such that $f_\omega(\infty) = \infty$ and $f_\omega(x) = e^{2i\pi\omega(|x|)}x$; here S^2 is identified with $\mathbb{C} \cup \{\infty\}$ and $\omega\colon \mathbb{R}_+ \to \mathbb{R}$ is a function which is constant in a neighborhood of 0 and outside some compact set. Let $a(r)$ be the spherical area (with the normalization $\mathrm{vol}(\mathbb{C}) = 1$) of the disc in $\mathbb{C}$ with radius r centered at 0. Set $u = 1 - 2a(r)$ and let $\widetilde{\omega}(u) = \omega(r)$. In [23, Lemma 5.3] Gambaudo and Ghys showed that for each $n \geq 2$,

$$\overline{\mathrm{Sign}}_{2n}(f_\omega) = \frac{n}{2}\int_{-1}^{1}(u^{2n-1} - u)\widetilde{\omega}(u)\,du. \tag{4}$$

Proof of Corollary 1.7 Let $H_\omega\colon S^2 \to \mathbb{R}$ be a smooth function supported away from the $\{\infty\}$ point and $f_{t,\omega}$ be a Hamiltonian flow generated by H_ω, such that $f_{1,\omega} = f_\omega$. Since $f_{t,\omega}$ is an autonomous flow, by (4) we have

$$\overline{\mathrm{Sign}}_{2n}(f_{t,\omega}) = t\,\frac{n}{2}\int_{-1}^{1}(u^{2n-1}-u)\widetilde{\omega}(u)\,du.$$

Let $d \in \mathbb{N}$. It follows from (4) that it is straightforward to construct a family of functions $\omega_i\colon \mathbb{R}_+ \to \mathbb{R}$ and $\{H_{\omega_i}\}_{i=1}^d$ supported away from the $\{\infty\}$ point such that:

- Each Hamiltonian flow f_{t,ω_i} is generated by H_{ω_i} and $f_{1,\omega_i} = f_{\omega_i}$.
- The functions $\{H_{\omega_i}\}_{i=1}^d$ have disjoint support and hence the diffeomorphisms f_{t,ω_i} and f_{s,ω_j} commute for all $s,t \in \mathbb{R}$, $1 \le i, j \le n$.
- The $(d \times d)$ matrix
$$\begin{pmatrix} \overline{\mathrm{Sign}}_4(f_{1,\omega_1}) & \cdots & \overline{\mathrm{Sign}}_4(f_{1,\omega_d}) \\ \vdots & \ddots & \vdots \\ \overline{\mathrm{Sign}}_{2d+2}(f_{1,\omega_1}) & \cdots & \overline{\mathrm{Sign}}_{2d+2}(f_{1,\omega_d}) \end{pmatrix}$$
is nonsingular.

It follows that there exists a family $\{\overline{\Phi}_i\}_{i=1}^d$ of homogeneous quasimorphisms on $\mathrm{Diff}_0(S^2,\sigma)$ such that $\overline{\Phi}_i$ is a linear combination of $\overline{\mathrm{Sign}}_4, \dots, \overline{\mathrm{Sign}}_{2d+2}$ and

$$\overline{\Phi}_i(f_{t,\omega_j}) = \begin{cases} t & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{5}$$

Let $I\colon \mathbb{R}^d \to \mathrm{Diff}_0(S^2,\sigma)$ be a map such that

$$I(v) := f_{v_1,\omega_1} \circ \cdots \circ f_{v_d,\omega_d}$$

for $v = (v_1, \dots, v_d)$. It follows from the construction of $\{f_{v_i,\omega_i}\}_{i=1}^d$ that I is a monomorphism. Let $A'_p := \max_i\, l_p(\{f_{t,\omega_i}\}_{0\le t\le 1})$. Then

$$\|f_{v_1,\omega_1} \circ \cdots \circ f_{v_d,\omega_d}\|_p \le A'_p\|v\|,$$

where $\|v\| = \sum_{i=1}^d |v_i|$ and $\|\cdot\|_p = d_p(\cdot\,,1)$ denotes the L^p–norm.

The diffeomorphisms $f_{v_1,\omega_1}, \dots, f_{v_d,\omega_d}$ pairwise commute. Hence, for each $1 \le i \le d$, by Corollary 1.5 and Equation (5) we have

$$\|f_{v_1,\omega_1} \circ \cdots \circ f_{v_d,\omega_d}\|_p \ge A_p^{-1}|\overline{\Phi}_i(f_{v_1,\omega_1} \circ \cdots \circ f_{v_d,\omega_d})| = A_p^{-1}\cdot|v_i||\overline{\Phi}_i(f_{1,\omega_i})|,$$

where A_p is the maximum over the Lipschitz constants (in Corollary 1.5) of the functions

$$\overline{\Phi}_i\colon \mathrm{Diff}_0(S^2,\sigma) \to \mathbb{R}.$$

It follows that

$$\|f_{v_1,\omega_1}\circ\cdots\circ f_{v_d,\omega_d}\|_p \geq \big((d\cdot A_p)^{-1}\min_i|\overline{\Phi}_i(f_{1,\omega_i})|\big)\|v\| = (d\cdot A_p)^{-1}\|v\|,$$

and the proof is complete. □

Appendix: The case of the torus

The proof of [20, Theorem A.1] or [11, Theorem 1.2] applied to $\mathrm{Ham}(T^2, dx\wedge dy)$, combined with the fact that $\mathrm{Ham}(T^2, dx\wedge dy)$ is simply connected as a topological space (see [35, Chapter 7.2.B]) shows that the diameter of $(\mathrm{Ham}(T^2, dx\wedge dy), d_{L^1})$ is infinite. Hence by Remark 1.11, the following statement implies that the diameter of $(\mathrm{Diff}_0(T^2, dx\wedge dy), d_{L^p})$ is infinite for all $p\geq 1$.

Proposition A.1 *The inclusion* $(\mathrm{Ham}(T^2, dx\wedge dy), d_{L^1})\hookrightarrow(\mathrm{Diff}_0(T^2, dx\wedge dy), d_{L^1})$ *is a quasi-isometry.*

Proof of Proposition A.1 We equip the torus T^2 with the standard flat Riemannian metric. We use the following instance of the flux exact sequence (see [4; 33]):

$$1\to\mathrm{Ham}(T^2, dx\wedge dy)\xrightarrow{\iota}\mathrm{Diff}_0(T^2, dx\wedge dy)\xrightarrow{\mathrm{Flux}}T^2\to 1.$$

It has the property that the monomorphism $\tau\colon T^2\to\mathrm{Diff}_0(T^2, dx\wedge dy)$ given by $\tau(a,b)\colon (x,y)\mapsto(x+b, y-a)$ satisfies $\mathrm{Flux}\circ\tau = \mathbf{1}_{T^2}$. In particular,

$$\mathrm{Diff}_0(T^2, dx\wedge dy) = \mathrm{Ham}(T^2, dx\wedge dy)\cdot\tau(T^2).$$

However, $d_{L^1}(\tau(a,b),\mathbf{1})\leq 1/\sqrt{2}$ for all $(a,b)\in T^2$, as is verified in an elementary manner. In particular, $\iota\colon\mathrm{Ham}(T^2, dx\wedge dy)\to\mathrm{Diff}_0(T^2, dx\wedge dy)$ has coarsely dense image.

We proceed to prove that ι is a bi-Lipschitz group monomorphism. First, $\iota^* d_{L^1}\leq d_{L^1}$ is immediate by definition of the L^1–distance. We claim that $c\cdot d_{L^1}\leq\iota^* d_{L^1}$ for some $0<c<1$. By right-invariance, it is sufficient to show that $c\cdot d_{L^1}(h,1)\leq d_{L^1}(\iota(h),1)$ for all $h\in\mathrm{Ham}(T^2, dx\wedge dy)$. Consider a smooth path $[0,1]\to\mathrm{Diff}_0(T^2, dx\wedge dy)$ such that $t\mapsto g_t$ with $g_0=\mathbf{1}$ and $g_1=\iota(h)$. Look at the path

$$[0,1]\to\mathrm{Diff}_0(T^2, dx\wedge dy),\qquad t\mapsto\tau_t=\tau\circ\mathrm{Flux}(g_t).$$

Notice that in fact it is a loop based at $\mathbf{1}\in\mathrm{Diff}_0(T^2, dx\wedge dy)$. We shall prove the following estimate of L^1–lengths.

Claim 1 *We have* $l_1(\{\tau_t^{-1}\})\leq c_0\cdot l_1(\{g_t\})$ *for some* $c_0>0$.

We defer the proof of this claim to the end of the section. Define the path

$$[0,1] \to \mathrm{Ham}(T^2, dx \wedge dy), \quad t \mapsto h_t = \iota^{-1}(\tau_t^{-1} g_t),$$

with $h_0 = \mathbf{1}$ and $h_1 = h$. Then, since τ_t are isometries, we see that

$$l_1(\{h_t\}) \le l_1(\{\tau_t^{-1}\}) + l_1(\{g_t\}) \le (1 + c_0) \cdot l_1(\{g_t\}),$$

by Claim 1. This finishes the proof, with $c = (1 + c_0)^{-1}$. □

Proof of Claim 1 First of all, since τ_t are isometries, $l_1(\{\tau_t^{-1}\}) = l_1(\{\tau_t\})$. Let $Y_t = a_t(x,y)\,\partial_x + b_t(x,y)\,\partial_y$ be the time-dependent symplectic vector field generating $\{g_t\}$. For $f \in C^\infty(T^2, \mathbb{R})$, denote its average by $\langle f \rangle = \int_{T^2} f\, dx \wedge dy$ (our area form has total area 1). We record that

$$|\langle f \rangle| \le |f|_{L^1}. \tag{6}$$

It follows quickly from the definition of Flux (and an explicit characterization of exact 1–forms on T^2) that the vector field $Z_t = \langle a_t \rangle\,\partial_x + \langle b_t \rangle\,\partial_y$ generates τ_t. Hence by (6), for each $0 \le t \le 1$ we have

$$|Z_t|_{L^1} \le \sqrt{2} \cdot |Y_t|_{L^1}.$$

Hence $l_1(\{\tau_t\}) \le \sqrt{2} \cdot l_1(\{g_t\})$, finishing the proof with $c_0 = \sqrt{2}$. □

References

[1] **V Arnol'd**, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*, Ann. Inst. Fourier (Grenoble) 16 (1966) 319–361 MR

[2] **V I Arnol'd**, *The cohomology ring of the colored braid group*, Mat. Zametki 5 (1969) 227–231 MR In Russian; translated in Math. Notes 5 (1969) 138–140

[3] **V I Arnol'd**, **B A Khesin**, *Topological methods in hydrodynamics*, Applied Mathematical Sciences 125, Springer (1998) MR

[4] **A Banyaga**, *The structure of classical diffeomorphism groups*, Mathematics and its Applications 400, Kluwer, Dordrecht (1997) MR

[5] **M Benaim**, **J-M Gambaudo**, *Metric properties of the group of area preserving diffeomorphisms*, Trans. Amer. Math. Soc. 353 (2001) 4661–4672 MR

[6] **M Bestvina**, **K Fujiwara**, *Bounded cohomology of subgroups of mapping class groups*, Geom. Topol. 6 (2002) 69–89 MR

[7] **J S Birman**, *On braid groups*, Comm. Pure Appl. Math. 22 (1969) 41–72 MR

[8] **M Brandenbursky**, *On quasi-morphisms from knot and braid invariants*, J. Knot Theory Ramifications 20 (2011) 1397–1417 MR

[9] **M Brandenbursky**, *Quasi-morphisms and L^p–metrics on groups of volume-preserving diffeomorphisms*, J. Topol. Anal. 4 (2012) 255–270 MR

[10] **M Brandenbursky**, **J Kędra**, *On the autonomous metric on the group of area-preserving diffeomorphisms of the 2–disc*, Algebr. Geom. Topol. 13 (2013) 795–816 MR

[11] **M Brandenbursky**, **J Kędra**, *Quasi-isometric embeddings into diffeomorphism groups*, Groups Geom. Dyn. 7 (2013) 523–534 MR

[12] **M Brandenbursky**, **E Shelukhin**, *On the large-scale geometry of the L^p–metric on the symplectomorphism group of the two-sphere*, preprint, IHES (2015) Available at `http://preprints.ihes.fr/2015/M/M-15-09.pdf`

[13] **M Brandenbursky**, **E Shelukhin**, *On the L^p–geometry of autonomous Hamiltonian diffeomorphisms of surfaces*, Math. Res. Lett. 22 (2015) 1275–1294 MR

[14] **E Calabi**, *On the group of automorphisms of a symplectic manifold*, from "Problems in analysis" (R C Gunning, editor), Princeton Univ. Press (1970) 1–26 MR

[15] **D Calegari**, *scl*, MSJ Memoirs 20, Math. Soc. Japan, Tokyo (2009) MR

[16] **J Crisp**, **B Wiest**, *Quasi-isometrically embedded subgroups of braid and diffeomorphism groups*, Trans. Amer. Math. Soc. 359 (2007) 5485–5503 MR

[17] **D G Ebin**, *Geodesics on the symplectomorphism group*, Geom. Funct. Anal. 22 (2012) 202–212 MR

[18] **D G Ebin**, **J Marsden**, *Groups of diffeomorphisms and the motion of an incompressible fluid*, Ann. of Math. 92 (1970) 102–163 MR

[19] **D G Ebin**, **G Misiołek**, **S C Preston**, *Singularities of the exponential map on the volume-preserving diffeomorphism group*, Geom. Funct. Anal. 16 (2006) 850–868 MR

[20] **Y Eliashberg**, **T Ratiu**, *The diameter of the symplectomorphism group is infinite*, Invent. Math. 103 (1991) 327–340 MR

[21] **A Fathi**, *Transformations et homéomorphismes préservant la mesure: systémes dynamiques minimaux*, PhD thesis, Université Paris-Sud, Orsay (1980)

[22] **J-M Gambaudo**, **E Ghys**, *Enlacements asymptotiques*, Topology 36 (1997) 1355–1379 MR

[23] **J-M Gambaudo**, **E Ghys**, *Commutators and diffeomorphisms of surfaces*, Ergodic Theory Dynam. Systems 24 (2004) 1591–1617 MR

[24] **J-M Gambaudo**, **M Lagrange**, *Topological lower bounds on the distance between area preserving diffeomorphisms*, Bol. Soc. Brasil. Mat. 31 (2000) 9–27 MR

[25] **P de la Harpe**, *Topics in geometric group theory*, Univ. of Chicago Press (2000) MR

[26] **T Ishida**, *Quasi-morphisms on the group of area-preserving diffeomorphisms of the 2–disk via braid groups*, Proc. Amer. Math. Soc. Ser. B 1 (2014) 43–51 MR

[27] **M Kapovich**, *RAAGs in Ham*, Geom. Funct. Anal. 22 (2012) 733–755 MR

[28] **C Kassel**, **V Turaev**, *Braid groups*, Graduate Texts in Mathematics 247, Springer (2008) MR

[29] **B Khesin**, **R Wendt**, *The geometry of infinite-dimensional groups*, Ergeb. Math. Grenzgeb. 51, Springer (2009) MR

[30] **S-h Kim**, **T Koberda**, *Anti-trees and right-angled Artin subgroups of braid groups*, Geom. Topol. 19 (2015) 3289–3306 MR

[31] **A N Kolmogorov**, **S V Fomīn**, *Introductory real analysis*, Dover, New York (1975) MR corrected reprinting

[32] **S G Krantz**, **H R Parks**, *Geometric integration theory*, Birkhäuser, Boston (2008) MR

[33] **D McDuff**, **D Salamon**, *Introduction to symplectic topology*, 2nd edition, Oxford Univ. Press (1998) MR

[34] **L Polterovich**, *Hofer's diameter and Lagrangian intersections*, Int. Math. Res. Not. 1998 (1998) 217–223 MR

[35] **L Polterovich**, *The geometry of the group of symplectic diffeomorphisms*, Birkhäuser, Basel (2001) MR

[36] **L Polterovich**, *Floer homology, dynamics and groups*, from "Morse theoretic methods in nonlinear analysis and in symplectic topology" (P Biran, O Cornea, F Lalonde, editors), NATO Sci. Ser. II Math. Phys. Chem. 217, Springer, Dordrecht (2006) 417–438 MR

[37] **P Py**, *Quasi-morphismes de Calabi et graphe de Reeb sur le tore*, C. R. Math. Acad. Sci. Paris 343 (2006) 323–328 MR

[38] **P Py**, *Quasi-morphismes et invariant de Calabi*, Ann. Sci. École Norm. Sup. 39 (2006) 177–195 MR

[39] **E Shelukhin**, *"Enlacements asymptotiques" revisited*, Ann. Math. Qué. 39 (2015) 205–208 MR

[40] **A I Shnirel'man**, *The geometry of the group of diffeomorphisms and the dynamics of an ideal incompressible fluid*, Mat. Sb. 128(170) (1985) 82–109, 144 MR In Russian; translated in Math. USSR-Sb. 56 (1987) 79–105

[41] **A I Shnirel'man**, *Generalized fluid flows, their approximation and applications*, Geom. Funct. Anal. 4 (1994) 586–620 MR

[42] **A Shnirelman**, *Microglobal analysis of the Euler equations*, J. Math. Fluid Mech. 7 (2005) S387–S396 MR

[43] **S Smale**, *Diffeomorphisms of the 2–sphere*, Proc. Amer. Math. Soc. 10 (1959) 621–626 MR

Department of Mathematics, Ben-Gurion University
Beer-Sheva, Israel

Department of Mathematics and Statistics, University of Montréal
Montréal, Canada

brandens@math.bgu.ac.il, shelukhin@dms.umontreal.ca

Proposed: Yasha Eliashberg
Seconded: Danny Calegari, Leonid Polterovich

Received: 8 July 2016
Accepted: 28 January 2017

Guidelines for Authors

Submitting a paper to Geometry & Topology

Papers must be submitted using the upload page at the GT 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 Geometry & 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 GT is available on the GT 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.

Geometry & Topology

Volume 21 Issue 6 (pages 3191–3810) 2017