Density of isoperimetric spectra

NOEL BRADY
MAX FORESTER

We show that the set of k–dimensional isoperimetric exponents of finitely presented groups is dense in the interval $[1,\infty)$ for $k \geqslant 2$. Hence there is no higher-dimensional analogue of Gromov's gap $(1,2)$ in the isoperimetric spectrum.

20F65; 20F69, 20E06, 57M07, 53C99

We dedicate this paper to the memory of John Stallings (1935–2008).

1 Introduction

Dehn functions of groups have been the subject of intense activity over the past two decades. The Dehn function $\delta(x)$ of a group G is a quasi-isometry invariant which describes the best possible isoperimetric inequality that holds in any geometric model for the group. Specifically, for a given x, $\delta(x)$ is the smallest number A such that every null-homotopic loop of length at most x bounds a disk of area A or less. One defines length and area combinatorially, based on a presentation 2–complex for G, and the resulting Dehn function is well defined up to coarse Lipschitz equivalence. If G is the fundamental group of a closed Riemannian manifold M, then ordinary length and area in M may be used instead, and one obtains an equivalent function. (This seemingly modest but non-trivial result is sometimes called the Filling Theorem; see Bridson [6] or Burillo and Taback [9] for a proof.)

Due in large part to the work of Birget, Rips and Sapir [24] we now have a fairly complete understanding of which functions are Dehn functions of finitely presented groups. In the case of power functions, one defines the *isoperimetric spectrum* to be the following (countable) subset of the line:

$$\mathrm{IP} \;=\; \{\, \alpha \in [1,\infty) \mid f(x) = x^{\alpha} \text{ is equivalent to a Dehn function} \,\}.$$

We know from Brady and Bridson [4] that the isoperimetric spectrum has closure $\{1\} \cup [2,\infty)$ and, from Brady, Bridson, Forester and Shankar [5], that it contains all rational numbers in $[2,\infty)$. Moreover, in the range $(4,\infty)$, it contains (almost exactly)

those numbers having computational complexity below a certain threshold [24]. The gap $(1,2)$ reflects Gromov's theorem to the effect that every finitely presented group with sub-quadratic Dehn function is hyperbolic, and hence has linear Dehn function. Several proofs of this result are known: see Gromov [16], Ol′shanskiĭ [20], Papasoglu [21] and Bowditch [3].

By analogy with ordinary Dehn functions, one defines the k–dimensional Dehn function $\delta^{(k)}(x)$, describing the optimal k–dimensional isoperimetric inequality that holds in G. Given x, $\delta^{(k)}(x)$ is the smallest V such that every k–dimensional sphere of volume at most x bounds a $(k+1)$–dimensional ball of volume V or less. One uses combinatorial notions of volume, based on a chosen k–connected model for G. Again, up to coarse Lipschitz equivalence, $\delta^{(k)}(x)$ is preserved by quasi-isometries, by Alonso, Wang and Pride [2], and in particular does not depend on the choice of model for G.

Precise details regarding the definition of $\delta^{(k)}(x)$ are given in Section 2. Nevertheless, it is worth emphasizing here that we are filling spheres with balls, which is quite different from filling spheres with chains, or cycles with chains (the latter of which leads to the *homological Dehn function*). It turns out that we do indeed need to make use of other variants (namely, the *strong Dehn function* – see Section 2), but for us the primary object of most immediate geometric interest is the Dehn function as described above.

In this paper we are concerned with the following question: what is the possible isoperimetric behavior of groups, in various dimensions? For each positive integer k one defines the *k–dimensional isoperimetric spectrum*:

$$\mathrm{IP}^{(k)} \;=\; \{\, \alpha \in [1,\infty) \mid f(x) = x^{\alpha} \text{ is equivalent to a } k\text{–dimensional Dehn function} \,\}.$$

Until recently, relatively little was known about $\mathrm{IP}^{(k)}$, especially when $k \geqslant 3$. A few results concerning $\mathrm{IP}^{(2)}$ were known. Alonso, Bogley, Burton, Pride and Wang [1, 27, 26] have shown that $\mathrm{IP}^{(2)}$ contains infinitely many points in the interval $[3/2,2)$, and they located various lower and upper bounds throughout $[2,\infty)$. Also Brady and Bridson [4] and Bridson [7] have shown that $\mathrm{IP}^{(2)} \cap [3/2,2)$ is dense in $[3/2,2)$ and that $2,3 \in \mathrm{IP}^{(2)}$.

The recent paper of Brady, Bridson, Forester and Shankar [5] established that $\mathrm{IP}^{(k)}$ is dense in $[1+\frac{1}{k},\infty)$ and contains all rational numbers in this range. The endpoint $1+\frac{1}{k}$ corresponds to the isoperimetric inequality represented by spheres in Euclidean space. The main purpose of the present paper is to address the sub-Euclidean range $(1,1+\frac{1}{k})$ and establish the existence of isoperimetric exponents throughout this interval, for $k \geqslant 2$.

To state our results we need some notation. If A is a non-singular $n \times n$ integer matrix, let G_A denote the ascending HNN extension of $\mathbb{Z}^n$ with monodromy A. Our first result is the following.

Theorem 1.1 *Let A be a 2×2 integer matrix with eigenvalues λ, μ such that $\lambda > 1 > \mu$ and $\lambda\mu > 1$. Then the 2–dimensional Dehn function of G_A is equivalent to $x^{2+\log_\lambda(\mu)}$.*

In Section 7 we show that the exponents arising in the theorem are dense in the interval $(1, 2)$. Thus, roughly half of these groups have sub-Euclidean filling volume for 2–spheres, occupying densely the desired range of possible behavior.

Given an $n \times n$ matrix A, the *suspension* ΣA of A is the $(n+1) \times (n+1)$ matrix obtained by direct sum with the 1×1 identity matrix. Since $G_{\Sigma A} \cong G_A \times \mathbb{Z}$, results from [5] imply the following (see Section 6 for details).

Theorem 1.2 *Let G_A be as in Theorem 1.1. Then the $(i+2)$–dimensional Dehn function of $G_{\Sigma^i A}$ is equivalent to x^s where $s = \frac{(i+1)\alpha - i}{i\alpha - (i-1)}$ and $\alpha = 2 + \log_\lambda(\mu)$.*

Given that the numbers α are dense in the interval $(1, 2)$, it follows that the exponents s are dense in $(1, (i+2)/(i+1))$. Together with Corollary E of [5], we have the following result, illustrated in Figure 1.

Corollary 1.3 $\mathrm{IP}^{(k)}$ *is dense in $[1, \infty)$ for $k \geqslant 2$.*

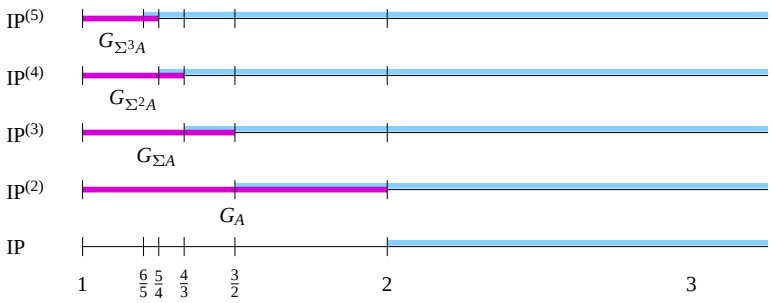

Figure 1: Isoperimetric exponents of $G_{\Sigma^i A}$. The blue intervals indicate isoperimetric exponents for the groups constructed in [5].

Methods

The methods used here to establish isoperimetric inequalities for G_A are quite different from those used in [5]. In the latter work, a slicing argument was used to estimate volume based on information coming from one-dimensional Dehn functions. This approach is rather less promising in the sub-Euclidean realm, since there are no one-dimensional Dehn functions there to reduce to. (Reducing to larger Dehn functions does not seem feasible, at least by similar methods.)

Instead we must find and measure least-volume fillings of 2–spheres in G_A directly, using properties of the particular geometry of this group. We work with a piecewise Riemannian cell complex with a metric locally modeled on a solvable Lie group $\mathbb{R}^2 \rtimes \mathbb{R}$. This metric is particularly simple from the point of view of the given coordinates, and these preferred coordinates make possible various volume and area calculations that are central to our arguments.

The preferred coordinates just mentioned do not behave well combinatorially, however. Coordinate lines pass through cells in an aperiodic manner, and this cannot be remedied by simply changing the cell structure. If one attempts to measure volume combinatorially, counting cells by passing between cells and their neighbors in an organized fashion (as with "t–corridor" arguments, for example), one loses the advantage of the preferred coordinates conferred by the special geometry of these groups. To count cells, therefore, we use integration and divide by the volume of a cell.

The combinatorial structure is still relevant, however. The piecewise Riemannian model is not a manifold, and its branching behavior is a prominent feature of the geometry of G_A. In order to make clean transitions between the combinatorial and Riemannian viewpoints, we use the transversality technology of Buoncristiano, Rourke and Sanderson [8]. This provides the appropriate notion of van Kampen diagrams for higher-dimensional spheres and fillings. Transversality also helps in dealing with singular maps, which otherwise present technical difficulties.

One other technical matter deserves mention: in order to apply results of [5] to deduce Theorem 1.2, we are obliged to find bounds for the *strong Dehn function*, which encodes uniform isoperimetric inequalities for fillings of surfaces by arbitrary 3–manifolds. See Section 2 for definitions and results concerning the strong Dehn function.

Remark/Conjecture 1.4 The groups G_A in Theorem 1.1 were classified up to quasi-isometry by Farb and Mosher [14]. At the time, none of the usual quasi-isometry invariants could distinguish these groups, but the two-dimensional Dehn function apparently does so quite well. We conjecture that it is a complete invariant for this class

of groups. What is missing is the knowledge that the real number $\log_\lambda(\mu)$ determines the diagonal matrix $\left(\begin{smallmatrix}\lambda & 0\\ 0 & \mu\end{smallmatrix}\right)$ up to a rational power. One needs to take into account the specific assumptions on the integer matrix A (eg having a contracting eigenspace), to rule out examples such as $\left(\begin{smallmatrix}4 & 0\\ 0 & 2\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}9 & 0\\ 0 & 3\end{smallmatrix}\right)$.

Acknowledgments Brady's work was funded in part by NSF grant DMS-0505707, and Forester's by NSF grant DMS-0605137.

2 Preliminaries

In this section we discuss in detail some of the key notions needed to carry out the proofs of the theorems. First we give a brief account of the transversality theory of Buoncristiano, Rourke and Sanderson. Then we discuss volume, Dehn functions of various types, and some basic results concerning these.

Handles and transverse maps

Using transversality, a map from a manifold to a cell complex can be put into a nice form, called a transverse map. Transverse maps induce *generalized handle decomopositions* of manifolds, which will play the role of van Kampen diagrams in higher dimensions. Whereas *admissible maps* were used for this purpose in [5], transverse maps have additional structure, incorporating combinatorial information dependent on the way cells meet locally in the target complex.

An *index i handle* (or *generalized handle*) of dimension n is a product $\Sigma^i \times D^{n-i}$, where Σ^i is a compact, connected i–dimensional manifold with boundary, and D^{n-i} is a closed disk. Let M be a closed n–manifold. A *generalized handle decomposition* of M is a filtration $\emptyset = M^{(-1)} \subset M^{(0)} \subset \cdots \subset M^{(n)} = M$ by codimension-zero submanifolds, such that for each i, $M^{(i)}$ is obtained from $M^{(i-1)}$ by attaching finitely many index i handles, as follows. To attach a single handle $H = \Sigma^i \times D^{n-i}$, choose an embedding $h\colon\ \partial\Sigma^i \times D^{n-i} \to \partial M^{(i-1)}$ and form the manifold $M^{(i-1)} \cup_h H$. Note that handle attachment is always along $\partial\Sigma^i \times D^{n-i}$, and never along $\Sigma^i \times \partial D^{n-i}$. To attach several handles, we require that the attaching maps have disjoint images in $\partial M^{(i-1)}$, so that the order of attachment does not matter. Note that both $M^{(i-1)}$ and the individual handles H are embedded in $M^{(i)}$.

If every Σ^i is a disk then this is the usual notion of handle decomposition arising in classical Morse theory. Some new things can occur by varying Σ^i, however. For

instance, we allow Σ^i to be closed, in which case the attaching map is empty and $M^{(i-1)} \cup_h H$ is the disjoint union $M^{(i-1)} \sqcup H$. Such a handle is called a *floating handle*. For example, $M^{(0)}$ is formed from $M^{(-1)} = \emptyset$ by attaching (floating) 0–handles $D^0 \times D^n$, and $M^{(0)}$ is simply several copies of D^n. (The lowest-index handles will always be floating ones.) Another phenomenon is that handles may be embedded in M in topologically interesting ways, as in the following example.

Example 2.1 Given a closed orientable 3–manifold M, we may construct a generalized handle decomposition as follows. Let $K \subset M$ be a knot or link in M. Let $M^{(1)}$ be a regular neighborhood of K and declare each component to be a (floating) 1–handle. Let Σ be a Seifert surface for K, and let $\{\Sigma_j\}$ be the components of $\Sigma \cap (M - \text{int}(M^{(1)}))$. The 2–handles will be regular neighborhoods of the surfaces Σ_j in $M - \text{int}(M^{(1)})$. Lastly, the 3–handles will be the components of $M - \text{int}(M^{(2)})$. This decomposition has no 0–handles, and its 1–handles are (obviously) knotted.

Now suppose M is an n–manifold with boundary. A *generalized handle decomposition of M* is a pair of filtrations $\emptyset = M^{(-1)} \subset M^{(0)} \subset \cdots \subset M^{(n)} = M$ and $\emptyset = N^{(-1)} \subset N^{(0)} \subset \cdots \subset N^{(n-1)} = \partial M$ by codimension-zero submanifolds, such that:

(1) the filtration $\emptyset = N^{(-1)} \subset N^{(0)} \subset \cdots \subset N^{(n-1)} = \partial M$ is a generalized handle decomposition of ∂M,

(2) for each i, $M^{(i)}$ is obtained from $M^{(i-1)} \cup N^{(i-1)}$ by attaching finitely many index i handles, and

(3) each index $i-1$ handle of ∂M is a connected component of the intersection of ∂M with an index i handle of M. In particular, $N^{(i-1)} = \partial M \cap M^{(i)}$ for all i.

In (2) each handle $H = \Sigma^i \times D^{n-i}$ is attached via an embedding $h\colon (\partial\Sigma^i \times D^{n-i}) \to (\partial M^{(i-1)} \cup N^{(i-1)})$. As before, we require the images of the attaching maps of the index i handles to be disjoint. It follows that the individual i–handles are embedded in M, and are disjoint from each other.

Let $f\colon M \to X$ be a map from a compact n–manifold to a CW complex. We say that f is *transverse* to the cell structure of X if M has a generalized handle decomposition such that the restriction of f to each handle is given by projection onto the second factor, followed by the characteristic map of a cell of X. Thus, index i handles map to $(n-i)$–dimensional cells. In particular, M maps into the n–skeleton of X. In a transverse map there may be floating handles of any index, and it may not be possible to modify f to eliminate these. By the same token, one must always allow for the possibility of knotted handles.

One virtue of transverse maps is that they can easily be proved to exist. However, to accomplish this, we must assume additional structure on the target complex X. We say that X is a *transverse CW complex* if the attaching map of every cell is transverse to the cell structure of the skeleton to which it is attached. The main existence result is the following:

Transversality Theorem (Buoncristiano–Rourke–Sanderson) *Let M be a compact smooth manifold and $f\colon M \to X$ a continuous map into a transverse CW complex. Suppose $f|_{\partial M}$ is transverse. Then f is homotopic rel ∂M to a transverse map $g\colon M \to X$.*

The theorem includes the case where M is closed: all maps of closed manifolds can be made transverse by a homotopy.

This theorem is proved in Buoncristiano–Rourke–Sanderson [8] for PL manifolds, and the proof in the smooth case is entirely analogous. The proof is a step by step application of smooth transversality, applied to preimages of open cells (considered as smooth manifolds themselves), starting with the top dimensional cells and working down. The first stage of the argument, in which the 0–handles are constructed, is explained fully in the proof of Lemma 2.3 of [5]. This is precisely the construction of admissible maps (defined below).

Remark 2.2 In order to apply the theorem one needs transverse CW complexes. Any CW complex can be made transverse by successively homotoping the attaching maps of its cells (by the Transversality Theorem and induction on dimension); this procedure preserves homotopy type. Moreover, in this paper, the complex X that we use can be made transverse in a more direct and controlled way, preserving both its homeomorphism type and its partition into open cells; see Section 3 and Figure 3.

Admissible maps and combinatorial volume

Recall from [5] the definition of an *admissible map*: it is a map $f\colon M^n \to X^{(n)} \subset X$ such that the preimage of every open n–cell is a disjoint union of open n–dimensional balls in M, each mapped by f homeomorphically onto the n–cell. The *combinatorial volume* of an admissible map, denoted $\mathrm{Vol}^n(f)$, is the number of open balls mapping to n–cells.

It is clear that transverse maps are admissible: the interiors of 0–handles are open balls, and the rest of M maps into $X^{(n-1)}$. Conversely, if one applies the proof

of the transversality theorem to an admissible map to make it transverse, then the preimages of the n–cells will not change (except possibly by being shrunk slightly), and combinatorial volume is preserved. For this reason, given an admissible map, the closures of the open balls mapping to n–cells will be called 0–handles.

Note that in an admissible map, 0–handles may intersect each other in their boundaries. For example, if M has a cell structure, then the identity map is admissible, with 0–handles equal to the closures of the top-dimensional cells.

In [5, Lemma 2.3] it is shown that every map from a smooth or PL manifold is homotopic to an admissible map. This is a special case of the Transversality Theorem, though it is not required that the target CW complex be transverse. The existence of admissible maps can also be proved without relying on a smooth or PL structure; see Epstein [11, Theorem 4.3].

Volume reduction

In this paper, generalized handle decompositions (and transverse maps) will serve as higher-dimensional analogues of van Kampen diagrams. Indeed, in dimension 2, transverse maps already provide an alternative to the combinatorial approach to diagrams, and they have several advantages. This is the viewpoint taken in Rourke [23] and Stallings [25], for example. With van Kampen diagrams one often considers *reduced* diagrams, where no folded cell pairs occur. The same type of cancellation process also works for admissible and transverse maps. One such process is given as follows.

Let $f\colon M^n \to X$ be an admissible map, and let $H_0, H_1 \subset M$ be 0–handles, and $\alpha \subset M - (\mathrm{int}(H_0) \cup \mathrm{int}(H_1))$ a 1–dimensional submanifold homeomorphic to an interval, with endpoints in H_0 and H_1 (we also allow the degenerate case in which α is a point in $H_0 \cap H_1$). Suppose that f maps α to a point and maps H_0 and H_1 to the same n–cell, with opposite orientations (relative to a neighborhood of $H_0 \cup \alpha \cup H_1$, which is always orientable). A typical example occurs when f is transverse and α is a fiber of a 1-handle joining H_0 and H_1.

Since H_0 and H_1 are 0-handles, there are homeomorphisms $h_i\colon H_i \to D^n$ such that $f|_{H_i} = \Phi \circ h_i$ for some characteristic map $\Phi\colon D^n \to X$. Now delete interiors of H_i from M to obtain M' with new boundary spheres S_i. Next delete the interior of a regular neighborhood $I \times D^{n-1}$ of α in M' (parametrized so that $f|_{\{0\}\times D^{n-1}} = f|_{\{1\}\times D^{n-1}}$). The new boundary becomes a union of two disks D_i and an annulus $A = I \times S^{n-2}$. Now collapse A to S^{n-2} and identify D_0 with D_1 via $h_0^{-1} \circ h_1$, to form M''. This new

space maps to X by f, and there is a homeomorphism $g\colon M \to M''$. Now $f \circ g$ is an admissible map $M \to X$ with two fewer 0-handles. Note that the other 0-handles are unchanged. If desired, this new map can then be made transverse, with the same 0-handles, and with its (lowered) volume unchanged.

Remark 2.3 There is, in fact, a more general procedure for cancelling H_0 and H_1 that does not require α to map to a point. This procedure is due to Hopf [19] and a detailed treatment was given by Epstein [11]. If X is 2–dimensional then the more general procedure is not particularly useful: new 0–handle pairs can be created when cancelling H_0 and H_1, and volume may fail to decrease. In higher dimensions, however, no new 0–handle pairs are created and the volume will always decrease by 2.

Riemannian volume

If N is a smooth manifold, M an oriented Riemannian manifold of the same dimension, and $f\colon N \to M$ a smooth map, then the *volume* of f can be defined. Following Gromov [17, Remarks 2.7 and 2.8$\frac{1}{2}$], let ν_M be the volume form on M and choose any Riemannian metric on N. We define

$$\mathrm{RVol}(f) \;=\; \int_N f^*(|\nu_M|).$$

The integral is independent of the choice of metric on N, by the change of variables formula. Note that we are using $|\mathrm{vol}|(f)$, not $\mathrm{vol}(f)$, in the notation of [17]. (The latter allows cancellation of volume, which is not appropriate in our setting.) In fact, we need not assume that M is oriented, since $|\nu_M|$ is still defined. If $\dim N = 2$ then RVol is also denoted RArea.

If f is an immersion then this definition amounts to giving N the pullback metric and taking the volume of N. More generally, if f fails to be an immersion at some $x \in N$, then $f^*(|\nu_M|)$ is zero at x, and does not contribute to volume. Hence, $\mathrm{RVol}(f)$ is the volume of the pullback metric on $U \subset N$, the set on which f is an immersion. Note that U is open, and hence is a Riemannian manifold. Generically, U has full measure in N when $\dim N \leqslant \dim M$; see eg Gromov [15, 1.3.1].

From this perspective, we can now define $\mathrm{RVol}(f)$ when $\dim N \neq \dim M$. We define $\mathrm{RVol}(f)$ to be the volume of $U \subset N$, the set on which f is an immersion, with the pullback metric. Note that $\mathrm{RVol}(f)$ measures *n–dimensional volume*, where $n = \dim N$.

Lastly, we wish to extend the definition of volume to allow a piecewise Riemannian CW complex in place of M. The complex $\widetilde{X}$ that interests us is a 3–complex with

branching locus a 2–manifold, homeomorphic to the product of $\mathbb{R}^2$ with a simplicial tree. In a neighborhood of any singular point one sees a union of half-spaces joined along their boundaries, naturally grouped into two collections, with a well defined common tangent space at the singular point. The situation is similar to that of a train track, or a branched surface from lamination theory (eg Calegari [10, Section 6.3]). There is a smooth structure, and $\widetilde{X}$ comes equipped with an immersion $q\colon \widetilde{X} \to M$ onto a Riemannian manifold M. (This immersion is not locally injective, but is injective on tangent spaces.) The Riemannian metric on $\widetilde{X}$ is the pullback under q of the metric on M. The volume $\mathrm{RVol}(f)$ can now be defined directly (as above) using this metric on $\widetilde{X}$, or equivalently by defining $\mathrm{RVol}(f) = \mathrm{RVol}(q \circ f)$.

Remarks 2.4 (1) If $\dim N > \dim M$ (or $\dim N > \dim \widetilde{X}$) then $\mathrm{RVol}(f)$ is zero, since f is an immerison nowhere. Similarly, if f factors through a manifold of smaller dimension, then the volume is zero.

(2) Any transverse map $f\colon N \to \widetilde{X}$ is piecewise smooth, and is a submersion on each handle. It will be an immersion only on the 0–handles. This latter statement also holds for admissible maps, since the complement of the 0–handles is mapped into a lower-dimensional skeleton.

Remark 2.5 We will be interested in finding least-volume maps extending a given boundary map. If the set of volumes of n–cells of a piecewise Riemannian CW complex is finite, then least-volume transverse maps of n–manifolds exist in any homotopy class. This is because the Riemannian volume of a transverse map is a *positive* linear combination of numbers in this set, and hence the set of such volumes is discrete, and well-ordered.

Dehn functions

Here we recall the definition of the n–dimensional Dehn function of a group from [5]. Note that these definitions all use combinatorial volume. Given a group G of type $\mathcal{F}_{n+1}$, fix an aspherical CW complex X with fundamental group G and finite $(n+1)$–skeleton (the existence of such an X is the meaning of "type $\mathcal{F}_{n+1}$"). Let $\widetilde{X}$ be the universal cover of X. If $f\colon S^n \to \widetilde{X}$ is an admissible map, define the *filling volume* of f to be the minimal volume of an admissible extension of f to B^{n+1}:

$$\mathrm{FVol}(f) \;=\; \min\{\, \mathrm{Vol}^{n+1}(g) \mid g\colon B^{n+1} \to \widetilde{X},\ g|_{\partial B^{n+1}} = f \,\}.$$

Note that extensions exist since $\pi_n(\widetilde{X})$ is trivial, and any extension can be made admissible, by [5, Lemma 2.3]. We define the *n–dimensional Dehn function* of X to be

$$\delta^{(n)}(x) \;=\; \sup\{\,\mathrm{FVol}(f) \mid f\colon\, S^n \to \widetilde{X},\ \mathrm{Vol}^n(f) \leqslant x\,\}.$$

Again, the maps f are assumed to be admissible.

In Alonso–Wang–Pride [2] it was shown that $\delta^{(n)}(x)$ is finite for each $x \in \mathbb{N}$, and that, up to coarse Lipschitz equivalence, $\delta^{(n)}(x)$ depends only on G. Thus the Dehn function will sometimes be denoted $\delta^{(n)}_G(x)$. (Recall that functions $f, g\colon\, \mathbb{R}_+ \to \mathbb{R}_+$ are *coarse Lipschitz equivalent* if $f \preccurlyeq g$ and $g \preccurlyeq f$, where $f \preccurlyeq g$ means that there is a positive constant C such that $f(x) \leqslant C\,g(Cx) + Cx$ for all $x \geqslant 0$.) If we wish to specify $\delta^{(n)}(x)$ exactly, we may denote it as $\delta^{(n)}_X(x)$.

Taking $n = 1$ yields the usual Dehn function $\delta(x)$ of a group G.

The strong Dehn function

The notion of n–dimensional Dehn function was modified in [5] to allow fillings by compact manifolds other than the ball B^{n+1}. In this way, every compact manifold pair $(M, \partial M)$ gave rise to a Dehn function $\delta^M(x)$. Several of the main results proved in [5] had hypotheses and conclusions involving the functions $\delta^M(x)$ "for all n–manifolds M." An equivalent way of formulating these results is by means of the *strong Dehn function*, defined as follows.

Given a compact $(n+1)$–manifold M and an admissible map $f\colon\, \partial M \to \widetilde{X}$, define

$$\mathrm{FVol}^M(f) \;=\; \min\{\,\mathrm{Vol}^{n+1}(g) \mid g\colon\, M \to \widetilde{X} \text{ admissible},\ g|_{\partial M} = f\,\}$$

and

$$\Delta^{(n)}(x) \;=\; \sup\{\,\mathrm{FVol}^M(f) \mid (M, \partial M) \text{ is a compact } (n+1)\text{–manifold},\ f\colon\, \partial M \to \widetilde{X} \text{ admissible},\ \mathrm{Vol}^n(f) \leqslant x\,\}.$$

We call $\Delta^{(n)}(x)$ the *strong n–dimensional Dehn function* of X. Note that the manifolds M appearing in the definition are not assumed to be connected. The statement $\Delta^{(n)}(x) \leqslant y$ means that for every compact manifold $(M, \partial M)$ and every admissible map $f\colon\, \partial M \to \widetilde{X}$ of volume at most x, there is an admissible extension to M of volume at most y. In particular, the bound y is uniform for all topological types of fillings (hence the word "strong"). Note that this is very different from *homological* Dehn functions, where only a single filling by an $(n+1)$–chain is needed, of some topological type.

The strong Dehn function has two principal features. The first is that it behaves well with respect to splittings and mapping torus constructions (as does the homological Dehn function). The next two theorems below are examples of this phenomenon. The second is that it (clearly) satisfies

(1) $$\delta^{(n)}(x) \leqslant \Delta^{(n)}(x)$$

and hence it may be used to establish upper bounds for $\delta^{(n)}(x)$. To this end, the following two theorems are proved in [5] (Theorems 7.2 and 8.1).

Theorem 2.6 (Stability for Upper Bounds) *Let X be a finite aspherical CW complex of dimension at most $n+1$. Let $f\colon X \to X$ be a π_1–injective map and let Y be the mapping torus of X using f. Then $\Delta_Y^{(n+1)}(x) \leqslant \Delta_X^{(n)}(x)$.*

Thus, any upper bound for $\Delta_X^{(n)}(x)$ remains an upper bound for $\Delta_Y^{(n+1)}(x)$. A similar result holds more generally (with the same proof) if Y is the total space of a graph of spaces whose vertex and edge spaces satisfy the hypotheses of X. Then the conclusion is that $\Delta_Y^{(n+1)}(x) \leqslant C\,\Delta_X^{(n)}(x)$ for some $C > 0$.

The next result provides a better bound in a special case.

Theorem 2.7 (Products with S^1) *Let X be a finite aspherical CW complex of dimension at most $n+1$. If $\Delta_X^{(n)}(x) \leqslant Cx^s$ for some $C > 0$ and $s \geqslant 1$ then $\Delta_{X\times S^1}^{(n+1)}(x) \leqslant C^{1/s}x^{2-1/s}$.*

It turns out that for $n \geqslant 3$ and for $n = 1$, there is no significant difference between the strong and ordinary Dehn functions. The precise relation between them is stated in the next theorem, which was essentially proved already in Remark 2.5(4) and Lemma 7.4 of [5].

However, we do indeed need to work specifically with the strong Dehn function in dimension 2, since we wish to apply Theorem 2.7 above. This case forms the base of the induction argument we use to show that $\mathrm{IP}^{(n)}$ is dense for all $n \geqslant 2$.

A function $f\colon \mathbb{N} \to \mathbb{N}$ is *superadditive* if $f(a)+f(b) \leqslant f(a+b)$ for all $a, b \in \mathbb{N}$. The *superadditive closure* of f is the smallest superadditive g such that $f(x) \leqslant g(x)$ for all x. An explicit recursive definition of g is given by

$$g(0) = f(0), \quad g(x) = \max\{\{g(i)+g(x-i) \mid i = 1,\dots,x-1\} \cup \{g(0)+f(x)\}\}.$$

It is easy to verify that $\Delta^{(n)}(x)$ is always superadditive, by considering fillings by non-connected manifolds.

Theorem 2.8 (Brady–Bridson–Forester–Shankar) $\Delta^{(n)}_X(x)$ *is the superadditive closure of* $\delta^{(n)}_X(x)$ *for* $n \geqslant 3$ *and for* $n = 1$.

It is not known whether there exist groups G for which $\delta^{(n)}_G(x)$ is not superadditive (up to coarse Lipschitz equivalence). Indeed, when $n = 1$, Sapir has conjectured that this does not occur [18]. So in all known examples, $\Delta^{(n)}$ and $\delta^{(n)}$ agree (for $n \geqslant 3$ or $n = 1$).

In contrast, Young [28] has shown that the statement of the theorem is false when $n = 2$. Specifically, he shows that for a certain group G, the strong Dehn function $\Delta^{(2)}_G(x)$ is not bounded by a recursive function, whereas $\delta^{(2)}_G(x)$ always satisfies such a bound, by Papasoglu [22]. The superadditive closure will inherit this property, since it is computable from $\delta^{(2)}_G(x)$.

Proof Let $s(x)$ be the superadditive closure of $\delta^{(n)}(x)$.

If $n = 1$ then the proof of Lemma 7.4 of [5] shows directly that for any compact 2–manifold M, one has $\delta^M(x) \leqslant \delta^{D^2 \sqcup \cdots \sqcup D^2}(x)$, where the number of disks equals the number of boundary components of M. For each admissible $f\colon S^1 \sqcup \cdots \sqcup S^1 \to X$ with length $x = \sum_i x_i$ we have $\mathrm{FVol}^{D^2 \sqcup \cdots \sqcup D^2}(f) \leqslant \sum_i \delta^{(1)}(x_i) \leqslant s(x)$, and so $\delta^M(x) \leqslant s(x)$. Therefore $\Delta^{(1)}(x) \leqslant s(x)$. Since $\Delta^{(1)}(x)$ is superadditive and $\delta^{(1)}(x) \leqslant \Delta^{(1)}(x)$, it follows that $\Delta^{(1)}(x) = s(x)$.

If $n \geqslant 3$ then the argument given in Remark 2.5(4) of [5] applies. Let $\{N_i\}$ be the components of ∂M and suppose that $g_i\colon N_i \to X$ are admissible maps of volume x_i, with union $g\colon \partial M \to X$ of volume $x = \sum_i x_i$. By the argument given in [5], for each i there is an admissible homotopy of $(n+1)$–dimensional volume at most $\delta^{(n)}(x_i)$ to an admissible map $g_i'\colon N_i \to X$ with image inside $X^{(n-1)}$. The union of these maps can be filled by a map $M \to X^{(n)}$, since $X^{(n-1)}$ is contractible inside $X^{(n)}$. This filling has zero $(n+1)$–dimensional volume, and hence $\mathrm{FVol}^M(g) \leqslant \sum_i \delta^{(n)}(x_i) \leqslant s(x)$. Since M and g were arbitrary, we have $\Delta^{(n)}(x) \leqslant s(x)$, and hence $\Delta^{(n)}(x) = s(x)$. □

Remark 2.9 (Lower bounds) As noted earlier, the strong Dehn function can be used to bound $\delta^{(n)}(x)$ from above. For a lower bound one needs explicit information about $\mathrm{FVol}(f)$ for admissible maps $f\colon S^n \to \widetilde{X}$. That is, one needs to identify *least-volume* extensions $g\colon B^{n+1} \to \widetilde{X}$. Suppose $\dim \widetilde{X} = n+1$ and $H_{n+1}(\widetilde{X}; \mathbb{Z}) = 0$. Then a simple homological argument, sketched in Remarks 2.2 and 2.6 of [5], shows that g is least-volume if g is injective on the interiors of 0–handles (i.e. no two 0–handles map to the same cell of $\widetilde{X}$). For convenience we provide the full argument here.

Let $C_{n+1}(\widetilde{X})$ be the cellular chain group for $\widetilde{X}$. Given an oriented manifold M^{n+1} and a transverse map $f\colon M^{n+1} \to \widetilde{X}$, there is a chain $[f] \in C_{n+1}(\widetilde{X})$ defined as follows. For each $(n+1)$–cell e_α, let σ_α be the corresponding generator of $C_{n+1}(\widetilde{X})$ and define $d_\alpha(f)$ to be the local degree of f at e_α (i.e. the number of 0–handles of f mapping to e_α, counted with respect to orientations). We define $[f] = \sum_\alpha d_\alpha(f)\sigma_\alpha$. Note that the boundary of $[f]$ in $C_n(\widetilde{X})$ is simply $[f|_{\partial M}]$. (Here the transversality structure is used: 0–handles in ∂M are joined to 0–handles in M by 1–handles, compatibly with boundaries of characteristic maps of cells in $\widetilde{X}$.)

Now suppose that $g\colon B^{n+1} \to \widetilde{X}$ is injective on 0–handles, and $h\colon B^{n+1} \to \widetilde{X}$ is another transverse map with $h|_{S^n} = g|_{S^n}$. These maps together define a transverse map $g - h\colon S^{n+1} \to \widetilde{X}$ by considering S^{n+1} as a union of two balls, with the orientation on one of the balls reversed. We have $[g-h] = [g] - [h]$ in $C_{n+1}(\widetilde{X})$, and so $\partial[g-h] = \partial[g] - \partial[h] = 0$, and $[g-h]$ is a cycle. Since $H_{n+1}(\widetilde{X}) = 0$ and $C_{n+2}(\widetilde{X}) = 0$, this cycle must be zero in $C_{n+1}(\widetilde{X})$. That is, $g-h$ has zero local degree at every $(n+1)$–cell. Hence $d_\alpha(g) = d_\alpha(h)$ for all α.

The injectivity assumption on g implies that $\mathrm{Vol}^{n+1}(g) = \sum_\alpha |d_\alpha(g)|$. Then we have

$$\mathrm{Vol}^{n+1}(h) \;\geqslant\; \sum_\alpha |d_\alpha(h)| \;=\; \sum_\alpha |d_\alpha(g)| \;=\; \mathrm{Vol}^{n+1}(g),$$

and hence g is least-volume.

3 The groups G_A and their model spaces

The model manifold M

Let M be the manifold $\mathbb{R}^3$ with the metric $ds^2 = \lambda^{-2z}dx^2 + \mu^{-2z}dy^2 + dz^2$, where $\lambda > 1$, $\mu < 1$, and $\lambda\mu > 1$. This is the left-invariant metric for the solvable Lie group $\mathbb{R}^2 \rtimes \mathbb{R}$, with $z \in \mathbb{R}$ acting on $\mathbb{R}^2$ by the matrix $\left(\begin{smallmatrix} \lambda^z & 0 \\ 0 & \mu^z \end{smallmatrix}\right)$. The geometry of M has much in common with that of SOL (the case $\lambda\mu = 1$), but with some important differences.

The group G_A and its model space X

Let $A \in M_2(\mathbb{Z})$ be a hyperbolic matrix with eigenvalues $\lambda > 1$ and $\mu < 1$ and determinant $d = \lambda\mu > 1$. Let $B \in GL_2(\mathbb{R})$ diagonalize A, so that $BAB^{-1} = \left(\begin{smallmatrix} \lambda & 0 \\ 0 & \mu \end{smallmatrix}\right)$. Call this diagonal matrix D. Then D preserves the lattice $\Gamma \subset \mathbb{R}^2$, defined to be the image of $\mathbb{Z} \times \mathbb{Z}$ under B.

Let G_A be the ascending HNN extension of $\mathbb{Z}\times\mathbb{Z}$ with monodromy A. That is,

$$G_A \;=\; \langle\, \mathbb{Z}\times\mathbb{Z}, t \mid tvt^{-1} = Av \;\text{ for all } v\in\mathbb{Z}\times\mathbb{Z}\,\rangle.$$

The matrix B defines an isomorphism from G_A to the (non-discrete) subgroup of $\mathbb{R}^2\rtimes\mathbb{R}$ generated by Γ and $1\in\mathbb{R}$ (corresponding to the stable letter $t\in G_A$).

The groups G_A are the main examples that interest us in this paper; our chief task will be determining their 2–dimensional Dehn functions $\delta^{(2)}(x)$. For this we need to construct a geometric model for G_A. Note that $\mathbb{R}^2\rtimes\mathbb{R}$ cannot serve as a model since the subgroup G_A is not discrete. (Indeed, this Lie group is not quasi-isometric to *any* finitely generated group, by Eskin, Fisher and Whyte [12].)

Topologically, our model is formed from $T^2\times I$ by glueing $T^2\times 0$ to $T^2\times 1$ by the d–fold covering map $T_A\colon\ T^2\to T^2$ induced by A. To put a piecewise Riemannian metric on this space, we use the geometry of M as follows. The construction is analogous to building the standard presentation 2–complex of a Baumslag–Solitar group from a "horobrick" in the hyperbolic plane, as in Farb–Mosher [13].

Let $Q\subset\mathbb{R}^2$ be the parallelogram spanned by the generators of Γ. Then $Q\times[0,1]$ is a fundamental domain for the action of Γ on $\mathbb{R}^2\times[0,1]\subset\mathbb{R}^2\rtimes\mathbb{R}$, with quotient homeomorphic to $T^2\times[0,1]$. The isometry $\mathbb{R}^2\times 0\to\mathbb{R}^2\times 1$ given by $(x,y,0)\mapsto(\lambda x,\mu y,1)$ is Γ–equivariant and induces a local isometry $\mathbb{R}^2/\Gamma\times 0\to\mathbb{R}^2/\Gamma\times 1$. This local isometry agrees precisely with the map $T_A\colon\ T^2\to T^2$ under the identification of $\mathbb{R}^2/\Gamma$ with T^2 induced by B. Thus, identifying opposite sides of $Q\times[0,1]$ to obtain a copy of $T^2\times[0,1]$, the glueing $T^2\times 0\to T^2\times 1$ is locally isometric, and the model for G_A is a piecewise Riemannian space. Call it X, and its universal cover $\widetilde{X}$.

Figure 2 below shows Q and the locally isometric glueing map for the example $A=\left(\begin{smallmatrix}4&2\\1&1\end{smallmatrix}\right)$. The diagonal matrix stretches horizontally and compresses vertically.

3.1 The cover $\widetilde{X}$ is tiled by isometric copies of $Q\times[0,1]$, with tiles meeting isometrically along faces. A generic point in the top face $Q\times 1$ of a tile meets d tiles in their bottom faces; side faces are joined in pairs. Topologically, $\widetilde{X}$ is a branched space homeomorphic to $\mathbb{R}^2\times T$, where T is the Bass–Serre tree corresponding to the splitting of G_A as an ascending HNN extension. The G_A–tree T has a fixed end η and there is an equivariant map $h_0\colon\ T\to\mathbb{R}$, sending η to $-\infty$ and all other ends to ∞, such that the induced G_A–action on $\mathbb{R}$ is by integer translations. The preimage of $\mathbb{Z}$ under this map is the set of vertices of T.

There is a locally isometric surjection $q\colon\ \widetilde{X}\to M$ which, viewed via the homeomorphisms $\widetilde{X}\cong\mathbb{R}^2\times T$ and $M\cong\mathbb{R}^2\times\mathbb{R}$, is given by the identity on $\mathbb{R}^2$ and the map

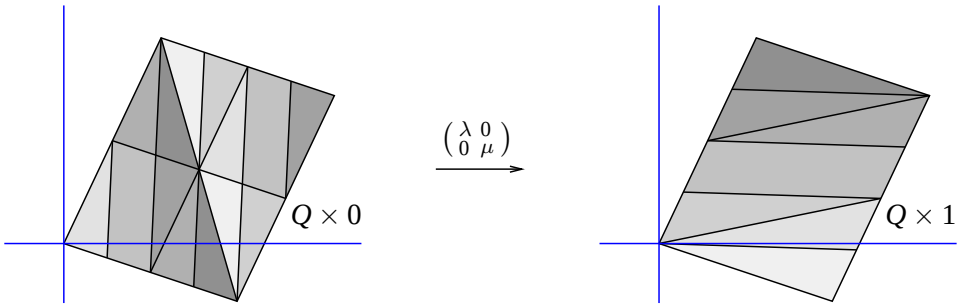

Figure 2: The region Q and the glueing map given by the diagonalized form of $A = \left(\begin{smallmatrix} 4 & 2 \\ 1 & 1 \end{smallmatrix}\right)$. Also shown is a cell structure (discussed below) for which this map is combinatorial.

$h_0 \colon T \to \mathbb{R}$ described above. The metric on $\widetilde{X}$ may be viewed as the pullback metric of M under this map. In particular, for any compact manifold W and any piecewise smooth map $f \colon W \to \widetilde{X}$, we have $\mathrm{RVol}(f) = \mathrm{RVol}(q \circ f)$.

If $L \subset T$ is a line mapping homeomorphically to $\mathbb{R}$ under h_0, then the subspace $\mathbb{R}^2 \times L \subset \widetilde{X}$ is isometric to M. This situation is completely analogous to that of the solvable Baumslag–Solitar groups, whose standard geometric models contain copies of the hyperbolic plane (cf Farb and Mosher [13]).

The map $h_0 \colon T \to \mathbb{R}$ also defines a *height function* $h \colon \widetilde{X} \to \mathbb{R}$ by composing with the projection $\widetilde{X} \cong \mathbb{R}^2 \times T \to T$.

Cell structure

The basic cell structure on X is the usual mapping torus cell structure, induced by the standard cell decomposition for the torus, but we will need to modify the attaching maps to make it a transverse CW complex.

First, consider $Q \times [0,1]$ combinatorially as a cube and give it the product cell structure (with eight 0–cells, twelve 1–cells, six 2–cells, and one 3–cell). The side-pairings are compatible with this structure, so we have a cell structure on $T^2 \times [0,1]$. Now subdivide the top and bottom faces $T^2 \times \{0,1\}$ into finitely many cells so that $T_A \colon T^2 \times 0 \to T^2 \times 1$ maps open cells homeomorphically to open cells (i.e. T_A becomes a *combinatorial map*). Note that $T^2 \times 0$ will have d times as many 2–cells as $T^2 \times 1$, since T_A is a d–fold covering. The pattern of subdivision is obtained by taking intersections of cells of $T^2 \times 1$ with cells of $T_A(T^2 \times 0)$. See Figure 2 for the example $A = \left(\begin{smallmatrix} 4 & 2 \\ 1 & 1 \end{smallmatrix}\right)$. Since T_A takes cells to cells, we now have a cell structure on X.

Next we make the cell structure transverse. In this case, the transversality procedure does not change the homeomorphism type of X, or even its partition into open cells. Thus, the piecewise Riemannian metric will still exist, exactly as described, with either cell structure.

Every map $S^0 \to X^{(0)}$ is transverse, so the 1–skeleton $X^{(1)}$ is already a transverse CW complex. For the 2–skeleton, note that for each attaching map $S^1 \to X^{(1)}$ in the original cell structure, there is a realization of S^1 as a graph such that the map is a graph morphism. To make this map transverse, expand each vertex into a closed interval (a 1–handle) to form a slightly larger circle. Let the new attaching map first collapse these intervals back into vertices, and then map to $X^{(1)}$ by the original attaching map. We have simply introduced some "slack" at the vertices. The 2–skeleton and its partition into open cells has not changed.

For the attaching map $S^2 \to X^{(2)}$ of the 3–cell, note again that S^2 has a cell structure for which this map is combinatorial (this is a property of our particular complex X). Expand every 0–cell into a small disk (a 2–handle) and then expand every 1–cell into a rectangle (a 1–handle), to abtain a new copy of S^2. The new transverse attaching map will collapse these new handles to 0– and 1–cells and then map to $X^{(2)}$ as before. See Figure 3. Again, the topology of X is unchanged. (This amounts to a claim that performing the collapses described above in the boundary of a ball results again in a ball.)

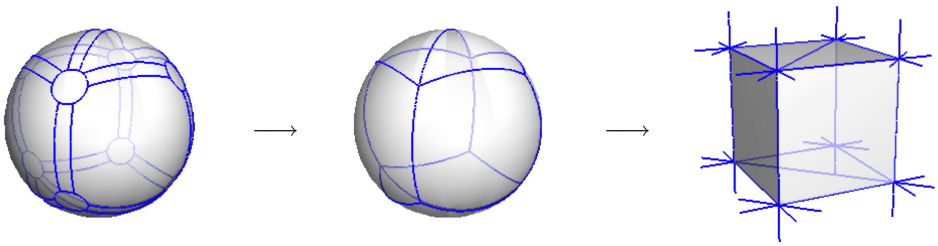

Figure 3: Transverse 3–cell attachment. The rightmost map is the original attaching map; the composition is the new (transverse) one.

The universal cover $\widetilde{X}$ is given the induced cell structure. Note that the closures of the 3–cells are exactly the copies of $Q \times [0,1]$ tiling $\widetilde{X}$ mentioned earlier. Also note that every 2–cell is either *horizontal* or *vertical*: in the product $\mathbb{R}^2 \times T$, it either projects to a point in T or to a line segment in $\mathbb{R}^2$. In the latter case, the projection of the 2–cell in T is exactly an edge.

4 The upper bound

We proceed now to establish an upper bound for the strong Dehn function $\Delta^{(2)}(x)$ of the group G_A.

Let W be a compact 3–manifold with boundary and $f\colon \partial W \to \widetilde{X}$ an admissible map, which we may make transverse without changing its combinatorial area (by a homotopy inside $\widetilde{X}^{(2)}$, of zero volume). Now let $g\colon W \to \widetilde{X}$ be a transverse extension of f of smallest Riemannian volume (cf Remark 2.5).

We need to measure the combinatorial volume of g and bound it in terms of the area of f. Note that every 0–handle of W has the same Riemannian volume, equal to the volume V of the single 3–cell in X. Thus, to count the 0–handles, we will instead measure the Riemannian volume of g by integration and divide by V. It turns out that the geometry of $\widetilde{X}$ is well-suited to this kind of measurement. We will also work with the Riemannian area of f, but again the relation to combinatorial area causes no difficulty.

The embedded case

First we discuss a special case in order to clarify the geometric ideas, before incorporating transverse maps into the argument. We will assume that W is a subcomplex of $\widetilde{X}$, with g the inclusion map.

Since W is a manifold, every 2–cell of W is either in ∂W or is adjacent to two 3–cells of W. Let $F \subset W$ (the *fold set*) be the smallest subcomplex whose 2–cells are the horizontal 2–cells σ such that $\sigma \not\subset \partial W$ and both adjacent 3–cells are *above* σ with respect to the height function $h\colon \widetilde{X} \to \mathbb{R}$. (The fold set may be empty, of course.)

Proposition 4.1 $\mathrm{RVol}(W) \leqslant \frac{1}{\ln(\lambda\mu)}(\mathrm{Area}(\partial W) + 2\,\mathrm{Area}(F))$.

Proof In M, integrating the volume element $(\lambda\mu)^{-z}\,dxdydz$ along a vertical ray from $z = 0$ to $z = \infty$ yields $\frac{1}{\ln(\lambda\mu)}$ times $dxdy$, the horizontal area element at the initial point of the ray. Also, at any point of ∂W, the surface area element is greater than or equal to the horizontal area element.

Consider a flow on $\widetilde{X} \cong \mathbb{R}^2 \times T$ which is towards the end η in the T factor and the identity in $\mathbb{R}^2$. This flow is semi-conjugate (by q) to a flow in M which is directly downward. Under this flow, every point p of W leaves W, either through ∂W or

through F. Let $\pi_-(p)$ be the first point of ∂W or F that p meets under this flow. This defines a map $\pi_-\colon W \to (\partial W \cup F)$, not necessarily continuous. Then W decomposes into two parts, $W_\partial = \pi_-^{-1}(\partial W)$ and $W_F = \pi_-^{-1}(F)$.

For any $p \in \partial W$, the fiber $\pi_-^{-1}(p)$ is a segment extending upward from p, and integrating along these fibers, we find that $\mathrm{RVol}(W_\partial) \leqslant \frac{1}{\ln(\lambda\mu)} \mathrm{Area}(\partial W)$. For $\mathrm{RVol}(W_F)$, the fiber of any point in F consists of *two* segments extending vertically, so $\mathrm{RVol}(W_F) \leqslant \frac{2}{\ln(\lambda\mu)} \mathrm{Area}(F)$. □

It now suffices to bound $\mathrm{Area}(F)$ from above in terms of $\mathrm{Area}(\partial W)$.

4.2 We need to make some definitions. Let $L = \log_\lambda(\mathrm{Area}(\partial W))$. We have the following properties.

$$\lambda^L = \mathrm{Area}(\partial W) \tag{2}$$

$$\mu^L = \mathrm{Area}(\partial W)^{\log_\lambda(\mu)} \tag{3}$$

$$(\lambda\mu)^L = \mathrm{Area}(\partial W)^{1+\log_\lambda(\mu)} \tag{4}$$

Equation (2) holds by definition, (4) follows from (2) and (3), and (3) is an instance of the identity $a^{\log_b(c)} = c^{\log_b(a)}$.

Let $v_1, \ldots, v_k \in V(T)$ be the vertices in the image of W under the projection $\pi_T\colon \widetilde{X} \to T$. We define several items associated to these vertices:

- $h_i = h_0(v_i)$, the *height* of v_i
- $F_i = \pi_T^{-1}(v_i) \cap F$, the *fold set at* v_i
- $T_i = \{x \in T \mid v_i \in [x, \eta)\}$, the *subtree above* v_i

and the following subsets of ∂W:

- $S_i = \partial W \cap \pi_T^{-1}(T_i)$, the *surface above* v_i
- $A_i = S_i \cap h^{-1}((h_i, h_i + 1))$, the *low slice* of S_i
- $B_i = S_i \cap h^{-1}((h_i + L, h_i + L + 1))$, the *high slice* of S_i.

Note that ∂S_i has height h_i, so A_i lies between heights 0 and 1 above ∂S_i, and B_i lies between heights L and $L+1$ above ∂S_i.

Lemma 4.3 *$A_i \cap A_j = B_i \cap B_j = \emptyset$ for $i \neq j$.*

Proof Consider the case of A_i and A_j first. If $h_i \neq h_j$ then $h(A_i) \cap h(A_j) = \emptyset$ since vertices have integer heights and the sets $h(A_i)$ have the form $(h_i, h_i + 1)$. If $h_i = h_j$ then $v_i \notin T_j$ and $v_j \notin T_i$, which implies that $T_i \cap T_j = \emptyset$, and hence A_i and A_j are disjoint. The case of B_i and B_j is similar. □

Recall that for each $p \in F$, the fiber $\pi_-^{-1}(p)$ is a pair of segments extending upward from p (it is an open subtree of $p_0 \times T \subset \mathbb{R}^2 \times T$, with no branching, since W is a manifold). Define a (non-continuous) map $\pi_+ \colon F \to \partial W$ by choosing $\pi_+(p)$ to be one of the two upper endpoints of the fiber $\pi_-^{-1}(p)$ for each $p \in F$. Note that π_+ is injective (since $\pi_- \circ \pi_+ = \mathrm{id}_F$), and $\pi_+(F_i) \subset S_i$. The choices of endpoints can be made so that π_+ is measurable.

We now express each fold set F_i as a union of two parts, the *low* and *high* parts, as follows:

$$\begin{aligned}(F_i)_{low} &= \{p \in F_i \mid h(\pi_+(p)) \leqslant h_i + L + 1\}\\ (F_i)_{high} &= \{p \in F_i \mid h(\pi_+(p)) \geqslant h_i + L + 1\}.\end{aligned}$$

Also define $F_{low} = \bigcup_i (F_i)_{low}$ and $F_{high} = \bigcup_i (F_i)_{high}$. Clearly, $F = F_{low} \cup F_{high}$.

Proposition 4.4 $\mathrm{Area}(F_{low}) \leqslant (\lambda\mu)\,\mathrm{Area}(\partial W)^{2+\log_\lambda(\mu)}$.

Proof We compare the areas of F_{low} and its image under π_+, which is a subset of ∂W. Since π_+ projects points of F_{low} upward a distance of at most $L+1$, the horizontal area element at $p \in F_{low}$ is at most $(\lambda\mu)^{L+1}$ times the horizontal area element at $\pi_+(p)$. Recall also that this latter area element is no larger than the surface area element of ∂W at $\pi_+(p)$. Since π_+ is injective, we now have $\mathrm{Area}(F_{low}) \leqslant (\lambda\mu)^{L+1}\,\mathrm{Area}(\pi_+(F_{low}))$. The proposition follows, by Equation 4 and the fact that $\mathrm{Area}(\pi_+(F_{low})) \leqslant \mathrm{Area}(\partial W)$. □

4.5 We need to introduce some further terminology. Recall that the map $q \colon \widetilde{X} \to M$ is the identity on the $\mathbb{R}^2$ factors of $\widetilde{X}$ and M. Thus the $\mathbb{R}^2$ factor of $\widetilde{X}$ has coordinates x, y coming from M. Let π_x, $\pi_y \colon \widetilde{X} = \mathbb{R}^2 \times T \to \mathbb{R}^2$ be the projection maps onto the x– and y–axes: $\pi_x(x,y,t) = (x,0)$ and $\pi_y(x,y,t) = (0,y)$.

Given $t \in T$ and a subset $S \subset \mathbb{R}^2 \times t$, let $\ell_x(S)$ be the length of $\pi_x(S) \times h_0(t)$ considered as a subset of M. This subset is contained in a line parallel to the x–axis, and its length in M will depend on the height of t. Similarly, let $\ell_y(S)$ be the length of $\pi_y(S) \times h_0(t)$. Since the metric on $\mathbb{R}^2 \times t$ is Euclidean, we have

$$\mathrm{Area}(S) \leqslant \ell_x(S)\,\ell_y(S). \tag{5}$$

Now consider two additional projection maps in M: the map $\Pi_x \colon M \to M$ given by $(x,y,z) \mapsto (x,0,z)$, and $\Pi_y \colon M \to M$ given by $(x,y,z) \mapsto (0,y,z)$. If we consider the image coordinate planes in their induced metrics, both of these maps are area-decreasing for surfaces in M.

We wish to estimate the area of $(F_i)_{high}$ using Equation 5. For this, we will relate $\ell_x((F_i)_{high})$ and $\ell_y((F_i)_{high})$ to the areas of A_i and B_i. Consider two more families of sets in $M = \mathbb{R}^2 \times \mathbb{R}$:

$$\begin{aligned} Q_i &= \pi_x((F_i)_{high}) \times (h_i, h_i + 1) \\ R_i &= \pi_y((F_i)_{high}) \times (h_i + L, h_i + L + 1). \end{aligned}$$

These sets are contained in the $xz-$ and $yz-$coordinate planes respectively, and their areas may be measured in the induced (hyperbolic) metrics.

Lemma 4.6 *For each i we have*

(1) $\ell_x((F_i)_{high}) \leqslant \lambda \operatorname{Area}(Q_i)$

(2) $\ell_y((F_i)_{high}) \leqslant \mu^L \operatorname{Area}(R_i)$.

Proof For (1), the induced metric on the $xz-$coordinate plane is given by $ds^2 = \lambda^{-2z}dx^2 + dz^2$, with area element $\lambda^{-z}dx\,dz$. Let $D_i \subset \mathbb{R}$ be the projection $\{x \in \mathbb{R} \mid (x,0) \in \pi_x((F_i)_{high})\}$. We have

$$\begin{aligned} \operatorname{Area}(Q_i) &= \int_{D_i}\int_{h_i}^{h_i+1} \lambda^{-z}dz\,dx \geqslant \int_{D_i}\int_{h_i}^{h_i+1} \lambda^{-h_i-1}dz\,dx \\ &= \lambda^{-1}\int_{D_i} \lambda^{-h_i}dx = \lambda^{-1}\ell_x((F_i)_{high}). \end{aligned}$$

The inequality holds since $\lambda > 1$, and the last equality holds since F_i has height h_i.

Part (2) is similar. The $yz-$plane has metric given by $ds^2 = \mu^{-2z}dy^2 + dz^2$ with area element $\mu^{-z}dy\,dz$. Let $E_i \subset \mathbb{R}$ be the projection $\{y \in \mathbb{R} \mid (0,y) \in \pi_y((F_i)_{high})\}$. Then

$$\begin{aligned} \operatorname{Area}(R_i) &= \int_{E_i}\int_{h_i+L}^{h_i+L+1} \mu^{-z}dz\,dy \geqslant \int_{E_i}\int_{h_i+L}^{h_i+L+1} \mu^{-h_i-L}dz\,dy \\ &= \mu^{-L}\int_{E_i} \mu^{-h_i}dy = \mu^{-L}\ell_y((F_i)_{high}). \end{aligned}$$

This time, the inequality holds because $\mu < 1$. □

Proposition 4.7 $\operatorname{Area}(F_{high}) \leqslant \lambda \operatorname{Area}(\partial W)^{2+\log_\lambda(\mu)}$.

Proof We will show that

$$\textbf{(6)} \qquad \operatorname{Area}((F_i)_{high}) \leqslant \lambda\mu^L \operatorname{Area}(A_i)\operatorname{Area}(B_i)$$

for all i. Then, summing over i and applying Lemma 4.3, we obtain

$$\operatorname{Area}(F_{high}) \leqslant \lambda\mu^L \operatorname{Area}(\partial W)^2$$

which implies the proposition by Equation 3.

To establish Equation 6 it suffices to show that $\mathrm{Area}(Q_i) \leqslant \mathrm{Area}(A_i)$ and $\mathrm{Area}(R_i) \leqslant \mathrm{Area}(B_i)$ and to apply Equation 5 and Lemma 4.6.

First we claim that $\Pi_y(q(B_i))$ contains R_i. Choose any $p \in (F_i)_{high}$ and $h \in (h_i + L, h_i + L + 1)$. Write p as $(p_0, t_0) \in \mathbb{R}^2 \times T$ and $\pi_+(p)$ as (p_0, t_1). The segment $p_0 \times [t_0, t_1]$ is part of the fiber $\pi_-^{-1}(p)$, and is contained in W. Since p is in the high part of F_i, the height of t_1 is at least $h_i + L + 1$, and there is a unique $t \in [t_0, t_1]$ of height h. Now we have $(p_0, t) \in W$. The line through (p_0, t) parallel to the x–axis must exit W, at some point $b \in B_i$. Now $\Pi_y(q(b)) = (\pi_y(b), h) = (\pi_y(p), h)$, and we have shown that $R_i \subset \Pi_y(q(B_i))$.

By a similar argument, $\Pi_x(q(A_i))$ contains Q_i (reverse the roles of x and y and choose $h \in (h_i, h_i + 1)$). Now recall that Π_x and Π_y are area-decreasing and q is locally isometric. It follows that $\mathrm{Area}(B_i) \geqslant \mathrm{Area}(R_i)$ and $\mathrm{Area}(A_i) \geqslant \mathrm{Area}(Q_i)$, as needed. □

Finally, putting together Propositions 4.1, 4.4, and 4.7, and consolidating constants (with the assumption that $\mathrm{Area}(\partial W) \geqslant 1$), we obtain

$$\mathrm{RVol}(W) \leqslant \left(\frac{2\lambda(\mu+1)+1}{\ln(\lambda\mu)}\right) \mathrm{Area}(\partial W)^{2+\log_\lambda(\mu)} \tag{7}$$

which has the form of the desired upper bound for $\Delta^{(2)}(x)$.

The general case

Now we return to the situation given at the beginning of this section, where $g\colon W \to \widetilde{X}$ is a least-volume transverse extension of $f\colon \partial W \to \widetilde{X}^{(2)}$. The proof will follow the same general outline as in the embedded case, and we will work with analogues of the various items F_i, A_i, B_i, Q_i, R_i, etc. The proof itself does not depend formally on the embedded case, though we will use several of the intermediate results obtained thus far.

4.8 We need to introduce some terminology related to the generalized handle decomposition of W. Recall that a 2–cell of $\widetilde{X}$ is either horizontal or vertical, accordingly as it maps to a vertex or an edge of the tree T.

A 1–handle is *horizontal* if it maps to a horizontal 2–cell of $\widetilde{X}$ and is not a floating 1–handle (i.e. it is homeomorphic to $I \times D^2$, and not to $S^1 \times D^2$). A 1–handle is *vertical* if it maps to a vertical 2–cell of $\widetilde{X}$ and is not a floating 1–handle. Thus, every 1–handle is either horizontal, vertical, or is a floating handle.

Remark 4.9 Every non-floating 1–handle either joins a 0–handle to a 0–handle, a 0–handle to ∂W, or ∂W to ∂W. In the first case, since the map g is least-volume, the two 0–handles map to *distinct* 3–cells of $\widetilde{X}$. For otherwise, the two neighboring 0–handles can be cancelled by the procedure described in Section 2, reducing the volume of g. No 1–handle joins a 0–handle to itself, since $\widetilde{X}$ has the property that no 2–cell appears more than once as a "face" of any single 3–cell; the closure of a 3–cell in $\widetilde{X}$ is an embedded ball with interior equal to the open 3–cell.

4.10 We will need to make use of some vector fields on W, obtained by pulling back the coordinate vector fields on M via the map $q \circ g\colon\ W \to M$. These vector fields will be denoted $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$, and $\frac{\partial}{\partial z}$, and they are defined on the interiors of the 0–handles. In particular, every 0–handle has an "upward" direction given by $\frac{\partial}{\partial z}$.

We say that a horizontal 1–handle H is *minimal* if $\frac{\partial}{\partial z}$ is directed *away* from H in both neighboring 0–handles. Such a 1–handle is a local minimum for the height function (the z–coordinate) on the tree T.

Since T branches only in the upward direction, and since horizontal 1–handles are joined to 0–handles mapping to distinct 3–cells in $\widetilde{X}$, there are no "maximal" 1–handles H (where $\frac{\partial}{\partial z}$ is directed toward H on both ends). Hence if a horizontal handle $H = I \times D^2$ is not minimal, then $\frac{\partial}{\partial z}$ on the neighboring 0–handles can be extended to a non-vanishing vector field on H, tangent to the I factor. Thus we will always regard $\frac{\partial}{\partial z}$ as being defined (and non-zero) on the union of the 0–handles and the non-minimal horizontal 1–handles.

Let $\mathscr{F}_z$ be the partial foliation on W whose leaves are the orbits of the flow along $\frac{\partial}{\partial z}$. Some leaves of $\mathscr{F}_z$ may terminate or originate in a 2– or 3–handle of W. These are the leaves whose images in $\widetilde{X}$ meet a 0– or 1–cell. In terms of transverse area, the set of such leaves has measure zero, and we will discard them from $\mathscr{F}_z$. Note that the remaining leaves of $\mathscr{F}_z$ still meet the 0–handles in a set of full measure. Let U_z denote the union of the leaves of $\mathscr{F}_z$.

Every vertical 2–cell of $\widetilde{X}$ is a face of exactly two 3–cells, and also is not tangent to the vector fields $\frac{\partial}{\partial x}$ or $\frac{\partial}{\partial y}$. (The sides of Q are not parallel to the x– or y–axes because the matrix A is hyperbolic.) These facts, together with Remark 4.9, imply that for any vertical 1–handle $H = I \times D^2$, the vector field $\frac{\partial}{\partial x}$ on the neighboring 0–handles extends to a non-vanishing vector field on H, tangent to the I factor. By adjusting lengths, we can arrange that this field is independent of the z–coordinate (this is already true in the 0–handles). The vector field $\frac{\partial}{\partial y}$ is defined similarly. We also define partial foliations $\mathscr{F}_x$ and $\mathscr{F}_y$ on the union of the 0–handles and vertical

1–handles, analogously to $\mathscr{F}_z$. Note that these two foliations coincide in the vertical 1–handles, even though they are transverse elsewhere. Again, we will discard all leaves terminating or originating in a 2– or 3–handle of W. Let U_x and U_y denote, respectively, the unions of the leaves of $\mathscr{F}_x$ and of $\mathscr{F}_y$.

4.11 Every leaf of $\mathscr{F}_z$ is homeomorphic to $\mathbb{R}$ and is oriented by the vector field $\frac{\partial}{\partial z}$. It terminates in a well-defined point of ∂W, and originates either at a point in ∂W or at a point in the boundary of a minimal 1–handle. Similarly, every leaf of $\mathscr{F}_x$ and $\mathscr{F}_y$ both originates and terminates on ∂W. For $p \in U_\alpha$ let $\tau_\alpha(p)$ denote the terminal point of the leaf of $\mathscr{F}_\alpha$ containing p (for $\alpha = x, y, z$). This defines maps $\tau_\alpha \colon U_\alpha \to \partial W$. Also let $o_\alpha(p)$ be the origination point of the leaf of $\mathscr{F}_\alpha$ containing p.

Definition 4.12 We wish to define the *fold sets* in W, which will be embedded surfaces with boundary (minus a measure zero set). Let $e_1, \dots, e_k$ be the closed edges of T which meet the image of $\pi_T \circ g$. Given e_i and a point p_i in the interior of e_i, the preimage $(\pi_T \circ g)^{-1}(p_i)$ is a properly embedded surface $\Sigma_i \subset W$, by transversality, and the preimage of the interior of e_i is an open regular neighborhood of Σ_i. The intersection of Σ_i with the handle decomoposition of W is a handle decomposition of Σ_i, and the map is transverse with respect to this structure. The closure of the preimage of the interior of e_i is a union of handles of W, and is a codimension-zero submanifold of W, homeomorphic to $\Sigma_i \times I$, with the product handle structure. That is, each 0–, 1–, or 2–handle of $\Sigma_i \times I$ is the product of a 0–, 1–, or 2–handle of Σ_i with I. The product structure $\Sigma_i \times I$ is chosen so that fibers $p \times I$ map by $q \circ g$ into vertical lines in M (in particular, I corresponds to the z–coordinate in the 0–handles).

Let v_i be the lower endpoint of e_i (with respect to the height function), and orient the I factor of $\Sigma_i \times I$ so that $\Sigma_i \times 0$ maps to v_i. The handles of W comprising $\Sigma_i \times I$ are all 0–, 1–, and 2–handles. Various 1–, 2–, and 3–handles (those mapping to v_i by $\pi_T \circ g$) may be attached in part to $\Sigma_i \times 0$. Let E_i be the intersection of $\Sigma_i \times 0$ with the union of all minimal 1–handles. It is a codimension-zero submanifold of $\Sigma_i \times 0$, equal to a union of attaching regions of minimal 1–handles. Every minimal 1–handle is attached to two surfaces E_i, E_j for some $i \neq j$, since the adjacent 0–handles are distinct and map to distinct edges of T. Lastly, define F_i to be $E_i \cap U_z$. Note that F_i has full measure in E_i.

Having defined F_i and v_i, note that various vertices v_i may now coincide (unlike the embedded case). Define the heights h_i exactly as before: $h_i = h_0(v_i)$. Define

$L = \log_\lambda(\mathrm{RArea}(f))$, and note that equations analogous to (2)–(4) hold.

$$\lambda^L = \mathrm{RArea}(f) \tag{8}$$

$$\mu^L = \mathrm{RArea}(f)^{\log_\lambda(\mu)} \tag{9}$$

$$(\lambda\mu)^L = \mathrm{RArea}(f)^{1+\log_\lambda(\mu)} \tag{10}$$

We redefine the subtrees T_i to be smaller than those from Section 4.2, by splitting along the edges above the vertex. That is, we now define

$$T_i = \{x \in T \mid \mathrm{int}(e_i) \cap [x, \eta) \neq \emptyset\}.$$

This is an *open* subtree of T, not containing v_i. Define S_i, A_i, and B_i as follows:

- $S_i = \partial W \cap \mathrm{closure}((g \circ \pi_T)^{-1}(T_i))$
- $A_i = S_i \cap (g \circ h)^{-1}((h_i, h_i + 1))$
- $B_i = S_i \cap (g \circ h)^{-1}((h_i + L, h_i + L + 1))$.

Note that S_i is a subsurface of ∂W and $\partial S_i = \partial W \cap (\Sigma_i \times 0)$. The next lemma has essentially the same proof as Lemma 4.3.

Lemma 4.13 *$A_i \cap A_j = B_i \cap B_j = \emptyset$ for $i \neq j$.* □

Now let $F = \bigcup_i F_i$, and define $\pi_+ \colon F \to \partial W$ to be the restriction $\tau_z|_F$. That is, π_+ flows F "upward" along $\frac{\partial}{\partial z}$ to ∂W. Note that π_+ is indeed defined on F, and is injective. Define the *low* and *high* parts of F as before:

$$(F_i)_{low} = \{p \in F_i \mid h(g(\pi_+(p))) \leqslant h_i + L + 1\}$$
$$(F_i)_{high} = \{p \in F_i \mid h(g(\pi_+(p))) \geqslant h_i + L + 1\}.$$

Also define $F_{low} = \bigcup_i (F_i)_{low}$ and $F_{high} = \bigcup_i (F_i)_{high}$.

Lemma 4.14 $\mathrm{RVol}(g) \leqslant \frac{1}{\ln(\lambda\mu)}(\mathrm{RArea}(f) + \mathrm{RArea}(g|_F))$.

Proof We have $\mathrm{RVol}(g) = \mathrm{RVol}(g|_{U_z})$ since U_z has full measure in the 0–handles of W. Note that every leaf of $\mathscr{F}_z$ starts on F or on ∂W, and ends in ∂W. Thus we may decompose U_z as $U_z^F \cup U_z^\partial$ where

$$U_z^F = \{p \in U_z \mid o_z(p) \in F\},$$
$$U_z^\partial = \{p \in U_z \mid o_z(p) \in \partial W\}.$$

Now $\mathrm{RVol}(g|_{U_z}) = \mathrm{RVol}(g|_{U_z^F}) + \mathrm{RVol}(g|_{U_z^\partial})$. By pulling back the metric from $\widetilde{X}$ and integrating along leaves of $\mathscr{F}_z$, we have

$$\mathrm{RVol}(g|_{U_z^F}) \leqslant \frac{1}{\ln(\lambda\mu)} \mathrm{RArea}(g|_F)$$

and

$$\mathrm{RVol}(g|_{U_z^\partial}) \leqslant \frac{1}{\ln(\lambda\mu)} \mathrm{RArea}(g|_{\partial W}) = \frac{1}{\ln(\lambda\mu)} \mathrm{RArea}(f). \qquad \square$$

Remark 4.15 In the current situation, there is no ambiguity or choice involved in the definition of π_+. The difference with the embedded case is that each minimal 1–handle has *two* attaching regions contributing to F, and there is a unique way to flow upward from each side. In effect, the fold set has been doubled, and this also accounts for the missing factor of 2 in Lemma 4.14 (compared with Proposition 4.1).

Our main task now is to bound $\mathrm{RArea}(g|_F)$ in terms of $\mathrm{RArea}(f)$. The next result is entirely analogous to Proposition 4.4, and has the same proof. The only difference is that here the area elements are pulled back from $\widetilde{X}$.

Proposition 4.16 $\mathrm{RArea}(g|_{F_{low}}) \leqslant (\lambda\mu)\,\mathrm{RArea}(f)^{2+\log_\lambda(\mu)}$. $\square$

Next we need an analogue of Equation 5. In order to define the lengths ℓ_x and ℓ_y for the sets $(F_i)_{high}$, we need to extend the vector fields $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ to the surfaces $\Sigma_i \times 0$. Recall that $\Sigma_i \times I$ has a product handle structure, and these vector fields are defined in the interiors of the 0–handles and 1–handles (all of which are vertical). Note that $\frac{\partial}{\partial x}$, in the interior of $\Sigma_i \times I$, is zero in the I factor and constant (as $t \in I$ is varied) in the Σ_i factor. Thus $\frac{\partial}{\partial x}$ extends continuously to $\Sigma_i \times 0$ as a non-vanishing field, defined on the interiors of the 0– and 1–handles of $\Sigma_i \times 0$. Any leaf of $\mathscr{F}_x$ meeting $\Sigma_i \times 0$ remains entirely within $\Sigma_i \times 0$, since $\frac{\partial}{\partial x}$ is tangent to this surface (indeed, every $\Sigma_i \times t$ has this property). The vector field $\frac{\partial}{\partial y}$ extends to $\Sigma_i \times 0$ in the same way. Lastly, we discard leaves of $\mathscr{F}_x$ and $\mathscr{F}_y$ meeting 2–handles of $\Sigma_i \times 0$, so that every leaf in $\Sigma_i \times 0$ begins and ends in ∂S_i. These remaining leaves have full measure in the 0–handles of $\Sigma_i \times 0$.

We now define $\ell_x((F_i)_{high})$ to be the transverse measure of the set of leaves of $\mathscr{F}_y$ meeting $(F_i)_{high}$. That is, we project $(F_i)_{high} \cap U_y$ to ∂S_i using τ_y, and then measure this set by integrating the pullback of the length element $\lambda^{-z}dx$ from M. Similarly, $\ell_y((F_i)_{high})$ is defined using the length element $\mu^{-z}dy$.

Proposition 4.17 $\mathrm{RArea}(g|_{(F_i)_{high}}) \leqslant \ell_x((F_i)_{high})\,\ell_y((F_i)_{high})$ *for each* i.

Proof First observe that the intersection of a leaf of $\mathscr{F}_x$ and a leaf of $\mathscr{F}_y$ is either one point (in a 0–handle of $\Sigma_i \times 0$), a closed interval (in a 1–handle of $\Sigma_i \times 0$), or is empty. To see this, map both leaves to M and project onto the x–axis. Each $\mathscr{F}_y$ leaf maps to a single point, whereas each $\mathscr{F}_x$ leaf maps monotonically, with point preimages equal to sets of the form described above.

It follows that the map

$$\tau_y \times \tau_x \colon (\Sigma_i \times 0) \cap U_x \cap U_y \;\to\; \partial S_i \times \partial S_i$$

is injective when restricted to the 0–handles of $\Sigma_i \times 0$.

Next define the map $g_i \colon \Sigma_i \times 0 \to \mathbb{R}^2$ to be $q \circ g \colon \Sigma_i \times 0 \to M$ followed by projection onto the first two coordinates of $M = \mathbb{R}^3$. Thus, $q(g(p)) = (g_i(p), h_i) \in M$ for all $p \in \Sigma_i \times 0$. Let $\pi_x, \pi_y \colon \mathbb{R}^2 \to \mathbb{R}$ be projections onto the first and second coordinates respectively. It is easily verified that g_i agrees with the following composition of maps:

$$(\Sigma_i \times 0) \cap U_x \cap U_y \xrightarrow{\tau_y \times \tau_x} \partial S_i \times \partial S_i \xrightarrow{g_i \times g_i} \mathbb{R}^2 \times \mathbb{R}^2 \xrightarrow{\pi_x \times \pi_y} \mathbb{R} \times \mathbb{R}.$$

(Write $q(g(p))$ as (x_p, y_p, h_i); both maps send p to (x_p, y_p).)

Recall that $\Sigma_i \times 0$ maps into $\mathbb{R}^2 \times h_i \subset M$, and so the surface area element being pulled back in the computation of $\mathrm{RArea}(g|_{(F_i)_{high}})$ is the horizontal area element of M. This element is just the product of the length elements $\lambda^{-z}dx$ and $\mu^{-z}dy$.

In the integrals below, $(F_i)_{high}$ is understood to be restricted to the 0–handles of $\Sigma_i \times 0$ (where area is supported). We have

$$\begin{aligned} \mathrm{RArea}(g|_{(F_i)_{high}}) &= \int_{(F_i)_{high}} (q \circ g)^*(\lambda^{-z}dx\, \mu^{-z}dy) \\ &= \int_{(F_i)_{high} \cap U_x \cap U_y} (\pi_x \times \pi_y \circ g_i \times g_i \circ \tau_y \times \tau_x)^*(\lambda^{-z}dx\, \mu^{-z}dy) \end{aligned}$$

which, by injectivity of $\tau_y \times \tau_x$, is at most

$$\int_{\tau_y((F_i)_{high} \cap U_x \cap U_y) \times \tau_x((F_i)_{high} \cap U_x \cap U_y)} (\pi_x \times \pi_y \circ g_i \times g_i)^*(\lambda^{-z}dx\, \mu^{-z}dy).$$

The latter is equal to

$$\int_{\tau_y((F_i)_{high} \cap U_x \cap U_y)} (\pi_x \circ g_i)^*(\lambda^{-z}dx) \int_{\tau_x((F_i)_{high} \cap U_x \cap U_y)} (\pi_y \circ g_i)^*(\mu^{-z}dy),$$

which is just $\ell_x((F_i)_{high})\, \ell_y((F_i)_{high})$. □

In Section 4.5 we defined the projection maps $\Pi_x, \Pi_y \colon M \to M$, sending (x,y,z) to the points $(x,0,z)$ and $(0,y,z)$ respectively. We also had projections $\pi_x, \pi_y \colon \widetilde{X} = \mathbb{R}^2 \times T \to \mathbb{R}^2$, mapping (x,y,t) to $(x,0)$ and $(0,y)$ respectively. Define the sets Q_i, $R_i \subset M = \mathbb{R}^2 \times \mathbb{R}$ as follows:

$$\begin{aligned} Q_i &= \pi_x(g((F_i)_{high})) \times (h_i, h_i+1) \\ R_i &= \pi_y(g((F_i)_{high})) \times (h_i+L, h_i+L+1). \end{aligned}$$

The claims of Lemma 4.6 remain true exactly as stated, and are proved in the same way. Thus:

Lemma 4.18 *For each i we have*

(1) $\ell_x((F_i)_{high}) \leqslant \lambda \operatorname{Area}(Q_i)$

(2) $\ell_y((F_i)_{high}) \leqslant \mu^L \operatorname{Area}(R_i)$. □

Next we adapt Proposition 4.7 to the current situation.

Proposition 4.19 $\operatorname{RArea}(g|_{F_{high}}) \leqslant \lambda \operatorname{RArea}(f)^{2+\log_\lambda(\mu)}$.

Proof As in the proof of Proposition 4.7, it suffices to show that $\operatorname{Area}(Q_i) \leqslant \operatorname{RArea}(f|_{A_i})$ and $\operatorname{Area}(R_i) \leqslant \operatorname{RArea}(f|_{B_i})$ for each i: since

$$\operatorname{RArea}(g|_{(F_i)_{high}}) \leqslant \lambda\mu^L \operatorname{Area}(A_i) \operatorname{Area}(B_i)$$

by Proposition 4.17 and Lemma 4.18, we then have

$$\operatorname{RArea}(g|_{(F_i)_{high}}) \leqslant \lambda\mu^L \operatorname{RArea}(f|_{A_i}) \operatorname{RArea}(f|_{B_i})$$

for all i. Summing over i, using Lemma 4.13, we obtain the desired inequality, by Equation 9.

We claim that $\Pi_y(q(f(B_i)))$ contains a subset of R_i of full measure. Given a point in R_i, it is determined by points $p \in (F_i)_{high}$ and $h \in (h_i+L, h_i+L+1)$. Let $p' \in W$ be a point on the leaf of $\mathscr{F}_z$ through p of height h; such a point exists since p has height h_i and $\pi_+(p)$ has height h_i+L+1 or greater. Write $q(g(p'))$ as $(x_{p'}, y_{p'}, h)$ in the coordinates of M, and note that $q(g(p)) = (x_{p'}, y_{p'}, h_i)$. Thus $\pi_y(g(p)) = (0, y_{p'})$.

If $p' \in U_x$ then $\tau_x(p')$ is defined and is in B_i, and

$$\Pi_y(q(f(\tau_x(p')))) = (0, y_{p'}, h) = (\pi_y(g(p)), h).$$

Therefore this point of R_i is indeed in the image of B_i under $\Pi_y \circ q \circ f$. Thus we want to verify that $p' \in U_x$ for almost all choices of $(\pi_y(g(p)), h) \in R_i$.

Let R_i' be the set of pairs $(\pi_y(g(p)), h) \in R_i$ such that h is not an integer. Let $K \subset \widetilde{X}$ be the intersection of $g(W)$ with the 1–skeleton of $\widetilde{X}$. It is a finite graph, and its image $\Pi_y(q(K))$ has measure zero in the yz–plane in M. Note also that all 2– and 3–handles of W map by g into K.

The point p' must be in the interior of a 0–handle or a horizontal 1–handle of W, since $p' \in U_z$. In the latter case, p' maps to a horizontal 2–cell of $\widetilde{X}$, and so h is an integer. In the former case, $\frac{\partial}{\partial x}$ is defined at p'. If $p' \notin U_x$ then the (discarded) leaf of $\mathscr{F}_x$ through p' meets a 2– or 3–handle. Then $\Pi_y(q(g(p')))$ is contained in the measure zero set $\Pi_y(q(K))$. But $\Pi_y(q(g(p')))$ is the original point $(\pi_y(g(p)), h) \in R_i$. The argument above therefore shows that $\Pi_y(q(f(B_i)))$ contains $R_i' - \Pi_y(q(K))$, a subset of R_i of full measure.

Thus $\operatorname{Area}\Pi_y(q(f(B_i))) \geqslant \operatorname{Area}(R_i)$. Since Π_y is area-decreasing and q locally isometric, we conclude that $\operatorname{RArea}(f|_{B_i}) \geqslant \operatorname{Area}(R_i)$. By a similar argument, $\operatorname{RArea}(f|_{A_i}) \geqslant \operatorname{Area}(Q_i)$. □

The bound

We can now determine an upper bound for $\Delta^{(2)}(x)$. Assembling Lemma 4.14 and Propositions 4.16, 4.19 and consolidating constants, we find that

$$\textbf{(11)} \qquad \operatorname{RVol}(g) \leqslant \left(\frac{1+\lambda(\mu+1)}{\ln(\lambda\mu)}\right) \operatorname{RArea}(f)^{2+\log_\lambda(\mu)}.$$

Recall that all 3–cells of $\widetilde{X}$ have the same volume V (and hence $\operatorname{Vol}^3(g) = \frac{1}{V}\operatorname{RVol}(g)$). Let C be the largest Riemannian area of a 2–cell of $\widetilde{X}$ (or equivalently, of X). Then $\operatorname{RArea}(f) \leqslant C\operatorname{Vol}^2(f)$, and by Equation 11 we have

$$\operatorname{Vol}^3(g) \leqslant \left(\frac{1+\lambda(\mu+1)}{V\ln(\lambda\mu)}\right)(C\operatorname{Vol}^2(f))^{2+\log_\lambda(\mu)}.$$

Therefore $\operatorname{FVol}^W(f) \leqslant D(\operatorname{Vol}^2(f))^{2+\log_\lambda(\mu)}$ for a constant D depending only on the original matrix A (which determined λ, μ, and the geometry of $\widetilde{X}$). Since the 3–manifold W was arbitrary, we have now established that $\Delta^{(2)}(x) \leqslant Dx^{2+\log_\lambda(\mu)}$, and therefore $\delta^{(2)}(x) \preccurlyeq \Delta^{(2)}(x) \preccurlyeq x^{2+\log_\lambda(\mu)}$.

5 The lower bound

To establish a lower bound for $\delta^{(2)}(x)$ we want a sequence of embedded balls $B_n \subset \widetilde{X}$ whose volume growth is as large as possible, relative to the growth of boundary area.

The optimal shape is a ball made from two half-balls, each contained in a copy of M inside $\widetilde{X}$, joined along their bottom faces. The half-balls in M will need to have large volume compared to "upper" boundary area.

For the half-balls, we begin by defining optimally proportioned regions $R_n \subset M$, which are easy to measure in the Riemannian metric. Then we approximate these regions combinatorially by subcomplexes S_n.

Extremal Riemannian regions

In the coordinates of M, define

$$R_n \;=\; [0,\lambda^n]\times[0,(\lambda\mu)^n]\times[0,n].$$

The volume of R_n is easily computed by integration. Each horizontal slice $[0,\lambda^n]\times[0,(\lambda\mu)^n]\times z$ has area $\lambda^n(\lambda\mu)^n(\lambda\mu)^{-z}$, and integrating in the z–coordinate yields

$$\mathrm{RVol}(R_n) \;=\; \frac{1}{\ln(\lambda\mu)}(\lambda^n(\lambda\mu)^n-\lambda^n). \tag{12}$$

Recall that $\lambda\mu=\det(A)\geqslant 2$. If $n\geqslant 1$ then $\frac{1}{2}(\lambda\mu)^n\geqslant 1$, whence $(\lambda\mu)^n-1\geqslant\frac{1}{2}(\lambda\mu)^n$. Together with Equation 12 this implies

$$\begin{aligned}\mathrm{RVol}(R_n) \;&\geqslant\; \frac{1}{2\ln(\lambda\mu)}\lambda^n(\lambda\mu)^n\\ &=\; \frac{1}{2\ln(\lambda\mu)}\left(\lambda^n\right)^{2+\log_\lambda(\mu)}\end{aligned} \tag{13}$$

for $n\geqslant 1$.

Next we consider the areas of the various faces of R_n. The top face has area λ^n (taking $z=n$, above). Next, the segment $[0,\lambda^n]\times y\times z$ has length $\lambda^n\lambda^{-z}$. Integrating with respect to z, we find that the faces $[0,\lambda^n]\times 0\times[0,n]$ and $[0,\lambda^n]\times(\lambda\mu)^n\times[0,n]$ each have area $\frac{1}{\ln(\lambda)}(\lambda^n-1)$. By a similar computation, the other two vertical faces each have area $\frac{1}{\ln(\mu)}\lambda^n(\mu^n-1)=\frac{1}{\ln(\mu^{-1})}\lambda^n(1-\mu^n)$. Since $\mu<1$, this quantity is less than $\frac{1}{\ln(\mu^{-1})}\lambda^n$. Now let ∂^+R_n denote the union of the five faces (omitting the bottom face) of R_n. We have shown that

$$\mathrm{RArea}(\partial^+R_n) \;\leqslant\; \big(1+(2/\ln\lambda)-(2/\ln\mu)\big)\lambda^n. \tag{14}$$

Extremal combinatorial regions

Recall that D is the matrix $BAB^{-1} = \left(\begin{smallmatrix} \lambda & 0 \\ 0 & \mu \end{smallmatrix}\right)$, and Γ is the lattice $B(\mathbb{Z}\times\mathbb{Z})$, preserved by D. Fix any standard copy of M inside $\widetilde{X}$, corresponding to a line $L \subset T$. Then M is a subcomplex of $\widetilde{X}$, and we need to understand its cell structure. Note that M is a union of subcomplexes $\mathbb{R}^2 \times [i-1, i]$ for $i \in \mathbb{Z}$. Consider the subcomplex $\mathbb{R}^2 \times [0,1]$. Possibly after a horizontal translation, the closed 3–cells are the sets $\gamma(Q) \times [0,1]$, for γ in Γ (recall that Q is a fundamental domain for Γ acting on $\mathbb{R}^2$). Figure 2 shows the top and bottom faces of one of these 3–cells, in the case of no translation.

To be more specific, let Γ' be the lattice $D^{-1}(\Gamma)$, and note that Γ' contains Γ as a subgroup of index d. Then the 3–cells of $\mathbb{R}^2 \times [0,1]$ are the sets $\gamma(Q) \times [0,1]$ where γ ranges over a single coset of Γ in Γ'.

Continuing upward, the closed 3–cells of $\mathbb{R}^2 \times [i-1, i]$ are the sets $\gamma(D^{i-1}(Q)) \times [i-1, i]$, where γ ranges over a coset of $D^{i-1}(\Gamma)$ in Γ'. The choice of coset depends on the path in T followed by L from height 0 to height i. (There are d^i such paths, and cosets.) Thus, the various copies of M inside $\widetilde{X}$ have differing cell structures (with respect to the standard coordinates), though at each height they agree up to horizontal translation.

For $i = 1, 2, \ldots$ let $\Lambda_i \subset \mathbb{R}^2$ be the union of the sides of $\gamma(D^{i-1}(Q))$ for γ in the appropriate coset of $D^{i-1}(\Gamma)$ in Γ'. Then $\Lambda_i \times i$ is a subcomplex of M, and in fact, so is $\Lambda_i \times [i-1, i]$. This latter subcomplex is the smallest subcomplex containing the vertical 1– and 2–cells of $\mathbb{R}^2 \times [i-1, i]$.

Definition 5.1 Let w be the diameter of Q (in $\mathbb{R}^2$, with the Euclidean metric). There is a constant k such that every horizontal or vertical line segment of length w intersects Λ_1 in at most k points. We will call k the *backtracking constant* for $\widetilde{X}$.

Lemma 5.2 *Let $W \subset \mathbb{R}^2$ be a region of the form $[a, a+w] \times \mathbb{R}$ or $\mathbb{R} \times [a, a+w]$. Let $\pi\colon W \to \mathbb{R}$ be projection onto the $\mathbb{R}$ factor. Then $W \cap \Lambda_1$ contains a properly embedded line ℓ, and the restricted map $\pi\colon \ell \to \mathbb{R}$ is at most k-to-one.*

Proof The components of $\mathbb{R}^2 - \Lambda_1$ are isometric copies of the interior of Q. For the first statement, note that an open set of diameter w cannot disconnect W, and so $W \cap \Lambda_1$ is connected and contains a line joining the two ends of W. The second statement is clear, since the fibers of π are horizontal or vertical segments of length w. □

Applying the map D^{i-1} (and possibly a translation) to Lemma 5.2 yields the following result. Note that D preserves the horizontal and vertical foliations of $\mathbb{R}^2$ by lines. In particular, D^{i-1} takes fibers of π to fibers.

Lemma 5.3 *Let $W \subset \mathbb{R}^2$ be a region of the form $[a, a+\lambda^{i-1}w] \times \mathbb{R}$ or $\mathbb{R} \times [a, a+\mu^{i-1}w]$. Let $\pi\colon W \to \mathbb{R}$ be projection onto the $\mathbb{R}$ factor. Then $W \cap \Lambda_i$ contains a properly embedded line ℓ, and the restricted map $\pi\colon \ell \to \mathbb{R}$ is at most k-to-one.* □

Now we can proceed to define subcomplexes approximating the regions R_n. Given an integer n, we will define "slabs" $S_{i,n} \subset \mathbb{R}^2 \times [i-1, i]$ for i between 1 and n. The union $\bigcup_i S_{i,n}$ will contain R_n, and will have comparable volume and surface area (the latter of which is controlled by the backtracking constant k). The slabs will not fit together perfectly: there will be under- and over-hanging portions, but the additional surface area arising in this way is not excessive.

Fix $n \in \mathbb{Z}_+$. For i between 1 and n, consider the four strips

$$\begin{aligned} W_i^1 &= \mathbb{R} \times [-\mu^{i-1}w, 0] \\ W_i^2 &= [\lambda^n, \lambda^n + \lambda^{i-1}w] \times \mathbb{R} \\ W_i^3 &= \mathbb{R} \times [(\lambda\mu)^n, (\lambda\mu)^n + \mu^{i-1}w] \\ W_i^4 &= [-\lambda^{i-1}w, 0] \times \mathbb{R} \end{aligned}$$

which surround the rectangle $[0, \lambda^n] \times [0, (\lambda\mu)^n]$. By Lemma 5.3, each of these strips contains a properly embedded line in Λ_i, projecting to the x– or y–axis in a k-to-one fashion, at most. Choose segments $\ell_i^j \subset W_i^j$ in these lines which meet each other only in their endpoints, forming an embedded quadrilateral in Λ_i enclosing $[0, \lambda^n] \times [0, (\lambda\mu)^n]$. Let D_i be the closed region bounded by this quadrilateral, and define the *slab* $S_{i,n}$ to be the subcomplex $D_i \times [i-1, i] \subset M$. Let $S_n = \bigcup_{i=1}^n S_{i,n}$.

Let $W_{i,n}$ be the rectangle delimited by the outermost sides of the strips W_i^1, W_i^2, W_i^3, W_i^4 and note that $W_{i,n}$ contains D_i. The maximum width of these rectangles is $\lambda^n + 2\lambda^{n-1}w = \lambda^n(1+2w/\lambda)$, and the maximum height is $(\lambda\mu)^n + 2w \leqslant (\lambda\mu)^n(1+2w)$. Let κ be the larger of $\log_\lambda(1 + 2w/\lambda)$ and $\log_{\lambda\mu}(1 + 2w)$. Then the rectangle with lower-left corner at $(-\lambda^{n-1}w, -w)$, of width $\lambda^{n+\kappa}$ and height $(\lambda\mu)^{n+\kappa}$, contains $W_{i,n}$ for all i. Let $R'_{n+\kappa}$ be $R_{n+\kappa}$, translated by $-\lambda^{n-1}w$ in the x–direction and by $-w$ in the y–direction. Then we have

$$R_n \subset S_n \subset R'_{n+\kappa}.$$

Let $\partial^+ S_n$ denote the largest subcomplex of the boundary of S_n which does not meet the interior of the base of R_n (that is, $(0, \lambda^n) \times (0, (\lambda\mu)^n) \times 0$). Note that $\partial^+ S_n$ has three

parts: the *top*, D_n; the *vertical part*, made of the sets $\ell_i^j \times [i-1,i]$; and the *horizontal part*, contained in the union of the annuli $\big(W_{i,n} \times i\big) - \big((0,\lambda^n) \times (0,(\lambda\mu)^n) \times i\big)$, for $i = 0,\dots,n-1$. This last part contains the horizontal 2–cells of height i in the symmetric difference $(D_i \times i) \triangle (D_{i-1} \times i)$, where the slabs fail to join perfectly.

Lemma 5.4 *There is a constant C such that the Riemannian area of the top and vertical parts of $\partial^+ S_n$ is at most $C\,\mathrm{RArea}(\partial^+ R'_{n+\kappa})$.*

Proof Translating D_n upward by κ, it becomes a subset of the top face of $R'_{n+\kappa}$. Therefore its area is at most $(\lambda\mu)^\kappa$ times the area of the top face of $R'_{n+\kappa}$. Next consider the coordinate projections of $\ell_i^j \times [i-1,i]$ onto the sides of $R'_{n+\kappa}$. These maps are at most k-to-one, by the construction of ℓ_i^j. Moreover, the Jacobians of these maps are bounded below by some $J > 0$, independent of n. To see this, consider for example the coordinate projection onto the xz–plane (the case of odd j). On each closed vertical 2–cell the Jacobian achieves a positive minimum, and there are finitely many such cells modulo isometries of M. These isometries preserve the xz–plane field, and hence also the Jacobian of this projection. The case of the yz–projection is similar. Now the Riemannian area of $\bigcup_{i=1}^n \ell_i^j \times [i-1,i]$ is at most k/J times the area of one of the four sides of $R'_{n+\kappa}$ (one side for each j). The result follows with $C = \max\{(\lambda\mu)^\kappa, k/J\}$. □

Lemma 5.5 *There is a constant D such that the Riemannian area of the horizontal part of $\partial^+ S_n$ is at most $D\lambda^n$.*

Proof Let $A_{i,n}$ be the annular region $\big(W_{i,n} \times i\big) - \big((0,\lambda^n) \times (0,(\lambda\mu)^n) \times i\big)$. Then

$$\begin{aligned}\mathrm{RArea}(A_{i,n}) &= (\lambda^{n-i} + 2w/\lambda)(\lambda^n\mu^{n-i} + 2w/\mu) - \lambda^{n-i}\lambda^n\mu^{n-i}\\ &= 2w\lambda^{n-1}\mu^{n-i} + 2w\lambda^{n-i}\mu^{-1} + 4w^2(\lambda\mu)^{-1}\\ &\leqslant 2w(\lambda^{n-1} + \lambda^{n-i}\mu^{-1}) + 4w^2.\end{aligned}$$

Hence the area of the horizontal part is at most

$$\begin{aligned}\sum_{i=0}^{n-1} \mathrm{RArea}(A_{i,n}) &\leqslant 2w\big(\lambda^{n-1} + \lambda(\lambda^n - 1)/\mu(\lambda - 1)\big) + 4w^2 n\\ &\leqslant 2w\big(\lambda^{-1} + \lambda/\mu(\lambda - 1)\big)\lambda^n + 4w^2 n.\end{aligned}$$

Lastly, $4w^2 n$ is less than $\frac{4w^2}{\ln\lambda}\lambda^n$, thus establishing the result. □

The bound

Recall that $\widetilde{X}$ contains isometric copies of M, corresponding to lines in T. Choose two such lines L_0, L_1 which coincide at negative heights and diverge at height 0. Let M_0, M_1 be the corresponding copies of M in $\widetilde{X}$. Let S_n^i be the subcomplex S_n of M_i constructed earlier (recall that the contruction depended on the cell structure of M_i, which varies with i). Let $B_n \subset \widetilde{X}$ be the subcomplex $S_n^0 \cup S_n^1$. It contains the two copies of R_n in M_0 and M_1 (which meet along their bottom faces), and its boundary is contained in $\partial^+ S_n^0 \cup \partial^+ S_n^1$.

Let a be the minimum Riemannian area of a 2–cell of $\widetilde{X}$. Combining Equation 14 with Lemmas 5.4 and 5.5, we have

$$\textbf{(15)} \qquad \mathrm{Vol}^2(\partial B_n) \leqslant (2/a)\Big(C\lambda^\kappa\big(1+(2/\ln\lambda)-(2/\ln\mu)\big)+D\Big)\lambda^n.$$

By Equation 13 we have

$$\mathrm{Vol}^3(B_n) \geqslant \frac{1}{V\ln(\lambda\mu)}\big(\lambda^n\big)^{2+\log_\lambda(\mu)}.$$

Thus there is a constant E such that $\mathrm{Vol}^3(B_n) \geqslant E(\mathrm{Vol}^2(\partial B_n))^{2+\log_\lambda(\mu)}$ for all n. By Remark 2.9, since S_n is embedded in $\widetilde{X}$, we have $\delta^{(2)}(x_n) \geqslant E(x_n)^{2+\log_\lambda(\mu)}$ for $x_n = \mathrm{Vol}^2(\partial B_n)$. Lastly, it remains to show that the sequence (x_n) is not too sparse. Recall that the top D_n of $\partial^+ S_n$ contains the top face of R_n, and the latter has area λ^n. Thus $\mathrm{Vol}^2(\partial B_n) \geqslant K\lambda^n$ for some constant K. Together with Equation 15 this implies that the ratios x_n/x_{n-1} are bounded. According to Remark 2.1 of [5], this property suffices to conclude that $\delta^{(2)}(x) \succcurlyeq x^{2+\log_\lambda(\mu)}$.

6 Proof of Theorem 1.2

Sections 4 and 5 established the proof of Theorem 1.1. Next we consider the groups $G_{\Sigma^i A} \cong G_A \times \mathbb{Z}^i$ and their $(i+2)$–dimensional Dehn functions. The following definition is taken from [5].

Definition 6.1 Let G be a group of type $\mathcal{F}_{k+1}$ and geometric dimension at most $k+1$. The k–dimensional Dehn function $\delta_G^{(k)}(x)$ *has embedded representatives* if there is a finite aspherical $(k+1)$–complex X, a sequence of embedded $(k+1)$–dimensional balls $B_i \subset \widetilde{X}$, and a function $F(x) \simeq \delta_G^{(k)}(x)$, such that the sequence given by $(n_i) = (\mathrm{Vol}^k(\partial B_i))$ tends to infinity and is exponentially bounded, and $\mathrm{Vol}^{k+1}(B_i) \geqslant F(n_i)$ for each i.

The Dehn functions $\delta^{(2)}(x)$ for the groups G_A have embedded representatives, as constructed in Section 5. We also have the following result from [5].

Proposition 6.2 *Let G be a group of type $\mathcal{F}_{k+1}$ and geometric dimension at most $k+1$. Suppose the k–dimensional Dehn function $\delta^{(k)}(x)$ of G is equivalent to x^s and has embedded representatives. Then $G \times \mathbb{Z}$ has $(k+1)$–dimensional Dehn function $\delta^{(k+1)}(x) \succcurlyeq x^{2-1/s}$, with embedded representatives.*

The proof of Theorem 1.2 now proceeds exactly as in Theorem D of [5]. Let $\alpha = 2 + \log_\lambda(\mu)$ and $s(i) = \frac{(i+1)\alpha - i}{i\alpha - (i-1)}$. We verify by induction on i the following statements for $G_{\Sigma^i A}$.

(1) $\Delta^{(i+2)}(x) \leqslant Cx^{s(i)}$ for some constant $C > 0$

(2) $\delta^{(i+2)}(x) \succcurlyeq x^{s(i)}$

(3) $\delta^{(i+2)}(x)$ has embedded representatives

The first two statements together yield the desired conclusion $\delta^{(i+2)}(x) \simeq x^{s(i)}$.

If $i = 0$ then (1) and (2) are the respective conclusions of Sections 4 and 5, and (3) holds as remarked above. For $i > 0$ note first that $s(i) = 2 - 1/s(i-1)$. Then statement (1) holds by Theorem 2.7 and property (1) of $G_{\Sigma^{i-1}A}$. Proposition 6.2 implies (2) and (3) by properties (1)–(3) of $G_{\Sigma^{i-1}A}$.

7 Density of exponents

In this section, A is a 2×2 matrix with integer entries. Denote the trace and determinant of A by t and d respectively. Note that the characteristic polynomial of A is given by $p(x) = x^2 - tx + d$, and the eigenvalues are $\lambda = \frac{t+\sqrt{t^2-4d}}{2}$ and $\mu = \frac{t-\sqrt{t^2-4d}}{2}$. The next lemma shows that under certain conditions, the leading eigenvalue can be roughly approximated by the trace.

Lemma 7.1 *If $t \geqslant 4$ and $t \geqslant d \geqslant 0$ then $\lambda, \mu \in \mathbb{R}$ and $t - 4 \leqslant \lambda \leqslant t$.*

Proof First, $t \geqslant 4$ and $t \geqslant d$ imply that $t^2 \geqslant 4d$, and therefore $\lambda, \mu \in \mathbb{R}$. Next, λ is the average of t and $\sqrt{t^2-4d}$, and so $\sqrt{t^2-4d} \leqslant \lambda \leqslant t$. It remains to show that $t - 4 \leqslant \sqrt{t^2-4d}$. Note that $\sqrt{t^2-4t}$ is the geometric mean of $t-4$ and t, and so it lies between $t-4$ and t. Since $t \geqslant d$, we now have $t - 4 \leqslant \sqrt{t^2-4t} \leqslant \sqrt{t^2-4d}$, as needed. □

Lemma 7.2 *The function $f(x,y) = \log_x(y)$ maps the set*

$$S \;=\; \{\,(t,d) \in \mathbb{N}\times\mathbb{N} \mid 2 \leqslant d \leqslant t-4\,\}$$

onto a dense subset of $(0,1)$.

Proof Given $\varepsilon > 0$, fix an integer $t > e^{2/\varepsilon}$. We will show that the points $(t,2)$, $(t,3)$, $\ldots$, $(t,t-4)$ map to an ε–dense subset of $(0,1)$.

Fixing $x = t$, the function $f(t,\,\cdot\,)$ maps $[1,t]$ homeomorphically onto $[0,1]$, and maps $[2,t]$ onto an interval containing $[\varepsilon,1]$, by the choice of t. Since $f_y = \frac{1}{y\ln(x)}$, we have $|f_y(t,y)| \leqslant \frac{1}{2\ln(t)} < \varepsilon/4$ for all $y \geqslant 2$, again by the choice of t. Therefore

$$|f(t,d) - f(t,d+1)| \;<\; \varepsilon/4$$

for all integers $d \geqslant 2$. Thus the image of the set $\{(t,2),(t,3),\ldots,(t,t)\}$ is $\varepsilon/4$–dense in (and includes the endpoints of) an interval containing $[\varepsilon,1]$. Omitting the last four points, the remaining set is ε–dense in $(0,1)$. □

Now we can prove the main result of this section.

Proposition 7.3 (Density) *Given $\alpha \in (1,2)$ and $\varepsilon > 0$, there is a matrix $A \in M_2(\mathbb{Z})$ with determinant $d \geqslant 2$ and eigenvalues λ, μ with $\lambda > 1 > \mu$ such that $\big|\big(2+\log_\lambda(\mu)\big) - \alpha\big| \;<\; \varepsilon$.*

Proof Given integers t and d, the matrix

$$A(t,d) \;=\; \begin{pmatrix} t & -d \\ 1 & 0 \end{pmatrix} \;\in\; M_2(\mathbb{Z})$$

has trace t and determinant d (and eigenvalues λ,μ). Note also that $\lambda\mu = d$ implies that $2+\log_\lambda(\mu) = 1+\log_\lambda(d)$. Thus we need to choose t and d so that $\log_\lambda(d)$ is within ε of $\alpha - 1$.

First, choose a number T such that

$$\textbf{(16)} \qquad \frac{4}{(t-4)\ln(t-4)} \;\leqslant\; \varepsilon/2$$

for all $t \geqslant T$.

Next, apply Lemma 7.2 to obtain t and d such that $|\log_t(d) - (\alpha-1)| < \varepsilon/2$ and $2 \leqslant d \leqslant t-4$. We may assume in addition that $t \geqslant T$, since only finitely many points

of S violate this condition, and omitting these from S does not affect the conclusion of the lemma. By Lemma 7.1 we have

(17) $$2 \leqslant d \leqslant t-4 \leqslant \lambda \leqslant t.$$

Note that $f(x,y) = \log_x(y)$ has partial derivative $f_x = \frac{-\ln(y)}{x\ln(x)\ln(x)}$. Along the segment $\{(x,y) \mid t-4 \leqslant x \leqslant t,\ y = d\}$ we have

$$|f_x| \leqslant \frac{\ln(d)}{(t-4)\ln(t-4)\ln(t-4)} \leqslant \frac{1}{(t-4)\ln(t-4)}.$$

This implies (with Equation 16) that

$$\left|\log_{t-4}(d) - \log_t(d)\right| \leqslant \frac{4}{(t-4)\ln(t-4)} \leqslant \varepsilon/2.$$

Now, since λ is between $t-4$ and t, we have

$$|\log_\lambda(d) - \log_t(d)| \leqslant \varepsilon/2,$$

and hence $\log_\lambda(d)$ is within ε of $\alpha - 1$.

Lastly, the inequality $\mu < 1$ reduces to $d < t-1$, which holds by Equation 17. The inequality $\lambda > 1$ is clear since $t \geqslant 2$. □

References

[1] **J M Alonso**, **W A Bogley**, **R M Burton**, **S J Pride**, **X Wang**, *Second order Dehn functions of groups*, Quart. J. Math. Oxford Ser. (2) 49 (1998) 1–30

[2] **J M Alonso**, **X Wang**, **S J Pride**, *Higher-dimensional isoperimetric (or Dehn) functions of groups*, J. Group Theory 2 (1999) 81–112

[3] **B H Bowditch**, *A short proof that a subquadratic isoperimetric inequality implies a linear one*, Michigan Math. J. 42 (1995) 103–107

[4] **N Brady**, **M R Bridson**, *There is only one gap in the isoperimetric spectrum*, Geom. Funct. Anal. 10 (2000) 1053–1070

[5] **N Brady**, **M R Bridson**, **M Forester**, **K Shankar**, *Snowflake groups, Perron-Frobenius eigenvalues, and isoperimetric spectra*, Geom. Topol. 13 (2009) 141–187

[6] **M R Bridson**, *The geometry of the word problem*, from: "Invitations to geometry and topology", Oxf. Grad. Texts Math. 7, Oxford Univ. Press, Oxford (2002) 29–91

[7] **M R Bridson**, *Polynomial Dehn functions and the length of asynchronously automatic structures*, Proc. London Math. Soc. (3) 85 (2002) 441–466

[8] **S Buoncristiano**, **C P Rourke**, **B J Sanderson**, *A geometric approach to homology theory*, Cambridge University Press, Cambridge (1976)London Mathematical Society Lecture Note Series, No. 18

[9] **J Burillo**, **J Taback**, *Equivalence of geometric and combinatorial Dehn functions*, New York J. Math. 8 (2002) 169–179 (electronic)

[10] **D Calegari**, *Foliations and the geometry of 3-manifolds*, Oxford Mathematical Monographs, Oxford University Press, Oxford (2007)

[11] **D B A Epstein**, *The degree of a map*, Proc. London Math. Soc. (3) 16 (1966) 369–383

[12] **A Eskin**, **D Fisher**, **K Whyte**, *Coarse differentiation of quasi-isometries I: spaces not quasi-isometric to Cayley graphs*Preprint, arxiv:math.GR/0607207

[13] **B Farb**, **L Mosher**, *A rigidity theorem for the solvable Baumslag-Solitar groups*, Invent. Math. 131 (1998) 419–451With an appendix by Daryl Cooper

[14] **B Farb**, **L Mosher**, *On the asymptotic geometry of abelian-by-cyclic groups*, Acta Math. 184 (2000) 145–202

[15] **M Gromov**, *Partial differential relations*, Springer-Verlag, Berlin (1986)

[16] **M Gromov**, *Hyperbolic groups*, from: "Essays in group theory", Math. Sci. Res. Inst. Publ. 8, Springer, New York (1987) 75–263

[17] **M Gromov**, *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*, Birkhäuser Boston Inc., Boston, MA (1999)Based on the 1981 French original [MR0682063 (85e:53051)], With appendices by M. Katz, P. Pansu and S. Semmes, Translated from the French by Sean Michael Bates

[18] **V S Guba**, **M V Sapir**, *On Dehn functions of free products of groups*, Proc. Amer. Math. Soc. 127 (1999) 1885–1891

[19] **H Hopf**, *Zur Topologie der Abbildungen von Mannigfaltigkeiten*, Math. Ann. 102 (1930) 562–623

[20] **A Y Ol′shanskiĭ**, *Hyperbolicity of groups with subquadratic isoperimetric inequality*, Internat. J. Algebra Comput. 1 (1991) 281–289

[21] **P Papasoglu**, *On the sub-quadratic isoperimetric inequality*, from: "Geometric group theory (Columbus, OH, 1992)", Ohio State Univ. Math. Res. Inst. Publ. 3, de Gruyter, Berlin (1995) 149–157

[22] **P Papasoglu**, *Isodiametric and isoperimetric inequalities for complexes and groups*, J. London Math. Soc. (2) 62 (2000) 97–106

[23] **C P Rourke**, *Presentations and the trivial group*, from: "Topology of low-dimensional manifolds (Proc. Second Sussex Conf., Chelwood Gate, 1977)", Lecture Notes in Math. 722, Springer, Berlin (1979) 134–143

[24] **M V Sapir**, **J-C Birget**, **E Rips**, *Isoperimetric and isodiametric functions of groups*, Ann. of Math. (2) 156 (2002) 345–466

[25] **J R Stallings**, *A graph-theoretic lemma and group-embeddings*, from: "Combinatorial group theory and topology (Alta, Utah, 1984)", Ann. of Math. Stud. 111, Princeton Univ. Press, Princeton, NJ (1987) 145–155

[26] **X Wang**, *Second order Dehn functions of finitely presented groups and monoids* (1996)PhD Thesis, University of Glasgow

[27] **X Wang**, **S J Pride**, *Second order Dehn functions and HNN-extensions*, J. Austral. Math. Soc. Ser. A 67 (1999) 272–288Group theory

[28] **R Young**, *A note on higher-order filling functions*Preprint, arxiv:0805.0584

Mathematics Department, University of Oklahoma
Norman, OK 73019, USA

nbrady@math.ou.edu, forester@math.ou.edu

http://math.ou.edu/~nbrady, http://math.ou.edu/~forester