FRACTIONAL DISTORTION IN HYPERBOLIC GROUPS

PALLAVI DANI AND TIMOTHY RILEY

Abstract. For all integers $p > q > 0$ and $k > 0$, and all non-elementary torsion-free hyperbolic groups H, we construct a hyperbolic group G in which H is a subgroup, such that the distortion function of H in G grows like $\exp^k(n^{p/q})$. Here, $\exp^k$ denotes the k-fold-iterated exponential function.

Contents

Date: March 13, 2024.

This work of the first author was supported by a grant from the Simons Foundation (#426932, P. D.) and by NSF Grant No. DMS-1812061. This work of the second author was supported by a grant from the Simons Foundation (#318301, T. R.).

A large part of this work was conducted in 2016–17, and the first author thanks the Simons Laufer Mathematical Sciences Institute (formerly MSRI) for its hospitality during the Semester Program on Geometric Group Theory (2016), as well as Cornell University Department of Mathematics and the Association for Women in Mathematics for the opportunity to visit Cornell University as a Michler Fellow in 2017. The second author is grateful for the hospitality of Cambridge University's DPMMS 2019–20.

1. Introduction

The landscape of subgroups of hyperbolic groups is poorly understood. Whether all one-ended hyperbolic groups have surface subgroups is a celebrated open question. What functions are Dehn functions of subgroups of hyperbolic groups is widely open. This article addresses another fundamental issue: What distortion can subgroups of hyperbolic groups exhibit? Indeed, in his 1998 survey [Mit98b] Mitra (now known as Mj) asked: "Given any increasing function $f : \mathbb{N} \to \mathbb{N}$, does there exist a hyperbolic subgroup H of a hyperbolic group G such that the distortion of H is of the order of $\exp(f(n))$."

Let $\exp^k$ denote the k-fold iterated exponential function $\mathbb{N} \to \mathbb{R}$ defined by $\exp^1(n) = \exp(n)$ and, for $k = 2, 3, \ldots$ by $\exp^k(n) = \exp(\exp^{k-1}(n))$. The notation $\simeq$ will be explained in Section 2. Our main result is:

Theorem 1.1. *Given integers $p > q > 0$ and $k > 0$, there exists a hyperbolic group G and free subgroup $H \le G$ of distortion $\mathrm{Dist}_H^G(n) \simeq \exp^k(n^{p/q})$.*

In Section 3 we outline our construction of these G and H and highlight the new techniques we introduce. Our G are of infinite height (so do not speak to an old open question of Swarup)—see Section 17. In the case $k = 1$ they can be made residually finite, $C'(1/6)$, CAT(-1), and virtually special—see Section 4.

In Section 16 we the leverage examples of Theorem 1.1 and of [BBD07, BDR13, Mit98a, Mit98b] so as to make the distorted subgroup be any given non-elementary torsion-free hyperbolic group:

Theorem 1.2. *Let H be any non-elementary torsion-free hyperbolic group and let f be any of the following functions:*

(1) $f(n) = \exp^m(n^{p/q})$, for any integers $m \ge 1$ and $p \ge q \ge 1$.
(2) f is any one of the Ackermann-function representatives of the successive levels of the Grzegorczyk hierarchy of primitive recursive functions.

Then there exists a hyperbolic group G with $H < G$ such that $\mathrm{Dist}_H^G \simeq f$.

This paper also contains results we needed to prove Theorem 1.2 which may be of independent interest. Theorem 15.4 assembles results of Bowditch, Dahmani, and Osin into a combination theorem for the hyperbolicity of amalgams $\Gamma = A *_C B$. Theorem 16.2 relates the distortion of C in A and of C in B to that of A in $\Gamma = A *_C B$. Building on the $k \ge 2$ case, proved by I. Kapovich in [Kap99], Lemma 16.1 states that in every non-elementary torsion-free hyperbolic group H there is, for any $k \ge 2$, a malnormal quasiconvex free subgroup F of rank k. Lemma 15.5 states that if a semi-direct product $G = F_l \rtimes F_m$ of finite rank free groups is hyperbolic, then the F_m-factor is quasiconvex and malnormal in G.

Background. At first sight, it is surprising that subgroups of hyperbolic groups can display any distortion given the tree-like geometry of the thin-triangle condition that defines hyperbolicity. Every $\mathbb{Z}$ subgroup of a hyperbolic group is undistorted—e.g., [BH99, III.Γ Corollary 3.10]. Finitely generated subgroups H of hyperbolic groups G are undistorted (meaning linear distortion, $\mathrm{Dist}_H^G(n) \simeq n$) if and only if they are quasi-convex, and in that event they are themselves hyperbolic. Above linear there is a gap in the spectrum of possible distortion functions: a consequence of the exponential divergence property of hyperbolic spaces is that if a finitely generated subgroup of a hyperbolic group is subexponentially distorted, then it is quasi-convex [Kap01, Proposition 2.6]. Theorem 1.1 sweeps out much of the landscape of possibilities above exponential.

Prior to Theorem 1.1, only sporadic examples of distortion functions for subgroups of hyperbolic groups were known. Subgroups of finite-rank free groups and of hyperbolic surface groups are undistorted [Pit93, Sho91]. Wise [Wis04b] generalized this result to fundamental groups of non-positively curved, piecewise Euclidean 2-complexes which enjoy a suitable negative sectional curvature condition. The free factor in any hyperbolic free-by-cyclic group is exponentially distorted [BF92, BF96, Bri00]. Mitra [Mit98a, Mit98b] constructed, for each integer $k \geq 1$, a hyperbolic group with a free subgroup distorted like $n \mapsto \exp^k(n)$, and an example with distortion growing faster than any iterated exponential. Barnard, Brady and Dani [BBD07] developed Mitra's constructions into more explicit examples that are also CAT(-1). Baker and Riley [BR13] exhibited a finite-rank free subgroup of a hyperbolic group that is distorted like $n \mapsto \exp^2(n)$ and is also pathological in that there is no Cannon–Thurston map. Brady, Dison, and Riley [BDR13] constructed, for every primitive recursive function, a hyperbolic 'hydra' group with a finite-rank free subgroup whose distortion outgrows that function. The Rips construction produces examples displaying yet more extreme distortion. Applied to a finitely presentable group with unsolvable word problem the construction yields a hyperbolic ($C'(1/6)$ small-cancellation) group G with a finitely generated subgroup N such that Dist_N^G is not bounded from above by a recursive function—see [AO02, §3.4], [Far94, Corollary 8.2], [Gro93, §3, 3.K_3''] and [Pit92].

The subgroup N in the Rips construction is not finitely presentable. In fact, it follows from a theorem of Bieri in [Bie81] that N is finitely presented if and only if the quotient Q is finite. So the Rips construction cannot be used to construct examples such as those in Theorem 1.1. Instead, we use a modification of the Rips construction: starting with a particular finitely presented group Q, we realize it as the quotient of a group presentation that satisfies $C'(1/6)$ and other small-cancellation conditions, and find a free subgroup which is distorted, but not normal. Several additional nuances in our construction guarantee that we get the desired distortion estimates. We outline this in Section 3.

In contrast to the situation with hyperbolic groups, a broad family of functions are known to be distortion functions of subgroups of CAT(0) groups. Indeed, Olshanskii and Sapir [OS01, Theorem 2] used a Mihailova-style construction to show that the set of distortion functions of finitely generated subgroups of $F_2 \times F_2$ coincides with the set of Dehn functions of finitely presented groups. Such functions are known to have wide scope thanks to the S-machines of [SBR02, Sap18].

In finitely presented groups, even $\mathbb{Z}$-subgroups can exhibit essentially any distortion: Olshanskii [Ol′97] showed that every computable function $\mathbb{N} \to \mathbb{N}$, satisfying some straight-forwardly necessary conditions, is $\simeq$-equivalent to the distortion function of such as subgroup.

Application to Dehn functions. What functions can be $\simeq$-equivalent to Dehn functions is understood in detail thanks to [BB00, BBFS09, Ol′97, SBR02]. However, because the most comprehensive results depend on deeply involved constructions, we note that our examples give some explicit examples as follows.

Corollary 1.3. *Our groups G yield explicit examples, for integers $p > q > 0$ and $k > 0$, of groups with Dehn functions growing $\simeq \exp^k(n^{p/q})$, namely the free product with amalgamation $G *_H G$ of two copies of G along H, and the HNN-extension $G*_\tau$ of G with stable letter τ that commutes with all elements of H.*

Proof. Theorem 6.20 in Chapter III.Γ of [BH99] gives upper and lower bounds on the Dehn functions of $G *_H G$ and $G*_\tau$ in terms of the Dehn function of G (which is $\simeq n$ because G is hyperbolic) and Dist_H^G. Up to $\simeq$, these bounds agree with each other and with Dist_H^G since Dist_H^G is super-exponential. $\square$

Next steps. A potential application of our examples is to constructing subgroups of CAT(0) groups or hyperbolic groups exhibiting a range of Dehn functions. One might, for example, look to embed the doubles of Corollary 1.3 in CAT(0) groups in the manner of [BT21]. However, our distorted subgroups not being normal is an obstacle to making this work.

Sapir's S-machines emulate general computing machines in appropriately constructed (and always non-hyperbolic) finitely presented groups. Looking yet further ahead, one might view the techniques we introduce here as groundwork for doing the same within appropriately constructed hyperbolic groups.

The organization of this article. Section 2 gives preliminaries on words, hyperbolicity, distortion, and the equivalence relation $\simeq$ on functions $\mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. Section 3 is an overview of our construction. Section 4 contains the definition of our groups G used to prove Theorem 1.1 in the case $m = 1$ and catalogs their small-cancellation conditions. Section 5 gives some immediate consequences of those small-cancellation conditions. Section 6 reviews the definition of a corridor in a van Kampen diagram and introduces a more general dual notion we call *tracks*, which may branch, unlike corridors. Section 7 gives two HNN-structures for G and establishes that H is free. Section 8 contains our proof of the lower bound on the distortion of H in G. Sections 9–14 prove the upper bound. Section 9 concerns how a van Kampen diagram Δ over G being reduced limits the patterns of tracks within it. Section 10 gives general results about paths across discs, which we will apply to tracks in Δ. Section 11 argues that tracks are further constrained in what we call a *distortion diagram*, meaning a Δ exhibiting how a word in the generators of H equals a shorter word in the generators of G. Section 12 concerns what we call (a_2, b_q)-tracks, which are a device we use to connect growth within Δ to the presence of certain edges in its boundary. Section 13 contains estimates which are made possible by the restrictions proved in 9–11 and which culminate in an upper bound on the distortion of H in G. Section 14 contains a calculation in a free-by-cyclic quotient Q of G that is postponed from the prior section and is where the fraction p/q ultimately enters. Section 15 promotes our examples to iterated exponential functions, and so completes our proof of Theorem 1.1. Section 16 explains how we leverage our examples to prove Theorem 1.2. Section 17 contains a proof that our examples have infinite height.

Acknowledgements. We are grateful to Ilya Kapovich and Mahan Mj for suggesting that we promote Theorem 1.1 to Theorem 1.2, and to Jason Manning for guidance on the associated literature.

2. Preliminaries

A *word w on* a set of letters $\mathcal{A}$ is an expression $a_1^{\varepsilon_1} \cdots a_m^{\varepsilon_m}$ where $m \geq 0$, $a_i \in \mathcal{A}$, and $\varepsilon_i = \pm 1$ for all i. It is *positive* when $\varepsilon_i = 1$ for all i. Its length $|w|$ is m. The word metric $d_S(g, h)$ on G gives the length of a shortest word on S that represents $g^{-1}h$. We use d_G or d in place of d_S when the generating set is understood from the context.

A finitely generated group is *hyperbolic* when its Cayley graph has the property that there exists $\delta > 0$ such that all geodesic triangles are δ*-thin*: that is, each of its three sides is in the δ-neighbourhood of the other two. The existence of such a δ does not depend on the finite generating set (but the values of δ for which the condition holds generally will). See, for example, [BH99, Gro87] for further background.

Suppose S and T are finite generating sets for a group G and subgroup H, respectively. The *distortion function* $\text{Dist}_H^G : \mathbb{N} \to \mathbb{N}$ measures how H sits as a metric space in G by comparing the restriction of the word metric d_S on G associated to S to the word metric d_T on H associated to T:

$$\text{Dist}_H^G(n) \ := \ \max\{\, d_T(e,g) \ \mid \ g \in H \text{ with } d_S(e,g) \leq n \,\}.$$

Replacing S and T by other finite generating sets will produce a distortion function that is $\simeq$-equivalent in the following sense. For $f, g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ write $f \preceq g$ when there exists $C > 0$ such that $f(n) \leq Cg(Cn + C) + Cn + C$ for all $n \geq 0$, and $f \simeq g$ when $f \preceq g$ and $g \preceq f$. Apply these relations to functions $\mathbb{N} \to \mathbb{R}_{\geq 0}$ by extending the domains to $\mathbb{R}_{\geq 0}$ and having the functions be constant on the intervals $[n, n+1)$.

The following two lemmas concern features of the $\simeq$-relation that will be important for us. The first is routine and we present it without proof.

Lemma 2.1.

(1) For $\alpha, \beta \geq 1$, $2^{n^\alpha} \simeq 2^{n^\beta}$ if and only if $\alpha = \beta$.
(2) For $\alpha \geq 1$ and $C > 1$, $C^{n+n^\alpha} \simeq C^{n^\alpha} \simeq 2^{n^\alpha}$.

For our proof of the lower bound in Theorem 1.1, we will exhibit a sequence of words that represent elements of H, but can only be expressed by long words on the generators of H. The force of the following lemma is that, despite the lengths of our words forming a sparse sequence, we can draw the desired conclusion.

Lemma 2.2. *Suppose H is a subgroup of G and both are finitely generated. Suppose $p > q > 0$ are integers, $C_1, C_2, C_3 > 0$ are constants, and $w_1, w_2, \ldots$ is a sequence of words on the generators of G. Suppose that w_n represents an element of H for all n, and*

$$C_1 n^q \leq |w_n| \leq C_2 n^q \quad \textit{but} \quad d_H(e, w_n) \geq C_3 2^{n^p}.$$

Then $\text{Dist}_H^G(n) \succeq 2^{n^{p/q}}$.

Proof. Remark 2.1 in [BBFS09] is that to verify $g \succeq f$ for $f, g : \mathbb{N} \to \mathbb{N}$, it suffices to have $g(m_n) \geq f(m_n)$ on a sequence (m_n) of integers such that $m_n \to \infty$ as $n \to \infty$ and such that there exists $C > 0$ with $m_{n+1} \leq C m_n$ for all n. If $C_4 = (q+1) \max_{i=0,\ldots,q} \binom{q}{i}$, then $(n+1)^q \leq C_4 n^q$ for all n. So there is a C such that the sequence $m_n = |w_n|$ satisfies this condition. Now

$$\text{Dist}_H^G(|w_n|) \ \geq \ d_H(e, w_n) \ \geq \ C_3 2^{n^p} \ \geq \ C_3 2^{\left(\frac{1}{C_2}|w_n|\right)^{p/q}}.$$

So $\text{Dist}_H^G(n) \succeq 2^{\left(\frac{n}{C_2}\right)^{p/q}}$, and the result then follows from Lemma 2.1(2) (by taking $C = 2^{(C_2^{-p/q})}$ and $\alpha = p/q$).

$\square$

We will work extensively with van Kampen diagrams. There are many introductory accounts in the literature.

3. Motivation for our construction

In this section, we offer some insights into the origins of our construction. The formal definition of our group-pair $H < G$, used to prove Theorem 1.1 in the case $m = 1$, follows in Section 4.

Our construction begins with the free-by-cyclic group

$$Q \ = \ \langle \ a_1, b_0, \ldots, b_p \mid a_1^{-1} b_i a_1 = \varphi(b_i) \ \ \forall i \ \rangle \tag{3.1}$$

where φ is the polynomially-growing automorphism of the free group $F = F(b_0, \ldots, b_p)$ mapping $b_i \mapsto b_{i+1} b_i$ for $i \neq p$ and $b_p \mapsto b_p$.

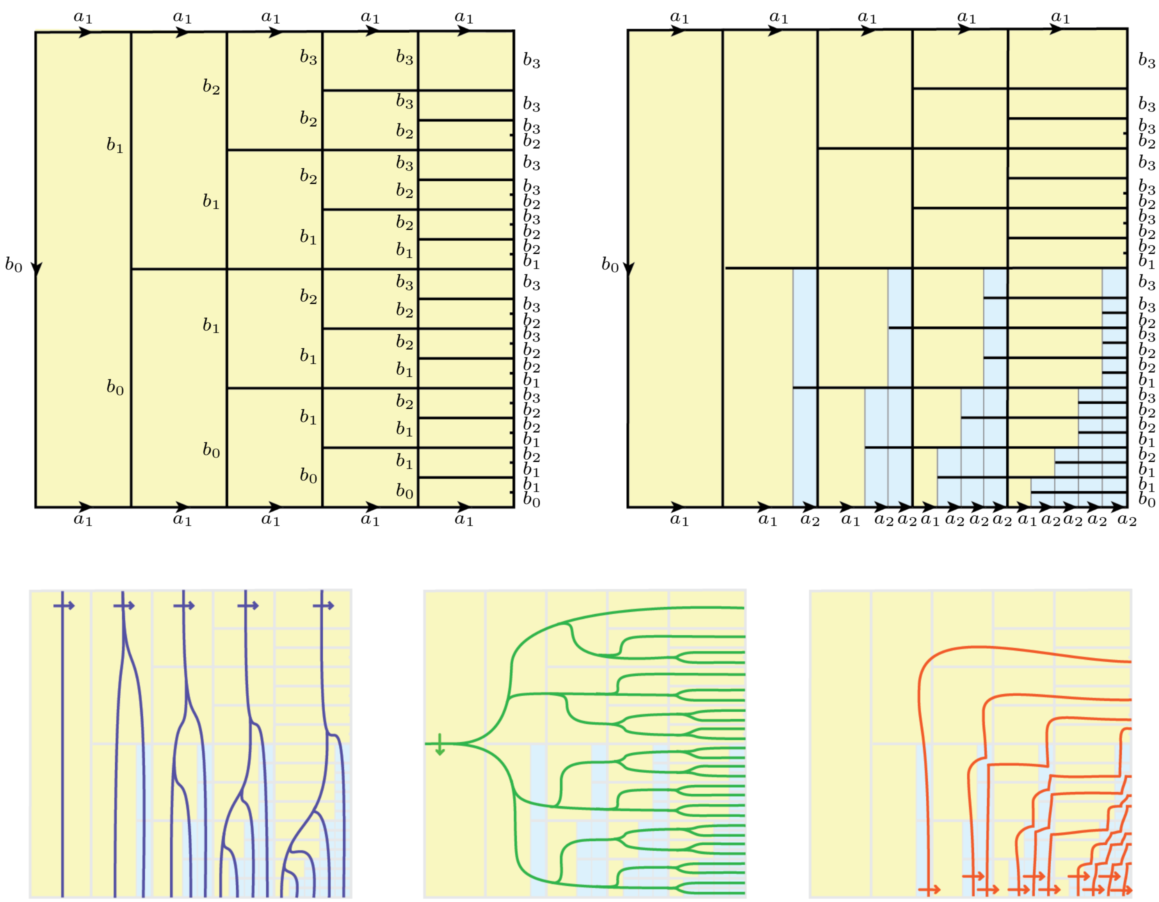


Figure 3.1. Top left: the van Kampen diagram D_0 over Q for $a_1^{-n} b_0 a_1^n = \varphi^n(b_0)$ when $n = 5$ and $p = 3$. Top right: the corresponding diagram D_2 over G_2. Lower left, middle and right: a-tracks, b-tracks, and (a_2, b_q)-tracks through D_2.

The van Kampen diagram D_0 over Q pictured top-left in Figure 3.1 (for the case $n = 5$ and $p = 3$) shows how $a_1^{-n} b_0 a_1^n = \varphi^n(b_0)$ equals a positive word λ on $b_0, b_1, \ldots, b_p$ which contains $\simeq n^i$ letters b_i for $i = 0, \ldots, p$ (Lemma 14.1). The contribution of b_p dominates, so the length of λ is $N = |\lambda| \simeq n^p$.

Next, we define

$$G_1 \ = \ \langle \ Q, x \mid b_j^{-1} x b_j = x^2 \ \ \forall j \ \rangle.$$

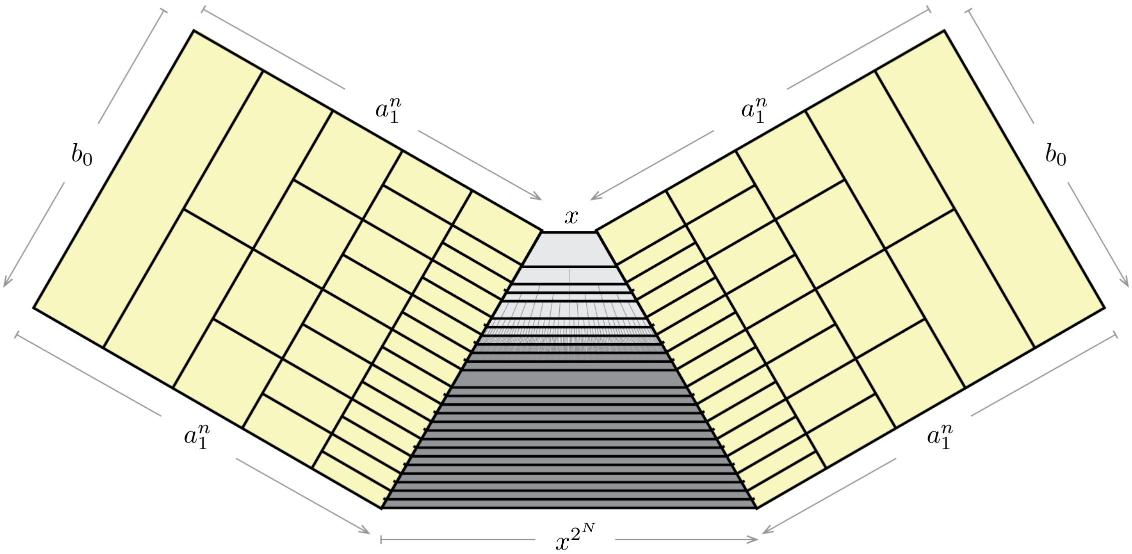


Figure 3.2. A van Kampen diagram Δ_1 over G_1 for $a_1^{-n} b_0^{-1} a_1^n x a_1^{-n} b_0 a_1^n = x^{2^N}$ when $n = 5$, $p = 3$, and $N = |\varphi^n(b_0)| = 26$.

As shown in Figure 3.2, attaching a copy of D_0 and a copy of its mirror image to a diagram for $\lambda^{-1} x \lambda = x^{2^N}$ along its two paths labelled λ gives a van Kampen diagram Δ_1 over G_1 for the relation

$$a_1^{-n} b_0^{-1} a_1^n x a_1^{-n} b_0 a_1^n \ = \ x^{2^N}. \tag{3.2}$$

This diagram illustrates that there is a word of length $\simeq 2^{n^p}$ in $H_1 = \langle x \rangle$, whose length in G_1 is $\simeq n$. As there is a family of such diagrams indexed by n, this shows that $\mathrm{Dist}_{H_1}^{G_1}(n) \succeq 2^{n^p}$.

Next we elaborate on this construction in a way that plays off the $\simeq n^p$ letters b_p against the $\simeq n^q$ letters b_q in λ. We introduce a new generator a_2 and we modify the relation $a_1^{-1} b_{q-1} a_1 = b_q b_{q-1}$ of G_1 to $a_1^{-1} b_{q-1} a_1 a_2 = b_q b_{q-1}$, so that for every new b_q created by φ within D_1, an a_2 is created as well. Furthermore, we add the relations that a_2 commutes with b_j for all j, allowing these newly created edges to *flow to the boundary* as shown in the diagram on the right in Figure 3.1. The resulting diagrams D_2 can be mapped onto D_1 by suitably collapsing all the a_2-edges and the commutator 2-cells in which they occur. As for the construction of Δ_1, assemble D_2, its mirror-image, and our diagram for $\lambda^{-1} x \lambda = x^{2^N}$ to get a diagram Δ_2 that demonstrates that x^{2^N} equates in a group G_2 to a word $a_1^{-n} b_0^{-1} a_1^n x a_1^{-n} b_0 a_1^n$ with $\simeq n^q$ letters $a_2^{\pm 1}$ inserted. This construction suggests that the distortion function of $\langle x \rangle$ in G_2 grows like $n^q \mapsto 2^{n^p}$, and therefore like $n \mapsto 2^{n^{p/q}}$.

Now, G_2 is not hyperbolic. So next we *hyperbolize* its presentation using an approach similar to Wise's version of the Rips construction [Wis03]. We add *noise* to each relation so that the resulting presentation satisfies small-cancellation conditions including $C'(1/6)$. This is achieved by replacing x by three letters t, x_1, x_2, and introducing a *noise word* on t, x_1, x_2 to each relation. We then add relations to allow the noise to *flow* to the boundary of the diagram and then (in the two triangles at the bottom of Figure 3.3) be moved past the $a_1^{\pm 1}, a_2^{\pm 1}$ and collected together. These additional relations play a similar role to the commuting relations involving a_2 introduced above; they allow noise to move past a- and b-letters (but only in one direction) at the expense of introducing additional noise. The resulting group G_3 admits diagrams Δ_3 which map onto Δ_2 on suitably collapsing the edges labelled by noise letters and suitably collapsing the 2-cells that allow the noise to *flow*. We take $H_3 = \langle t, x_1, x_2 \rangle$.

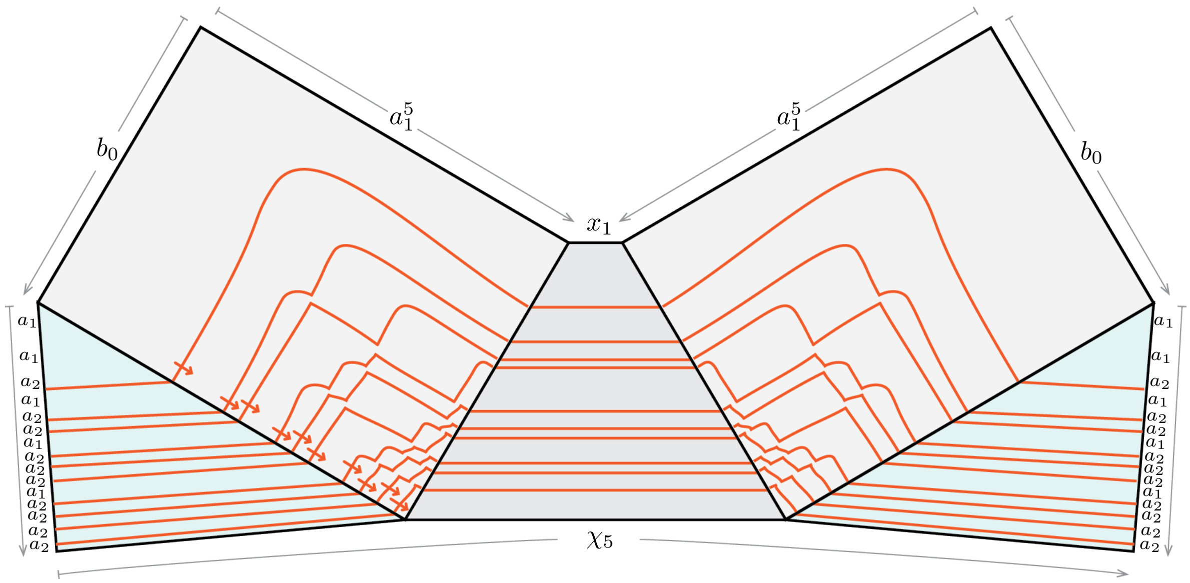


FIGURE 3.3. A schematic of a van Kampen diagram Δ over G for $w_5 = (a_1\, a_1a_2\, a_1a_2^2\, a_1a_2^3\, a_1a_2^4)^{-1} b_0^{-1} a_1^5 \ x_1 \ a_1^{-5} b_0 (a_1\, a_1a_2\, a_1a_2^2\, a_1a_2^3\, a_1a_2^4)$ on the generators of G equals a word χ_5 on the *noise* letters. Its (a_2, b_q)-tracks are shown. Each meets the boundary at a pair a_2-edges.

The diagram of Figure 3.3 shows the $n = 5$ instance of a family of diagrams demonstrating how words w_n on a_1, a_2, b_0, a_1, x_1 represent the same elements of G_3 as words χ_n on t, x_1, x_2. Because the effect is so pronounced, the figure cannot do justice to the exponential expansion in the direction of χ_n.

While this family of diagrams provides the desired $2^{n^{p/q}}$ lower bound on the distortion of H_3 in G_3, some issues remain. Firstly, with the presentation described, we cannot get a matching $2^{n^{p/q}}$ upper bound on distortion. If we replace the two b_0 letters in (3.2) with b_i, where $i < q$, and then construct diagrams Δ_3 as described above, then they will exhibit $n \mapsto 2^{n^{(p-i)/(q-i)}}$ distortion of H_3, which is greater than $2^{n^{p/q}}$. Secondly, allowing the noise letters to interact with both a- and b-letters prevents us from establishing an HNN-structure on the group (the iterated HNN-structure of Proposition 7.4) which will allow us to prove that our distorted subgroup H is free.

Both issues are solved by making the role of the noise more nuanced. We introduce two pairs of noise letters, x_1, x_2 and y_1, y_2 (in addition to the noise letter t). For $i > 0$, b_i interacts with x_1 and x_2 but not y_1 and y_2, while a_1 and a_2 interact with y_1 and y_2, and not x_1 and x_2. Conjugation by b_0 converts x_1 and x_2 to words on y_1 and y_2. This way we arrive at our group G whose defining relations are set out in Figure 4.1. We take H to be the subgroup generated by t, y_1, y_2.

Over G there are diagrams Δ of the form shown in Figure 3.3 exhibiting $2^{n^{p/q}}$-distortion. This construction is the heart of our proof in Section 8 that $\mathrm{Dist}_H^G(n) \succeq 2^{n^{p/q}}$.

As for the reverse bound $\mathrm{Dist}_H^G(n) \preceq 2^{n^{p/q}}$, the aforementioned diagrams yielding larger distortion no longer exist because if we replace b_0 with b_i where $i > 0$ in the construction of Δ, then $\partial\Delta$ has a long word in a_1, a_2, t along with x_1, x_2 rather than along with y_1, y_2. We have long words on letters that are not all generators for H and we can no longer attach the triangular subdiagrams that separate the a_1, a_2 from the the noise letters.

However, to establish the upper bound we must prove that no other "bad" diagrams exist. To achieve this we study what we call *distortion diagrams*—reduced diagrams Δ, subject to natural simplifying assumptions, which exhibit how a word χ on t, y_1, y_2 can be represented by a shorter word w on the generators of G. We show in Sections 9–12 that such a Δ is subject to considerable rigidity. Our argument shows that Δ is so constrained that it strongly resembles the diagrams described above and is thereby subject to estimates that yield the $2^{n^{p/q}}$ upper bound.

Three features of G impose this rigidity.

(1) *Noise in Δ must flow towards χ and orthogonally to tracks.* This refers to the propagation of ("noise") letters t, x_1, x_2, and y_1, and y_2 through Δ. Figures 3.1, 3.3 and 3.4 show *tracks* through the various diagrams we constructed above. Introduced in Section 6, tracks are generalizations of *corridors*. We will be concerned with four types: a-tracks, b-tracks, t-tracks, and (a_2, b_q)-tracks.

An a-track is a path in the dual of Δ that crosses successive edges labelled by a-letters (meaning a_1 and a_2). A b-track is the same, but for edges labelled by $b_0, \ldots, b_p$. A t-track crosses t-edges—the use of t is a distinctive feature of Wise's version of the Rips construction; it renders the group an HNN-extension of a free group, with t the stable letter (see Proposition 7.1). This extra structure, manifested in the geometry of t-tracks, facilitates analysis of G. We will describe (a_2, b_q)-tracks in (2) below. As there are three a-letters or three b-letters in some of the defining relators, a-tracks and b-tracks can branch.

As noise advances across successive tracks it increases exponentially in length. A consequence of the small-cancellation condition enjoyed by the Rips words used in the defining relators is that noise cannot substantially cancel within a diagram—it must instead emerge on the boundary. Therefore, if we assume that w is of minimal length among all words on the generators of G that equal χ in G, then almost all this noise must emerge in χ. If many noise letters emerge in w, then their blow up en route there would result in it being possible to cut a subdiagram out of Δ to get a new diagram that demonstrated a shorter word than w equals χ in G.

This also has helpful consequences for the orientation of tracks—see Lemma 11.2. In short, they must be oriented towards χ because otherwise they would act as blockades for the flow of noise.

(2) *(a_2, b_q)-tracks.* The subject of Section 12, these are paths through van Kampen diagrams that cross successive a_2- and b_q-edges. Examples are found in Figures 3.1 and 3.3. In most defining relators of G there are either zero or two a_2-letters, and ditto for b_q-letters. If an (a_2, b_q)-track enters a 2-cell labelled by such a relator across an a_2-edge, then it exists across the other a_2-edge, and ditto for b_q-edges. However our presentation for G has a defining relator ($r_{1,q-1}$ of Figure 4.1) with one a_2-letter and one b_q-letter, and a defining relator ($r_{2,q}$ of Figure 4.1) that has two a_2-letters and two b_q-letters. On entering the 2-cell of the former type across its a_2-edge it exits across its b_q-edge (or vice versa). On entering a 2-cell of

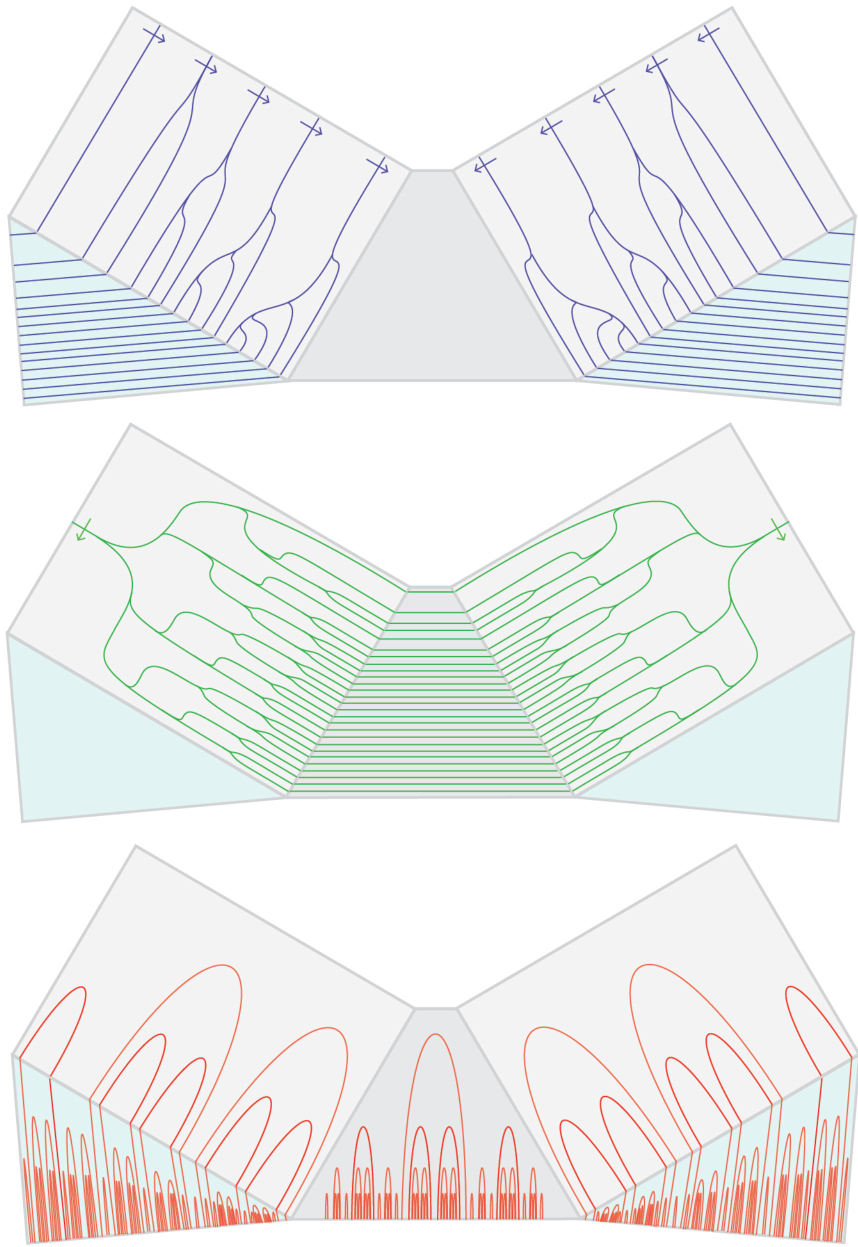


FIGURE 3.4. Top, middle, lower: a-tracks, b-tracks, and t-tracks through the diagram Δ of Figure 3.3. The lower diagram is intended only to convey the nesting pattern of the t-tracks. The pattern expands too rapidly towards χ to be displayed accurately.

the latter type across an a_2-edge (resp. b_q-edge), it exits across the b_q-edge (resp. a_2-edge) that is oriented the same way. These conventions ensure that every a_2- and b_q-edge in a van Kampen diagram over G is crossed by exactly one (a_2, b_q)-track, no (a_2, b_q)-track can cross itself, and no two (a_2, b_q)-tracks can cross each other. So (a_2, b_q)-tracks associate to every b_q-edge in a diagram Δ a pair of edges labelled by a_2 or b_q on the boundary. So, if the automorphism φ gives $\sim n^p$ growth within Δ, then the length of w must be $\succeq n^q$.

(3) *x- versus y-noise, and b_0-tracks.* It is significant that our generating set for H consists of the noise letters t, y_1, y_2 but omits x_1 and x_2. It is possible for x-noise to flow across b-tracks but impossible for y-noise. And x-noise becomes y-noise when (and only when) it crosses b_0-tracks (particular examples of b-tracks). This means that stacks of nested b-tracks must include at most one b_0-track and that b_0-track must be the closest to χ.

In Section 13 we use these ideas to reduce the problem of bounding $|\chi|$ from above to establishing an inequality concerning the quotient Q of (3.1) (specifically, we reduce it to Lemma 13.11), and this is where the "$n^{p/q}$" in our distortion functions is ultimately established, as we explain in Section 14. Combined with the blow-up that comes from the flow of noise through Δ, it gives our $2^{n^{p/q}}$ upper bound on the distortion of H in G.

We leverage our examples to get iterated exponential distortion functions and complete our proof of Theorem 1.1 in Section 15. The strategy is to amalgamate G with a chain of hyperbolic free-by-free groups following Brady and Tran [BT21], and then prove and apply a combination theorem for the hyperbolicity of amalgams.

In Section 16 we show that the distorted subgroup H need not be free of rank 3, but rather can be taken to be any torsion-free non-elementary hyperbolic group, proving Theorem 1.2. For this we establish the existence (in Lemma 16.1, after [Kap99]) of undistorted free subgroups of any rank in torsion-free non-elementary hyperbolic groups, apply the same combination theorem to amalgamate these with our examples in a new hyperbolic group, and then we prove the estimates on the distortion function by means of an appropriate general theorem (Theorem 16.2) concerning distortion in amalgams.

4. The definition of our groups

Here we will define the group G which will prove Theorem 1.1 in the case $k = 1$. In Section 15 we will explain how the case $k = 1$ leads to the result for other k.

We fix integers $p > q > 0$. Then G has presentation

$$\mathcal{P} \;=\; \langle\, a_1, a_2, b_0, \ldots, b_p, t, x_1, x_2, y_1, y_2 \;\mid\; \mathcal{R} \,\rangle$$

where $\mathcal{R}$ is the set of $5p + 11$ defining relators displayed in Figure 4.1. Our notation X_* and Y_* is intended to indicate indexing that we have chosen to suppress. Every element of $\mathcal{R}$ is a word of the form $t^{-1}utv^{-1}$ where u and v are words on generators other than t. Each has two or three *Rips subwords*, denoted X_* or Y_*, from sets $\mathcal{X} = \{X_1, X_2, \ldots, X_{14p}\}$ and $\mathcal{Y} = \{Y_1, Y_2, \ldots, Y_{30}\}$ of pairwise disjoint subwords of the infinite Rips words $x_1x_2^1\,x_1x_2^2\,x_1x_2^3\cdots$ and $y_1y_2^1\,y_1y_2^2\,y_1y_2^3\cdots$, respectively, chosen in a manner we will explain momentarily. We stress that each X_* and Y_* occurs once in $\mathcal{P}$ and does so as a subword of one defining relator. So, if an X_* or Y_* can be read around a portion of the boundary circuit of a 2-cell in a van Kampen diagram (see Section 6) over $\mathcal{P}$, then that Rips word uniquely determines the defining relator that 2-cell corresponds to. This use of t and Rips words is a variation on Wise's [Wis03] HNN-version of Rips' Construction [Rip82]. (Our example G departs in some respects from Wise's framework. Wise has two X_* subwords in each defining relator, has only two 'noise' generators x_1 and x_2, and has additional defining relators that ensure that $\langle t, x_1, x_2\rangle$ is a normal subgroup.)

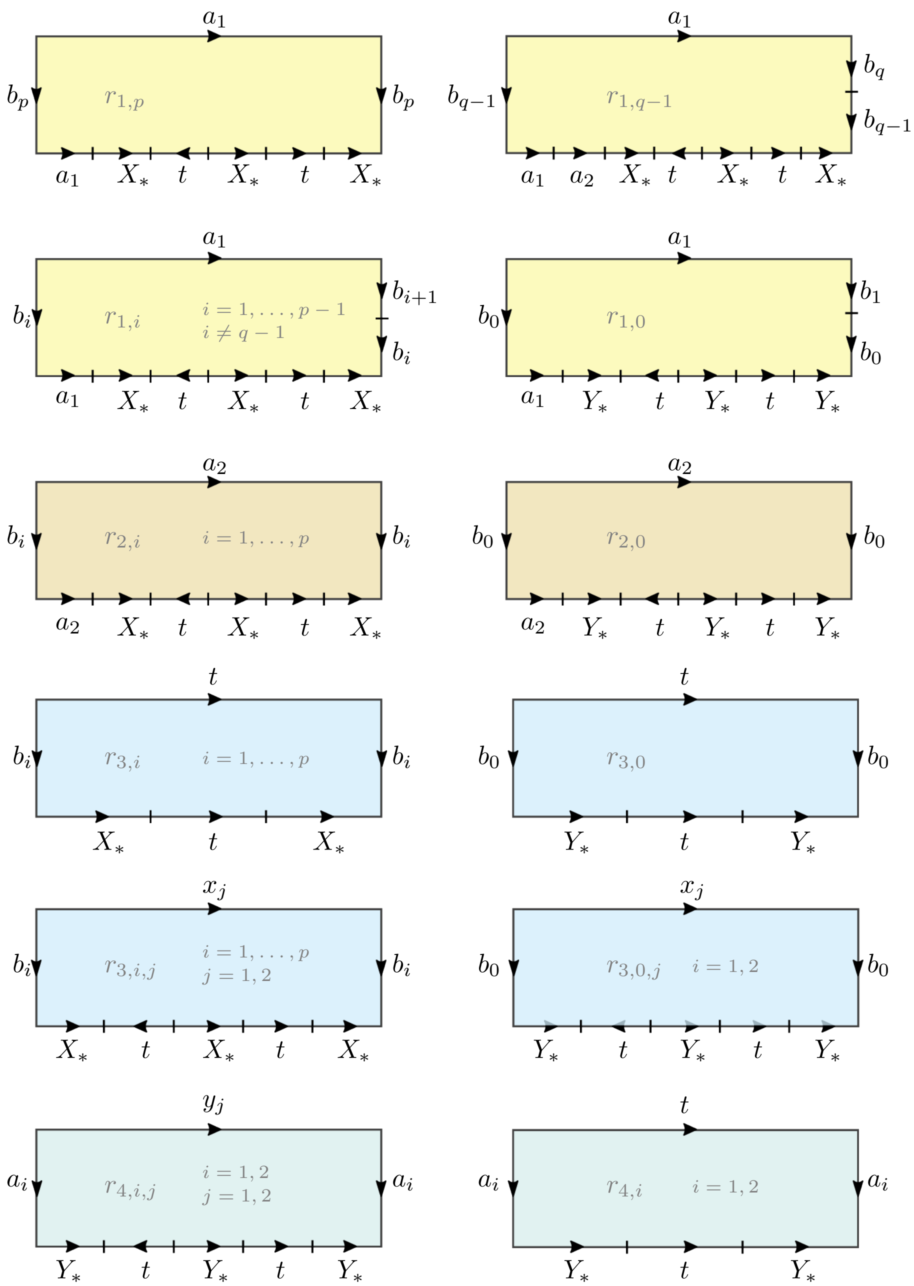


FIGURE 4.1. Defining relators for our group G

Suppose S is a set of words on $A \cup A^{-1}$ for some alphabet A. A cyclic conjugate of a word w is a word $s_2 s_1$ such that s_1 is a prefix of w and s_2 a suffix such that $s_1 s_2 = w$. Let $\mathcal{C}(S)$ be the set of all cyclic conjugates of words in $S^{\pm 1}$. Assume that all elements of $\mathcal{C}(S)$ are reduced. A *piece* is a common prefix π of a pair of distinct words πu and πv in $\mathcal{C}(S)$.

We choose the Rips subwords X_* and Y_* so that each has length at least 100 and we have:

i. *The uniform $C'(1/6)$-condition for $\mathcal{R}$.* Every piece has length strictly less than a sixth of the length of the shortest relator in $\mathcal{R}$.

ii. *The $C(3)$-condition for the union S of the 3- and 5-element generating sets of the terminal vertex groups of Table 7.1.* No element of $\mathcal{C}(S)$ is a concatenation of fewer than 3 pieces.
iii. *The $C'(1/4)$ condition for the set of Rips words $\mathcal{X} \cup \mathcal{Y}$.* Every piece has length strictly less than a quarter of the length of each element of $\mathcal{C}(\mathcal{X} \cup \mathcal{Y})$ in which it occurs.
iv. *The $C(5)$-condition for $\mathcal{U} = \{ u, v \mid t^{-1}utv^{-1} \in \mathcal{R} \}$.* No element of $C(\mathcal{U})$ is a concatenation of fewer than 5 pieces.

This can be achieved for instance by adapting the example of [Wis03, Remark 3.2] so that $\mathcal{X}$ is the set of words

$$X_i \;:=\; x_1 x_2^{200ip} x_1 x_2^{200ip+1} \cdots x_1 x_2^{200ip+200p-1}$$

for $1 \leq i \leq 14p$ and $\mathcal{Y}$ is the set of words

$$Y_i \;:=\; y_1 y_2^{200ip} y_1 y_2^{200ip+1} \cdots y_1 y_2^{200ip+200p-1}$$

for $1 \leq i \leq 30$. Then $\mathcal{R}$ satisfies $C'(1/6)$ because the longest pieces in $\mathcal{R}$ have the form $x_2^{\alpha-1} x_1 x_2^{\alpha}$ or $y_2^{\alpha-1} y_1 y_2^{\alpha}$ (or the inverse thereof) for some $\alpha \in \mathbb{N}$. The longest piece appears either in X_{14p} with $\alpha = 200(14p) + 200p - 2$ or in Y_{30} with $\alpha = 200(30) + 200p - 2$. Its length is 2α, which (in either case, since $p > 1$) is strictly less than $12,400p$. On the other hand, the shortest defining relator has length at least $2|X_1|$ (see Figure 4.1) which is certainly bigger than $80,000p^2$, and this number is already bigger than six times $12,400p$. Conditions ii–iv hold similarly.

Condition i is used in the next paragraph and will be used to achieve $\mathrm{CAT}(-1)$ in Remark 15.6. Condition ii will be used in Lemma 5.1 towards establishing HNN-structures for G. Condition iii will restrict cancellation in Section 8, where we prove a lower bound on distortion, and in Sections 5 and 9, towards showing certain configurations of tracks do not arise in reduced diagrams. Condition iv achieves residual finiteness as we now explain.

All $C'(1/6)$ groups satisfy a linear isoperimetric inequality and so are hyperbolic [Ger99]. By [Wis04a] they are cubical, and then, by [Ago13], they are virtually special, and so are residually finite. Their residually finiteness is more directly apparent via [Wis03, Theorem 2.1], given the $C(5)$-condition for $\mathcal{U}$.

Our distorted subgroup is

$$H \;=\; \langle t, y_1, y_2 \rangle.$$

5. Consequences of small-cancellation

Here we give three lemmas that are proximate consequences of the small-cancellation conditions in Section 4.

Part (1) of the first of these lemmas will be used in our proof of Proposition 7.4. Part (2) will imply Proposition 7.1. We prove it using the $C(3)$-condition for $\mathcal{U}$, which is weaker than the $C(5)$-condition we have for $\mathcal{U}$ in Section 4. It is a special case of [Wis01, Theorem 2.11], but we include our own proof here because the result is central to our argument and the following short argument is available in our context.

Lemma 5.1. *(Cf.* [Wis01, Theorem 2.11]*)*

(1) Let S be the union of the 3- and 5-element generating sets of the terminal vertex groups of Table 7.1 (that is, S is the set of all words appearing in the final column). Then S freely generates a free subgroup of the free group $F = F(\mathcal{A})$, where $\mathcal{A} = \{a_1, a_2, t, x_1, x_2, y_1, y_2\}$.

(2) The set

$$\mathcal{U} \;=\; \left\{\, u, v \;\mid\; t^{-1}utv^{-1} \in \mathcal{R} \,\right\}$$

freely generates a free subgroup in the free group

$$F \;=\; F(a_1, a_2, b_0, \ldots, b_p, x_1, x_2, y_1, y_2).$$

Proof. Both parts are instances of the same general result, which we will prove here in the notation of (1). Suppose $w_1, \ldots, w_m \in S^{\pm 1}$ are such that $W = w_1 \cdots w_m$ is a non-empty reduced word on S but W freely reduces to the empty word when viewed as a word on the generators of F. We will show that the existence of this W contradicts $C(3)$.

There is a planar tree T whose edges are directed and are labelled by generators of F so that around the perimeter of T we read W. As each w_i is a reduced word on $\mathcal{A}$, the portion of the perimeter of T along which one reads w_i can only include a leaf of T at its start or end. It follows that if T is a line, then the shorter of w_1 and w_m^{-1} is subword of the other, and so is a piece, contrary to $C(3)$.

Assume, then, that T is not a line. There must be a pair of leaves v_1 and v_2 in T such that the geodesic ρ from v_1 to v_2 visits exactly one branching (i.e. valence at least 3) vertex b. So the word u one reads along ρ is $w_j \cdots w_k$ for some $1 \leq j \leq k \leq m$. In the remainder of our argument, read indices modulo m. The portion of ρ along which we read w_j must pass b else whichever of w_{j-1} and w_j is shorter would be a piece. And, in fact, then w_j must be u, else w_k or w_{k+1} would be a piece. So $j = k$. But then, as neither w_{j-1}^{-1} nor w_{k+1}^{-1} can be a subword of w_j (else they would be pieces), w_j must be concatenation of two pieces: one that it shares with w_{j-1}^{-1} and one that it shares with w_{k+1}^{-1}. Again, this is contrary to $C(3)$. $\square$

In our next lemma, a stronger small-cancellation hypothesis allows the same conclusion for further subsets of free groups. We will call on it in Lemma 7.6 en route to our proof of Proposition 7.4.

Lemma 5.2. *Suppose $Z_1, Z_2, Z_3, Z_1', Z_2', Z_3', Z_{p1}, Z_{p2}, Z_{p3}, Z_{p4}, Z_{p5}$ are words of the form $Y_* t^{-1} Y_* t Y_*$ or $Y_* t Y_*$ and each is a subword of a different defining relation from Figure 4.1 (so no Y_* appears twice). We will refer to these as Z-words. Then*

$$\mathcal{S}_1 \;=\; \left\{t, x_1, x_2, Z_1, Z_2, Z_3, Z_1', Z_2', Z_3'\right\}$$

freely generate a free subgroup of $F = F(t, x_1, x_2, y_1, y_2)$. The same is true of

$$\mathcal{S}_2 \;=\; \left\{Z_1, Z_2, Z_3, Z_1', Z_2', Z_3', Z_{p1}, Z_{p2}, Z_{p3}, Z_{p4}, Z_{p5}\right\}.$$

Proof. Suppose for a contradiction that w is a reduced word on $\mathcal{S}_1$ or $\mathcal{S}_2$ that represents the identity in F and includes at least one of the Z-words. Express each Y_* as the concatenation $P_* S_*$ of a prefix and a suffix whose lengths differ by at most one.

Consider a first $P_*^{\pm 1}$ or $S_*^{\pm 1}$ that is completely cancelled away on freely reducing w in F by removing successive inverse pairs of adjacent letters. It must have cancelled into a neighbouring $P_*^{\pm 1}$ or $S_*^{\pm 1}$. But then, because of the $C'(1/4)$-condition on the set of Rips words $\mathcal{X} \cup \mathcal{Y}$, some neighbouring pair of Z-words are inverses, contrary to w being reduced as a word on $\mathcal{S}_1$ or $\mathcal{S}_2$. $\square$

We will use the following variation on Lemma 5.2 in our proof of Lemma 9.3.

Lemma 5.3. *Suppose*

$$\overline{v} \;=\; x_{\lambda_0}^{\epsilon_0} X_{\xi_1}^{\mu_1} x_{\lambda_1}^{\epsilon_1} \cdots X_{\xi_m}^{\mu_m} x_{\lambda_m}^{\epsilon_m}$$

is a word on $\mathcal{X} \cup \{x_1, x_2\}$ in which $m \geq 1$, each $X_i \in \mathcal{X}$, each $\lambda_i \in \{1, 2\}$, each $\mu_i \in \{\pm 1\}$, and each $\epsilon_i \in \{0, \pm 1\}$. If $\overline{v}$ freely equals the empty word in $F(x_1, x_2)$, then for any sequence Σ of free-reduction moves (successive removals of $x_j^{\pm 1} x_j^{\mp 1}$ subwords) that takes $\overline{v}$ to the empty word, there is some i such that a subword consisting of at least a quarter of the letters of $X_{\xi_i}^{\mu_i}$ cancels with subword consisting of at least a quarter of the letters of $X_{\xi_{i+1}}^{\mu_{i+1}}$.

Proof. Express each word $X_{\xi_i}^{\mu_i}$ as the concatenation $P_i S_i$ of a prefix and a suffix whose lengths differ by at most one. Let i be the index of a first P_i or S_i to be completely cancelled away in the course of Σ. Assume it is S_i. (The argument for P_i is essentially the same.) Then S_i cancels with a prefix of $x_{\lambda_i}^{\epsilon_i} X_{\xi_{i+1}}^{\mu_{i+1}}$. But then, $C'(1/4)$ and the fact that the X_* all have length at least 100 together imply the result. $\square$

6. Van Kampen diagrams, corridors, and tracks

Suppose w is a word on the generators of a group which is given by a presentation. A *van Kampen diagram* for w with respect to that presentation is a finite planar 2-complex in which every edge is directed and labelled by a generator in such a way that around the perimeter of the diagram (in some direction from some starting vertex) one reads w and around the perimeter of each 2-cell (in some direction from some starting vertex) one reads a defining relator. A word w admits a van Kampen diagram if and only if it represents the identity in the group. Many introductory texts discuss van Kampen diagrams—e.g., [BH99].

Definition 6.1. (Reduced diagrams) A van Kampen diagram is *reduced* when it does not contain a pair of back-to-back cancelling cells—that is, a pair of cells with a common edge e such that the word read clockwise around the perimeter of one of these cells starting from e is the same as that read anticlockwise around the other starting from e.

Definition 6.2. (Corridors) Suppose z is a generator. Suppose $C_1, \ldots, C_m$ is a maximal set of distinct 2-cells in a van Kampen diagram Δ such that for all i, around ∂C_i one reads a word $u_i z v_i^{-1} z^{-1}$ and the z in ∂C_i is the z^{-1} in ∂C_{i+1}. Then the $C_1, \ldots, C_m$ concatenate in Δ to form an *z-corridor* $\mathcal{C}$, as shown in Figure 6.1. A z-edge in $\partial\Delta$ that is not part of the boundary of a 2-cell is a corridor with no 2-cells.

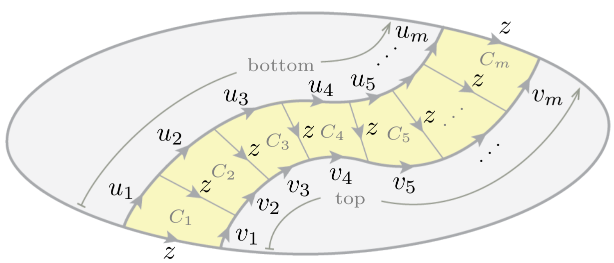


Figure 6.1. A corridor in a van Kampen diagram.

An assumption commonly made when defining corridors is that every defining relator containing a z or z^{-1}, contains exactly one z and one z^{-1}. Then z-corridors cannot cross or self-intersect, and each one either connects a pair of z-edges on $\partial\Delta$ or closes up to form a *z-annulus*. In our presentation $\mathcal{P}$ for G this assumption is met by the letters a_1, b_0, and t, but not, for example, by

a_2, b_1, …, or b_p: an a_2-corridor can terminate at an $r_{1,q-1}$-cell and a b_i-corridor, for $i \neq 0$, can terminate at an $r_{1,i-1}$-cell.

The *words along the top and bottom* of $\mathcal{C}$ are $v_1 \cdots v_m$ and $u_1 \cdots u_m$, respectively.

We will reframe and generalize the definition of a corridor via the dual of a van Kampen diagram. Let Δ^+ be Δ with one additional 2-cell e_∞ "*at infinity*" attached along its boundary cycle. So Δ^+ is homeomorphic to a 2-sphere. Let $\mathcal{G}^+$ be the 1-skeleton of the 2-complex dual to Δ^+. Let $\mathcal{G}$ be the graph obtained from $\mathcal{G}^+$ by removing the interior of e_∞. So the vertex dual to e_∞ is absent from $\mathcal{G}$ and instead $\mathcal{G}$ has a vertex in the middle of every edge in $\partial e_\infty = \partial\Delta$.

While the following definition could be presented in more general terms, we prefer to specialize to van Kampen diagrams Δ over our presentation $\mathcal{P}$ for G.

Definition 6.3. (Tracks, subtracks, and compound tracks) An *a- or b-edge* in a van Kampen diagram Δ over $\mathcal{P}$ is an edge labelled by a_i or b_i, respectively, for some i. An *s-subtrack* is a path $\rho : [0, k] \to \mathcal{G}$, where $k > 0$ is an integer, with the following properties:

(1) For each integer i in $[0, k-1]$, the image $\rho([i, i+1])$ is an edge of $\mathcal{G}$ dual to an s-edge of Δ.
(2) All s-edges of Δ dual to ρ are oriented the same way as one travels along ρ (i.e., cross ρ all right-to-left or all left-to-right).
(3) The map ρ is injective on $(0, k)$.

An *s-track* is an s-subtrack that is maximal—i.e., it cannot be extended to a longer path with properties (1)–(3). For $s = a_1, b_0, \ldots, b_p, t$ an s-track traverses the 2-cells of an s-corridor. When s is a or b, it gives a more general notion. Figures 3.1, 3.3 and 3.4 show examples of tracks. As seen in these figures, a- or b-tracks could merge. We impose a smoothness condition on these merges, which we now discuss.

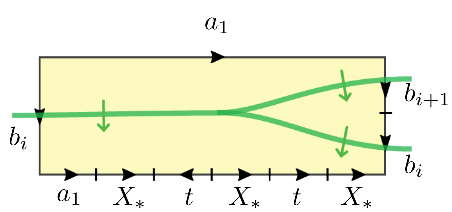


Figure 6.2. A train-track junction.

Let $\mathcal{G}_a$ and $\mathcal{G}_b$ be the subgraphs of $\mathcal{G}$ made up of all edges dual to a- and b-edges, respectively. We give $\mathcal{G}_a$ and $\mathcal{G}_b$ "train-track" structures by rendering some paths in them smooth and others not. As the defining relators in $\mathcal{P}$ each have zero, two or three b-letters, the valence-1 vertices of $\mathcal{G}_b$ are precisely those in the interior of e_∞. The valence-2 vertices are those dual to 2-cells of Δ that have (for some i) one b_i and one b_i^{-1} in their boundary word. We term the valence-3 vertices *junctions*. They are the vertices dual to 2-cells of Δ that have (for some i) one b_{i+1}, one b_i, and one b_i^{-1} in its boundary word. Paths γ in $\mathcal{G}_b$ can only fail to be smooth at junctions: per Figure 6.2 we make γ smooth at a junction if and only if the orientations of the b-edges it crosses before and after v agree. So a b-track is a maximal path $\rho : [0, k] \to \mathcal{G}_b$ that is injective and smooth on $(0, k)$. We will see below that if ρ closes up, then ρ must in fact be a smooth map of a circle into $\mathcal{G}$. Corresponding statements apply to $\mathcal{G}_a$.

Figure 6.3 shows how we consider a-, b-, and t-tracks to intersect when they traverse the same 2-cell.

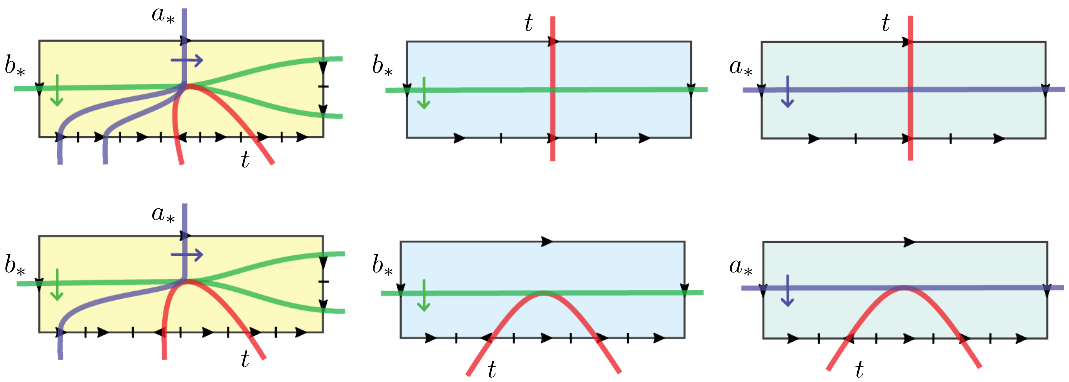


FIGURE 6.3. How a-tracks, b-tracks, and t-tracks intersect in a 2-cell. In the first, second, fourth and sixth cases, the t-track through the cell touches but does not cross the other tracks.

A *compound track* is a concatenation of a-, b-, and t-subtracks (the orientations of which are not required to agree). The corridor or annulus associated to a (compound) track ρ in a van Kampen diagram Δ is the subcomplex made up of all the 2-cells through which ρ passes. There are words along its top and bottom as for a standard corridor as explained above.

We will see in Section 9 that the hypothesis that a van Kampen diagram Δ over $\mathcal{P}$ is reduced significantly restricts the behaviours of its tracks. Then in Section 11 the tracks are yet more sharply restricted in diagrams pertinent to establishing upper bounds on the distortion of H in G. Here is a first observation in that direction.

Lemma 6.4. (No teardrops) *An s-track cannot be a teardrop—i.e., if $\rho : [0, k] \to \mathcal{G}$ is an s-track with $\rho(0) = \rho(k)$, then ρ induces a smooth map from S^1 to $\mathcal{G}$.*

Proof. Were the image of ρ a teardrop, the point $\rho(0) = \rho(k)$ would be a junction. However, as all the s-edges along an s-track are oriented the same way (in this case, either into or out of the teardrop) this would violate the orientation condition at the junction; see Figure 6.2 ☐

Definition 6.5. (Tracks forming loops) A track that closes up is a *loop*. In light of Lemma 6.4, a track closes up without introducing a corner, and so loops are smooth.

7. HNN-STRUCTURES FOR G

We will give two HNN-structures for G. The first is an immediate consequence of Lemma 5.1(2).

Proposition 7.1. *G is an HNN-extension:*

$$G \;=\; F \mathop{\ast}_{t} \quad \text{where} \quad F = F(a_1, a_2, b_0, \ldots, b_p, x_1, x_2, y_1, y_2)$$

and the $r = 5p + 11$ defining relators displayed in Figure 4.1 dictate the associated isomorphism between the vertex groups, both of which are rank-r free subgroups of F.

We will call on the following corollary in our proof of Lemma 9.14. It holds because the elements of $\mathcal{U}$ are *reduced* words with no t-letters.

Corollary 7.2. *Non-trivial subwords of elements of $\mathcal{U}$ represent non-identity elements in G.*

We will learn later (in Corollary 9.17) that F is undistorted in G, and it will follow that the same is true of the two vertex subgroups.

Our second HNN-structure for G is:

$$G = \left(\cdots \left(\left(F(t, x_1, x_2, y_1, y_2) \underset{a_1,a_2}{*} \right) \underset{b_p}{*} \right) \cdots \right) \underset{b_0}{*}$$

in the manner detailed in Proposition 7.4 and Table 7.1 below.

We use the notation $K *_{s_1,\ldots,s_l}$ to denote an l-fold HNN-extension with vertex group K, stable letters $s_1, \ldots, s_l$ and subgroups $I_i, T_i < K$ for $i = 1, \ldots, l$, such that $s_i^{-1} I_i s_i = T_i$. We call I_i and T_i the *initial* and *terminal* groups respectively, and say that the stable letter s_i *conjugates* I_i to T_i.

Definition 7.3. Let F be the free group on $\{t, x_1, x_2, y_1, y_2\}$. Note that this is a departure from our definition of F in Proposition 7.1. Let G_{-1} be the group generated by $\{t, x_1, x_2, y_1, y_2, a_1, a_2\}$ subject to the two $r_{4,*}$- and four $r_{4,*,*}$-defining-relators of Figure 4.1. Then, for $i = 0, \ldots, p$, define G_i to be the group generated by $\{t, x_1, x_2, y_1, y_2, a_1, a_2, b_{p-i}, \ldots, b_p\}$ subject to all the relators of Figure 4.1 in which only these letters appear. In particular, $G = G_p$.

We will establish that $G_{-1} = F *_{a_1,a_2}$ and $G_i = G_{i-1} *_{b_{p-i}}$ for $i \geq 0$, where the initial and terminal groups at each stage are as shown in Table 7.1. More precisely:

Proposition 7.4. *For $G_{-1}, G_0, \ldots, G_p$ as per Definition 7.3:*

(1) G_{-1} is a double HNN-extension over F with stable letters a_1 and a_2 conjugating the initial group $\langle t, y_1, y_2 \rangle$ to the first and second terminal groups listed in Row 1 of Table 7.1, respectively.

(2) For $i \geq 0$, the group G_i is an HNN-extension over G_{i-1} with stable letter b_{p-i} conjugating the group $K_i < G_{i-1}$ from Table 7.1 to the group $L_i < G_{i-1}$ from Table 7.1.

Recall that, per Section 4, we have chosen to suppress the indexing in our notation for the small cancellation words appearing in our construction. Thus, in Table 7.1, different instances of X_* or Y_* represent different small cancellation words, and the collection $\mathcal{X} \cup \mathcal{Y}$ of all such words satisfies $C'(1/4)$.

Before we prove Proposition 7.4, we observe that it yields:

Corollary 7.5. *The subgroup $H = \langle t, y_1, y_2 \rangle$ of G is a free group of rank 3.*

Proof. Since F is free on t, x_1, x_2, y_1, y_2, it is clear that $\langle t, y_1, y_2 \rangle$ is rank-3 free in F. As vertex groups inject into HNN-extensions, Proposition 7.4 yields: $H \hookrightarrow F \hookrightarrow G_{-1} \hookrightarrow G_0 \hookrightarrow \cdots \hookrightarrow G_p = G$. □

Proof of Proposition 7.4(1). The group $\langle t, y_1, y_2 \rangle < F$ is free of rank 3. The two terminal vertex groups in the G_{-1} row of Table 7.1 are free of rank 3 by Lemma 5.1(1). Thus the described HNN-structure follows from the definition of G_{-1}. □

To establish the HNN-structure of G_i for $i \geq 0$ (thereby completing the proof of Proposition 7.4(2)), we must show that the groups K_i and L_i listed in Table 7.1 are free of rank 5 in G_{i-1}. As a first step, we show:

Lemma 7.6. *The groups K_0 and L_i for $i = 0, \ldots, p$ are rank-5 free subgroups of G_{-1}.*

Table 7.1. Iterated HNN structure of G

Group	Stable letter(s)	Vertex group	Initial group	Terminal group(s)
G_{-1}	a_1, a_2	F	$\langle t, y_1, y_2\rangle$	$\langle Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*\rangle$, $\langle Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*\rangle$
G_i $i \geq 0$	b_{p-i}	G_{i-1}	$K_0 = \langle a_1, a_2, t, x_1, x_2\rangle$ $K_i = \langle a_1 b_{p-i+1}, a_2, t, x_1, x_2\rangle$ $i > 0$	$L_i = \langle a_1 X_* t^{-1} X_* t X_*,\ a_2 X_* t^{-1} X_* t X_*,\ X_* t X_*,\ X_* t^{-1} X_* t X_*,\ X_* t^{-1} X_* t X_*\rangle$ $i \neq p,\ p-q+1$ $L_{p-q+1} = \langle a_1 a_2 X_* t^{-1} X_* t X_*,\ a_2 X_* t^{-1} X_* t X_*,\ X_* t X_*,\ X_* t^{-1} X_* t X_*,\ X_* t^{-1} X_* t X_*\rangle$ $L_p = \langle a_1 Y_* t^{-1} Y_* t Y_*,\ a_2 Y_* t^{-1} Y_* t Y_*,\ Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*,\ Y_* t^{-1} Y_* t Y_*\rangle$

Proof. We begin with K_0. If a_1, a_2, t, x_1, x_2 do not generate a free subgroup of G_{-1}, then there is a non-empty freely reduced word on these letters which represents the identity in G_{-1}. Let w be the shortest such word and let Δ be a reduced van Kampen diagram with boundary label w.

Observe that the group F injects into G_{-1}, as it is the vertex group in the HNN-structure for G_{-1}, by Proposition 7.4(1). Thus $\langle t, x_1, x_2\rangle < F < G_{-1}$ is free, and so no non-empty freely reduced word on these letters represents the identity. Thus we may assume that w has at least one a_1- or a_2-letter, and so Δ has at least one a_1- or a_2-corridor. Moreover, we can assume Δ is homeomorphic to a 2-disc, because otherwise it could be broken into two subdiagrams for two words which are shorter than w and represent the identity, and cannot both be freely reduced to the empty word (since w cannot be). In particular, every a_1- and a_2-corridor is *non-degenerate*, by which we mean that it is not a single a_1- or a_2-edge that is part of a 1-dimensional portion of Δ.

Let $\langle Z_1, Z_2, Z_3\rangle$ and $\langle Z_1', Z_2', Z_3'\rangle$ denote the two terminal groups in the construction of G_{-1} as shown in Table 7.1. No two a_1- or a_2-corridors can cross or branch in Δ, so dual to them there is an oriented tree $\mathcal{T}$ which has a vertex for each complimentary region and an edge for each corridor. Give the edges of $\mathcal{T}$ orientations that match the directions of the a_1- or a_2-corridors they cross. Then $\mathcal{T}$ necessarily has a sink vertex (a vertex with the property that all its incident edges are oriented towards it), and the boundary of the subdiagram Δ_0 of Δ corresponding to this vertex consists of parts of $\partial\Delta$ between a_1- or a_2-edges at the ends of corridors and the top boundaries of a_1- or a_2-corridors. Thus, read around $\partial\Delta_0$ is a word v on

$$t, x_1, x_2, Z_1, Z_2, Z_3, Z_1', Z_2', Z_3'.$$

By Lemma 5.2 these elements form a basis for a free subgroup F' of F and therefore of G_{-1}. Now v is non-empty (since every corridor is non-degenerate) and represents the identity in G_{-1}, and therefore in the free group F' (since $F' \hookrightarrow G_{-1}$). So v is not freely reduced, i.e., it has a subword of the form uu^{-1} for some letter or inverse letter u. Since the subwords of v on t, x_1, x_2 come from w, which is freely reduced, u is one of the remaining generators of F'. Then uu^{-1} must be a subword of the top boundary of a single a_1- or a_2-corridor (because, if u and u^{-1} came from different corridors, w would have a subword $a_1^{\pm 1}a_1^{\mp 1}$ or $a_2^{\pm 1}a_2^{\mp 1}$, contradicting the fact that it is freely reduced). This means the corridor has adjacent cells that are identical and oppositely oriented, contradicting the fact that Δ is reduced. Thus K_0 is a free subgroup of G_{-1}.

A near identical proof shows that $L_p < G_{-1}$ is free. Denoting the generators of L_p by

$$a_1 Z_{p1}, a_2 Z_{p2}, Z_{p3}, Z_{p4}, Z_{p5},$$

let w be the shortest non-empty freely reduced word on these generators which represents the identity in G_{-1}. Let Δ be a reduced van Kampen diagram over G_{-1} with boundary label w. Since $\langle Z_{p3}, Z_{p4}, Z_{p5}\rangle < F < G_{-1}$ is free (using Lemma 5.1(1)), we may assume as before that w has at least one $a_1 Z_{p1}$ or $a_2 Z_{p2}$. Hence Δ has at least one a_1- or a_2-corridor. Furthermore, we conclude as before that all a_1- or a_2-corridors are non-degenerate. Considering a sink region of the oriented dual tree as above, we see that the boundary label of the sink region is a word v on

$$Z_1, Z_2, Z_3, Z_1', Z_2', Z_3', Z_{p1}, Z_{p2}, Z_{p3}, Z_{p4}, Z_{p5}$$

which represents the identity in G_{-1}. (The first six of these words appear along top boundaries of a-corridors while the last five appear in parts of v coming from w.) By Lemma 5.2, these elements form a basis for a free subgroup of F, and therefore of G_{-1} (since $F \hookrightarrow G_{-1}$). Then we argue as in the previous paragraph to arrive at a contradiction.

Finally, for $i \neq p$, Lemma 5.1(1) implies that L_i is a rank-5 free subgroup of K_0. Thus L_i is a rank-5 free subgroup of G_{-1} as $K_0 \hookrightarrow G_{-1}$. □

In order to prove that K_i is free for $i > 0$ and complete the proof of Proposition 7.4 we need three technical lemmas.

Lemma 7.7. *In G_i of Definition 7.3, $b_p, b_{p-1}, \ldots, b_{p-i}$ freely generate a free subgroup.*

Proof. By examining the relators of G_i, we see that there is a quotient homomorphism

$$G_i \twoheadrightarrow Q_i = \langle\, b_p, b_{p-1}, \ldots, b_{p-i}, a \mid a_1^{-1} b_j a_1 = b_{j+1} b_j \text{ for } j < p;\ a_1^{-1} b_p a_1 = b_p \,\rangle$$

mapping $b_j \mapsto b_j$, $a_1 \mapsto a_1$ and killing every other generator. This quotient Q_i is free-by-cyclic: the generator a of the cyclic part acts by conjugation on a free group generated by $b_p, \ldots, b_{p-i}$ by an automorphism. Moreover, the restriction of this homomorphism to the subgroup $\langle b_p, \ldots, b_{p-i}\rangle < G_i$ is a surjection onto the rank-$(i+1)$ free subgroup $\langle b_p, \ldots, b_{p-i}\rangle < Q$. The result follows. □

The next lemma restricts the possible b-track systems in certain van Kampen diagrams over G_i.

Lemma 7.8. *For $i = 0, \ldots p-1$, let Δ be a reduced van Kampen diagram over the group G_i of Definition 7.3 with boundary labelled by a word on $a_1, a_2, t, x_1, x_2, b_p, b_{p-1}, \ldots, b_{p-i}$. Then*

(1) Δ has no $r_{4,*,*}$- or $r_{4,*}$-cells (per Figure 4.1).
(2) Δ has no a_1-annuli.
(3) If the word read around $\partial\Delta$ contains no letters $a_1^{\pm 1}$, then the track system $\mathcal{G}_b$ of Δ has no junctions. Thus $\mathcal{G}_b$ consists of a collection of disjoint tracks, each dual to a b_j-corridor for some j such that $0 < p - i \leq j \leq p$.

Proof. For (1), we suppose Δ_0 is a maximal subdiagram of Δ that contains no b-edges and is homeomorphic to a 2-disc. Any $r_{4,*,*}$- or $r_{4,*}$-cell must be in some such Δ_0. All its 2-cells must be of type $r_{4,*,*}$ or $r_{4,*}$ since every other type of 2-cell has a b-edge. So, arguing that there are no 2-cells in Δ_0 will establish (1).

There can be no y-edges in $\partial\Delta_0$ because such a y-edge would have to be either in $\partial\Delta$ (contrary to hypothesis) or in the boundary of a 2-cell of Δ that is not of type $r_{4,*,*}$- or $r_{4,*}$ (impossible because the only such 2-cells from Figure 4.1 have b_0-edges, and $b_0 \notin G_i$ when $i < p$). So $\partial\Delta_0$ is labelled by a word v on a_1, a_2, t. Now, v represents the identity in

$$\langle a_1, a_2, t, y_1, y_2 \mid r_{4,i,j}, r_{4,i};\ i,j = 1,2 \rangle \ = \ F(a_1, a_2, y_1, y_2) \underset{t}{\ast} .$$

There can be no t-annulus in Δ_0 since the word read around the inner boundary of an innermost t-annulus would be a word on $\mathcal{U}$ that freely equals the empty word, and Lemma 5.1(2) would imply that there must be cancellation of a pair of 2-cells, contrary to Δ being reduced. And if there is a t-corridor in Δ_0, then there is one that is outermost in that the freely reduced form of the word along its top or bottom follows a path in $\partial\Delta_0$. But (since Δ_0 is reduced and homeomorphic to a 2-disc) the word along the top or bottom any t-corridor in Δ_0 must contain y-letters, so this contradicts there being no y-letters in $\partial\Delta_0$.

Next we deduce (2). Were there such an a_1-annulus, in light of (1), one of its boundaries would be labelled by a word on $b_p, b_{p-1}, \ldots, b_{p-i}$ representing the identity in G_i. It would then follow from Lemma 7.7 that this word would freely reduce to the empty word. This would imply that the annulus would have adjacent 2-cells that are identical but with opposite orientation, contrary to Δ being reduced.

Finally, for (3), suppose the word read around $\partial\Delta$ contains no letters $a_1^{\pm 1}$. If the track system $\mathcal{G}_b$ had a junction, that junction would be in a 2-cell of Δ with an a_1 on its boundary, and this 2-cell would be part of an a_1-corridor or a_1-annulus. However, there are no a_1-corridors since the label of $\partial\Delta$ has no a_1 and there are no a_1-annuli by (2). $\square$

Lemma 7.9. *For $i = 0, \ldots p-1$, in the group G_i of Definition 7.3, we have*

$$\langle b_p, b_{p-1}, \ldots, b_{p-i} \rangle \cap \langle a_2, x_1, x_2, t \rangle \ = \ \{1\} .$$

Proof. Suppose there is a non-trivial element in $\langle b_p, b_{p-1}, \ldots, b_{p-i} \rangle \cap \langle a_2, x_1, x_2, t \rangle$. Then there are non-empty freely reduced words $u = u(a_2, x_1, x_2, t)$ and $v = v(b_p, b_{p-1}, \ldots, b_{p-i})$ such that $u = v$ in G_i, and there is a reduced van Kampen diagram Δ with boundary label uv^{-1}. Observe that Δ satisfies the hypotheses of Lemma 7.8(3) since the word read around $\partial\Delta$ has no instances of $a_1^{\pm 1}$. Thus the track system $\mathcal{G}_b$ of Δ consists of a union of disjoint tracks, each dual to a b_j-corridor for some j. Since u has no instances of b_j for any j, each of these tracks has both ends on the part of $\partial\Delta$ labelled v. Since these b-tracks cannot cross each other, there must be at least one that is innermost in that it begins and ends at consecutive letters in v. This implies that v has a subword $b_j^{\pm 1} b_j^{\mp 1}$, which contradicts v being freely reduced. $\square$

We can now prove the following lemma, which establishes Proposition 7.4(2).

Lemma 7.10. *For $i = 0, \ldots, p$,*

(1) the subgroups $K_i, L_i \leq G_{i-1}$ are free of rank 5,
(2) the group G_i is an HNN-extension over G_{i-1} with stable letter b_{p-i} conjugating K_i to L_i.

Proof. We induct on i. In the case $i = 0$, Lemma 7.6 gives (1), and then (2) follows by definition of G_0. We now prove the induction step. Assume the result holds up to some value of the index $i < p$. We will show that (1) and (2) hold with the index i elevated by 1.

In Lemma 7.6 we showed that L_{i+1} is a free subgroup of G_{-1} of rank 5. By statement (2) of the induction hypothesis, $G_{-1}, G_0, \ldots, G_i$ are successive HNN extensions. So $G_{-1} \hookrightarrow G_0 \hookrightarrow \cdots \hookrightarrow G_i$ are injective inclusions and L_{i+1} is a rank-5 free subgroup of G_i as well.

Likewise, K_0 is a rank-5 free subgroup of G_i. We will show that $K_{i+1} = \langle a_1 b_{p-i}, a_2, t, x_1, x_2 \rangle$ is also a rank-5 free subgroup of G_i. This will prove (1), and then (2) will immediately follow.

Let w be a non-empty freely reduced word on the generators of K_{i+1} such that $w = 1$ in G_i. Assume that w is minimal in the sense that no shorter non-empty freely reduced word on the generators of K_{i+1} represents the identity in G_i. Let Δ be a reduced van Kampen diagram for w over G_i. It contains no 2-cells of type $r_{4,*,*}$ or $r_{4,*}$ by Lemma 7.8(1).

The word w must include at least one instance of $a_1 b_{p-i}$, as otherwise w would be a non-empty freely reduced word representing the identity in the free group $K_0 < G_i$, a contradiction. Consequently, Δ has at least one a_1-corridor. Moreover, every a_1-corridor is non-degenerate, as a degenerate corridor would cut $\partial\Delta$ into two loops (both non-trivial as w is non-empty and freely reduced) and one of these would be labelled by a shorter freely reduced word on the generators of K_0, contradicting the minimality of w. As Δ has no 2-cells of type $r_{4,*,*}$ or $r_{4,*}$, every a_1-corridor is made up of $r_{1,i}$-cells, where $1 \leq i \leq p$. (We exclude $r_{1,0}$ since $i < p$.)

Let C be an innermost a_1-corridor in Δ, i.e. an a_1-corridor whose complement in Δ has a component Δ' without a_1-corridors. Then $\partial\Delta'$ is composed of two paths between the same pair of points: a top or bottom boundary γ of C with label v (which is non-empty since C is non-degenerate) and a path δ along $\partial\Delta$. The labels γ and δ represent the same element of G_i.

There are two cases, depending on the orientation of C. If C points away from Δ', then γ is its bottom boundary and v is a non-empty word on $b_{p-i}, \ldots, b_p$, which is freely reduced since Δ is reduced. In this case δ is labelled by a freely reduced word u on t, x_1, x_2, a_2, which is non-empty since otherwise w would have an $a_1^{-1} a_1$ subword and not be freely reduced. Now $u = v$ in G_i, which contradicts Lemma 7.9.

On the other hand, if C points towards Δ', then γ is its top boundary and v is a word on elements of the form $b_{j+1} b_j X_*^{-1} t^{-1} X_*^{-1} \epsilon^{-1}$, where $b_{j+1} = 1$ if $j = p$, and $\epsilon = a_2$ if $j = q - 1$ and 1 otherwise. In this case δ is labelled by a word of the form $b_{p-i} u b_{p-i}^{-1}$, where u is a word on t, x_1, x_2, a_2.

We consider the track system $\mathcal{G}_b'$ of Δ'. Lemma 7.8(3) applies to Δ', because it has no a_1-corridors, and we conclude that $\mathcal{G}_b'$ is a disjoint union of tracks. Each of these tracks is dual to a b_j-corridor for some j such that $0 < p - j \leq j \leq p$ and inherits its label.

Suppose there exists a b-track with both ends on γ. Consider an innermost such track, i.e. one for which the subword of v between its endpoints has no b-letters, and suppose it is labelled b_m for some m. Since each 2-cell of C has at least one b-letter and at most one b_m, this track must begin and end at neighboring cells of C. Examining the $r_{1,*}$-cells of Figure 4.1 we see that the only possibility is that these are identical cells with opposite orientation, which contradicts Δ being reduced. Thus tracks of $\mathcal{G}_b'$ have at most one end on γ.

Since δ is labelled by $b_{p-i} u b_{p-i}^{-1}$, where u has no b-letters, there are at most two tracks ending on δ. Since C is non-degenerate, there is at least one track starting at γ, which rules out the possibility of a track with both endpoints on δ. We conclude that $\mathcal{G}_b'$ has exactly two tracks, each with one end on δ and one on γ, and both with label b_{p-i}. It follows that C has exactly two 2-cells, both of type $r_{1,p}$, and $i = 0$ (as every other possible 2-cell has both b_j and b_{j+1} for some j in its top boundary). Moreover, since tracks preserve orientation, and the two edges of δ labelled b_p are oppositely oriented, it follows that the two 2-cells of C are oppositely oriented. This contradicts Δ being reduced.

We have arrived at contradictions in all cases. It follows that no such w can exist, and that K_{i+1} is free of rank 5, completing the induction. □

8. The lower bound on distortion

In this section, we will establish the lower bound on distortion of Theorem 1.1 in the case $k = 1$. In the manner outlined by the figures in this section, we prove that for all $n \in \mathbb{N}$, there is a freely reduced word χ_n on $t^{\pm 1}$, $y_1^{\pm 1}$, and $y_2^{\pm 1}$ of length $\simeq 2^{n^p}$ which represents the same group element as a word w_n in the generators of G of length $\simeq n^q$. These length estimates emerge from calculations tracing through the construction, with small-cancellation arguments ensuring that χ_n does not lose too much length through free reduction. As t, y_1 and y_2 freely generate H (Corollary 7.5), no shorter word than χ_n on $t^{\pm 1}$, $y_1^{\pm 1}$ and $y_2^{\pm 1}$ equals w_n in G. Via Lemma 2.2, this will establish that $\mathrm{Dist}_H^G(n) \succeq 2^{n^{p/q}}$.

For w a word, $|w|$ denotes the number of letters in w and $|w|_x$ the exponent sum of the x in w. So, if w is a *positive* word, which is to say it contains no inverse letters, then $|w|_x$ is the number of x in w.

Recall that killing $a_2, t, x_1, x_2, y_1, y_2$ maps G onto the free-by-cyclic group

$$Q \ = \ \langle\, a_1, b_0, b_1, \ldots, b_p \ | \ a_1^{-1} b_j a_1 = \varphi(b_j) \text{ for all } j \,\rangle,$$

where φ is the automorphism of $F(b_0, \ldots, b_p)$ mapping $b_j \mapsto b_{j+1} b_j$ for $j = 0, \ldots, p-1$ and $b_p \mapsto b_p$. The following lemma describes a lift of an equality $a_1 \varphi(u b_0) = u b_0 a_1$ in Q to an equality $a_1 \varphi(u b_0) = u b_0 a_1 \tau$ in G.

Lemma 8.1. *Given a positive word $u = u(b_1, \ldots, b_p)$, there is a freely reduced word $\tau = \tau(a_2, t^{\pm 1}, y_1^{\pm 1}, y_2^{\pm 1})$ such that*

$$a_1 \varphi(u b_0) \ = \ u b_0 a_1 \tau \ \text{ in } G \tag{8.1}$$

$$|\varphi(u b_0)|_{b_i} \ = \ |u b_0|_{b_i} + |u b_0|_{b_{i-1}} \qquad \text{for } i = 1, \ldots, p \tag{8.2}$$

$$|\varphi(u b_0)|_{b_0} \ = \ |u b_0|_{b_0} \ = \ 1 \tag{8.3}$$

$$|\tau|_{a_2} \ = \ |\varphi(u b_0)|_{b_q} - |u b_0|_{b_q}. \tag{8.4}$$

Moreover, τ has as a suffix κ that is also a long *suffix of one of the Rips words Y_* used in the presentation $\mathcal{P}$ of G—by* long *we mean that $|\kappa|$ is at least* $(3/4)|Y_*|$.

Proof. The statements (8.1)–(8.4) are easily verified when u is empty. Assuming $|u| \geq 1$, express u as $b_i u_0$ where b_i is the first letter of u and u_0 is the remainder of the word. The structure of a van Kampen diagram for (8.1) is displayed in Figure 8.1. It is constructed inductively, the base step being provided by the case where u is empty. The top cell in Figure 8.1 encodes the relation $a_1 \varphi(b_i) = b_i a_1 \sigma$, where σ is a word on a_2, t, x_1 and x_2 that contains no a_2^{-1}. The bottom left block comes from applying the induction hypothesis to u_0, so $\tau_0 = \tau_0(a_2, t^{\pm 1}, y_1^{\pm 1}, y_2^{\pm 1})$. The bottom right block encodes the result of moving $\phi(u_0 b_0)$ past σ. That σ_0, and therefore τ, contains letters

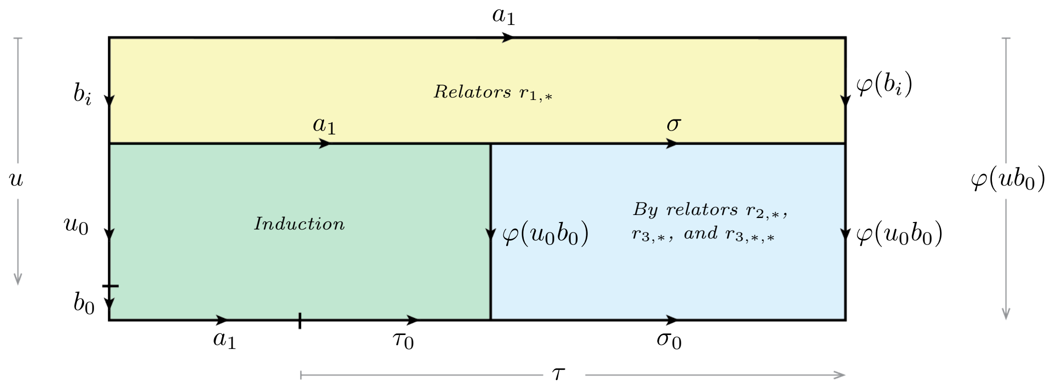


FIGURE 8.1. A diagram for $a_1\varphi(ub_0) = ub_0a_1\tau$ in G.

$a_2, t^{\pm1}, y_1^{\pm1}, y_2^{\pm1}$ but not $x_1^{\pm1}, x_2^{\pm1}$ is due to b_0 conjugating a_2, $t^{\pm1}$, $x_1^{\pm1}$ and $x_2^{\pm1}$ to words on a_2, $t^{\pm1}$, $y_1^{\pm1}$ and $y_2^{\pm1}$. (See the $r_{2,*}$-, $r_{3,*}$-, and $r_{3,*,*}$-cells of Figure 4.1.)

The equalities (8.2) and (8.3) follow from the definition of φ.

We get (8.4) by induction, as follows. Assume (8.4) holds for u_0. Examining the $r_{1,*}$-defining relators of Figure 4.1, we see that $|\sigma|_{a_2} = |\varphi(b_i)|_{b_q} - |b_i|_{b_q}$ for any i. Moreover, $|\sigma_0|_{a_2} = |\sigma|_{a_2}$ in the bottom right block of Figure 8.1 as each $r_{2,*}$-, $r_{3,*}$-, and $r_{3,*,*}$-defining relator of Figure 4.1 satisfies this property. Combining these observations with the induction hypothesis, we get: $|\tau|_{a_2} = |\tau_0|_{a_2} + |\sigma_0|_{a_2} = |\varphi(u_0b_0)|_{b_q} - |u_0b_0|_{b_q} + |\sigma|_{a_2} = |\varphi(u_0b_0)|_{b_q} - |u_0b_0|_{b_q} + |\varphi(b_i)|_{b_q} - |b_i|_{b_q} = |\varphi(b_iu_0b_0)|_{b_q} - |b_iu_0b_0|_{b_q}$, which completes the inductive step (since $u = b_iu_0$) and proves (8.4).

When u is empty, Figure 8.1 is a single $r_{1,0}$-cell and τ is $Y_*t^{-1}Y_*tY_*$, which satisfies the suffix condition by construction. For u non-empty we may assume by induction that τ_0 is reduced and its final letter is positive (since $\xi_\bullet$ or $\overline{\mu}_\bullet$ have positive final letters). Now σ is one of the subwords $X_*t^{-1}X_*tX_*$ of an $r_{1,*}$-defining relator of Figure 4.1 (as $Y_*t^{-1}Y_*tY_*$ is excluded since $b_i \neq b_0$). Thus σ has positive first letter and ends with x_1 or x_2. It follows, via the $C'(1/4)$-condition for $\mathcal{X} \cup \mathcal{Y}$ of Section 4, that the successive words we obtain from σ by conjugating by a b_i with $i \neq 0$ and then freely reducing have positive first letters and end with x_1 or x_2. Finally σ_0 is obtained by conjugating by b_0 and freely reducing, so it has a positive first letter and a suffix that is a *long* suffix of some $Y_*t^{-1}Y_*tY_*$ (again by $C'(1/4)$ for $\mathcal{X} \cup \mathcal{Y}$). Therefore there is no cancellation between τ_0 and σ_0, and so σ_0 gives τ the required *long* suffix. $\square$

For all $j \geq 0$, define u_j to be the positive word on $b_1, \ldots, b_q$ such that $u_jb_0 = \varphi^j(b_0)$ as words. In particular u_0 is the empty word ε, and $u_{j+1}b_0 = \varphi(u_jb_0)$. Now let $n \geq 1$. For $j = 0, \ldots, n-1$, let τ_{j+1} be as per Lemma 8.1 so that $a_1u_{j+1}b_0 = u_jb_0a_1\tau_{j+1}$ in G. Let $v_n = a_1\tau_1 \cdots a_1\tau_n$.

For our next lemma, we understand the binomial coefficient $\binom{n}{i}$ to be zero when $i > n$.

Lemma 8.2. *For all $n \geq 1$, the word v_n is freely reduced and*

$$a_1^n u_n b_0 \;=\; b_0 v_n \text{ in } G \tag{8.5}$$

$$|v_n|_{a_1} \;=\; n \tag{8.6}$$

$$|u_n b_0|_{b_i} \;=\; \binom{n}{i} \text{ for } i = 0, \ldots, p \tag{8.7}$$

$$|u_n b_0| \;=\; \binom{n}{0} + \cdots + \binom{n}{p}. \tag{8.8}$$

$$|v_n|_{a_2} \;=\; |u_n b_0|_{b_q} \;=\; \binom{n}{q} \tag{8.9}$$

Proof. The reason v_n is freely reduced is that each τ_i is freely reduced and contains no $a_1^{\pm 1}$ letters by Lemma 8.1. Then (8.5) holds as per Figure 8.2 and (8.6)–(8.9) all follow straightforwardly from Lemma 8.1. □

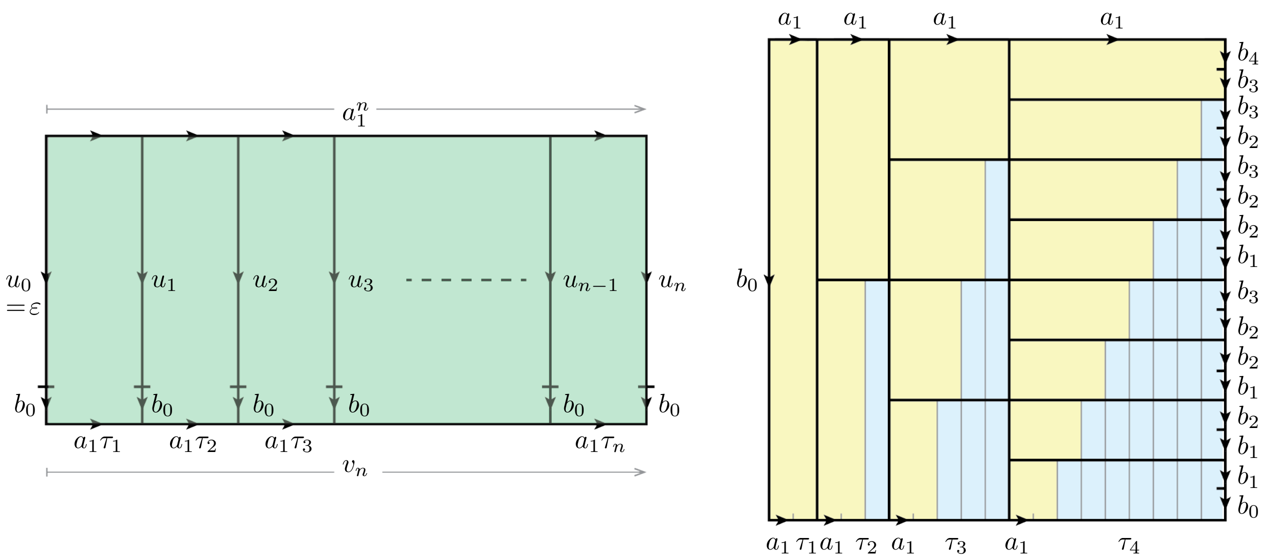


FIGURE 8.2. Why $a_1^n u_n b_0 = b_0 v_n$ in G. The diagram on the left is assembled from n instances of the diagram from Figure 8.1. That on the right shows it in finer detail in the case $n = 4$ and $q \geq 4$.

Let $\hat{v}_n$ be v_n with all $t^{\pm 1}$, $y_1^{\pm 1}$ and $y_2^{\pm 1}$ deleted.

Lemma 8.3. *For all $n \geq 1$, there is a freely reduced word $\mu_n = \mu_n(t^{\pm 1}, y_1^{\pm 1}, y_2^{\pm 1})$, whose final letter is positive, and such that*

$$v_n \;=\; \hat{v}_n \mu_n \text{ in } G. \tag{8.10}$$

Proof. Use the $r_{4,*,*}$- and $r_{4,*}$-defining relators of Figure 4.1 to shuffle the a_1 and a_2 through v_n to its start to make a prefix $\hat{v}_n$. In the process, the intervening letters $t^{\pm 1}, y_1^{\pm 1}, y_2^{\pm 1}$ become various $(Y_* t Y_*)^{\pm 1}$ and $(Y_* t^{-1} Y_* t Y_*)^{\pm 1}$.

By Lemma 8.1, τ_n, and therefore v_n, has a suffix κ that is a *long* suffix of some Y_*. The $Y_*^{\pm 1}$ that are created in the shuffling process are different from any that arise in Lemmas 8.1–8.3 (those lemmas do not use the relators $r_{4,*}$ or $r_{4,*,*}$). So, by $C'(1/4)$ for $\mathcal{X} \cup \mathcal{Y}$ (see Section 4), cancellation with these $Y_*^{\pm 1}$ cannot erode all of κ. So the final letter of μ_n is the final letter of κ, and so of some Y_*, and so is positive. □

Lemma 8.4. *There exists $K_1 > 1$ with the following property. For all $n \geq 1$, there is a reduced word Z_n on t, y_1, and y_2, whose first letter is positive, such that*

$$(u_n b_0)^{-1} x_1 \, u_n b_0 \; = \; Z_n \text{ in } G \tag{8.11}$$

$$K_1^{|u_n b_0|} \; \leq \; |Z_n|. \tag{8.12}$$

Proof. The word Z_n is the result of successive conjugations of x_1 by the letters of u_n (which are $b_1, \ldots, b_p$) and then by b_0. The relators $r_{3,*}$ and $r_{3,*,*}$ describe the effect: conjugation produces successive words on t and the X_* (so on t, x_1 and x_2) until the final conjugation by b_0, which results in a word on t and the Y_* (so on t, y_1 and y_2). In any one of these words, free reduction between adjacent $X_*^{\pm 1}$ (or adjacent $Y_*^{\pm 1}$) can only reduce the word's length by at most a half on account of the $C'(1/4)$ condition on $\mathcal{X} \cup \mathcal{Y}$ (see Section 4). So, if we take K_1 to be half the length of the shortest of the X_* and Y_*, then each conjugation increases reduced length by a factor of at least K_1. The $C'(1/4)$-condition for $\mathcal{X} \cup \mathcal{Y}$ also implies that free reduction cannot erode the first letter of the word at every stage, and as the initial x_1 is positive and so are first letters of each X_* and Y_*, it follows that the first letter of Z_n is positive. □

Lemma 8.5. *There exist $K_2 > 0$ and $K_3 > 1$ with the following properties. For all $n \geq 1$, the word*

$$w_n \; = \; \hat{v}_n^{-1} b_0^{-1} a_1^n x_1 a_1^{-n} b_0 \hat{v}_n$$

has length at most $K_2 n^q$ and equals in G a word $\chi_n = \chi_n(t^{\pm 1}, y_1^{\pm 1}, y_2^{\pm 1})$. Moreover, freely reducing χ_n gives a word of length at least $K_3^{(n^p)}$.

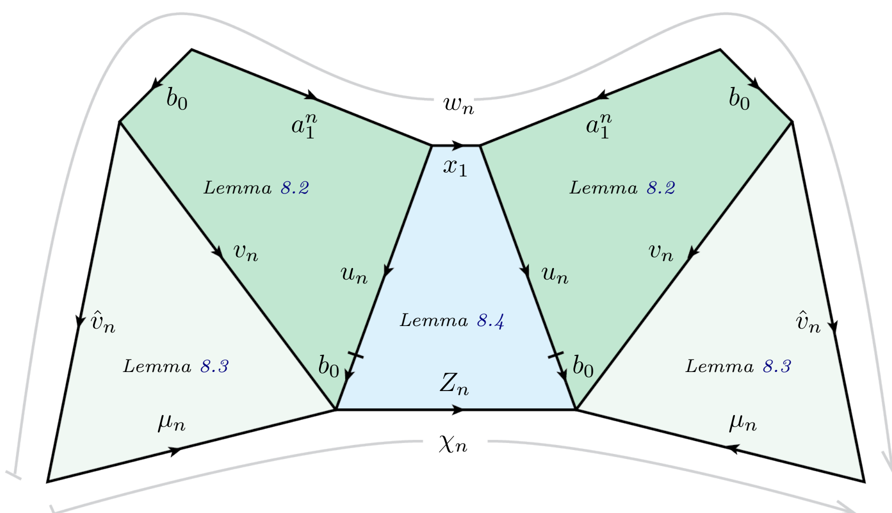


FIGURE 8.3. A diagram demonstrating that the word $w_n = \hat{v}_n^{-1} b_0^{-1} a_1^n x_1 a_1^{-n} b_0 \hat{v}_n$ on the generators of G and word $\chi_n = \mu_n Z_n \mu_n^{-1}$ on the generators of H represent the same element of G.

Proof. We have $|\hat{v}_n| = |\hat{v}_n|_{a_1} + |\hat{v}_n|_{a_2}$, which equals $|v_n|_{a_1} + |v_n|_{a_2} = n + \binom{n}{q}$ by (8.6) and (8.9). So $|w_n| = 2\binom{n}{q} + 2n + (2n+3)$, which is at most $K_2 n^q$ for a suitable constant $K_2 > 0$.

Figure 8.3 sets out why $\chi_n = \mu_n Z_n \mu_n^{-1}$ equals w_n in G. Consider freely reducing χ_n by freely reducing μ_n, Z_n, and μ_n^{-1}, and then performing all available cancellations where they meet. As the final letter of the freely reduced form of μ_n and the first letter of the freely reduced form of Z_n are both positive (by Lemmas 8.3 and 8.4), there is no cancellation between μ_n and Z_n. There may be cancellation between Z_n and μ_n^{-1} (indeed, a priori, all of Z_n could cancel into μ_n^{-1}). But for every letter of Z_n that cancels into μ_n^{-1}, there is a letter of μ_n that survives in the freely reduced form of χ_n. Therefore the length of the freely reduced form of χ_n is at least the length of the freely reduced form of Z_n. So the existence of a suitable $K_3 > 1$ follows from (8.12) and the fact that, by (8.8), $|u_n b_0|$ is a least a constant times n^p. □

9. Tracks in reduced van Kampen diagrams

As explained in Section 6, a van Kampen diagram is *reduced* when it does not contain a pair of back-to-back cancelling 2-cells. If a van Kampen diagram is reduced, then so are its subdiagrams. Here, we will explore the restrictions this hypothesis leads to on the arrangement of tracks in van Kampen diagrams over our presentation $\mathcal{P}$ for G of Section 4.

Definition 9.1. A *region* in a van Kampen diagram Δ is a closed subset that is homeomorphic to a 2-disc. We will consider regions that have boundary circuits comprised of portions of $\partial\Delta$, other paths in the 1-skeleton $\Delta^{(1)}$, and subtracks. Figure 9.1 shows two examples. Because tracks pass through the interiors of 2-cells, regions need not be subdiagrams. When we say a 1-cell or 2-cell of Δ is *in* R, we mean that it is a subset of R.

Before we give our first lemma, here is an overview of this section. Every 2-cell in a reduced van Kampen diagram Δ over $\mathcal{P}$ has some x- or y-letters (we call these "noise" letters) in its boundary word. We find it helpful to think of this noise to be *flowing though the diagram and expanding* in that, for the 2-cells to fit together, the adjacent cells must have more noise (in total), and those in the next layer further beyond those have yet more noise. This continues until the noise spills out into the boundary of the diagram.

Tracks in Δ mediate this flow of noise and provide a structure via which we can put this intuition on a firm foundation. All x-noise flows across b-tracks in the direction of their orientations, except that on crossing a b_0-track, the noise is converted to y-noise. And y-noise flows across a-tracks in the direction of their orientations. So, when a region has boundary that prevents the escape of noise, that region cannot occur in a reduced diagram. Lemmas 9.3, 9.4 and 9.6 are results of this nature. As for t-tracks, they have noise on both sides and reflect the HNN-structure $G = F*_t$. Lemma 9.2 is a consequence. It exemplifies the following idea, which reappears in Lemma 9.9 in a more complicated guise. If a certain feature is present (in this case, a t-loop), then there is an innermost instance, but an innermost instance must include cancelling 2-cells, contrary to the hypothesis that the diagram is reduced.

Lemmas 9.13 and 9.14 dig further into the structure of t-corridors and provide groundwork for Lemmas 9.15 and 9.16, which detail circumstances in which tracks and corridors show diagrams to flare out towards a portion of their boundary. These results will let us (in Lemma 11.2) simplify diagrams that demonstrate distortion.

Lemma 9.2. *Reduced van Kampen diagrams Δ over $\mathcal{P}$ contain no t-loops.*

Proof. Were there a t-loop in Δ, there would be one with no t-loop in its interior. The 2-cells it traverses would form an annular corridor. Around its inner boundary we read a word which, viewed as a word on the generators of the appropriate vertex group of the HNN-structure $G = F*_t$ of Proposition 7.1, would freely equal the empty word. So some adjacent pair of those generators would cancel. As those generators uniquely determine the 2-cells along whose sides they are read, a pair of 2-cells in the annulus would cancel, contrary to the diagram being reduced. $\square$

Our next lemma sets out circumstances in which x-edges being absent from the boundary of a region R forces there to be no x-edge anywhere in R. The lemma further explains that regions that do not contain a y-edge and are bounded only by a-subtracks, inward-oriented b-subtracks, and t-subtracks take a highly constrained form, examples of which are shown in Figure 9.1.

Lemma 9.3. (Trapped x-noise) *Suppose R is a region in a reduced van Kampen diagram Δ over $\mathcal{P}$ such that R contains no y-edges and is bordered by a-subtracks, inward-oriented b-subtracks, t-subtracks, and paths in $\Delta^{(1)}$.*

(1) If there is an x-edge in R, then there is an x-edge in ∂R.

(2) If there is no x-edge in ∂R (in particular, if ∂R is made up of only a-subtracks, inward-oriented b-subtracks, and t-subtracks), then

- *(a) Each t-subtrack in ∂R crosses only a single edge; indeed, it crosses between an $r_{4,1}$-cell and an $r_{4,2}$-cell as in the example in Figure 9.1(right) and must transition to an outward-oriented a_1-subtrack in the $r_{4,1}$-cell and to an outward-oriented a_2-subtrack in the $r_{4,2}$-cell.*
- *(b) Each b-subtrack in ∂R only crosses a single edge. It transitions to an outward-oriented a_1-subtrack at one end and to an outward-oriented a_2-subtrack at the other.*
- *(c) There is at least one b- or t-subtrack in ∂R.*
- *(d) The a-subtracks in ∂R are all outward oriented. Together, they cross at least one a_1-edge and at least one a_2-edge*

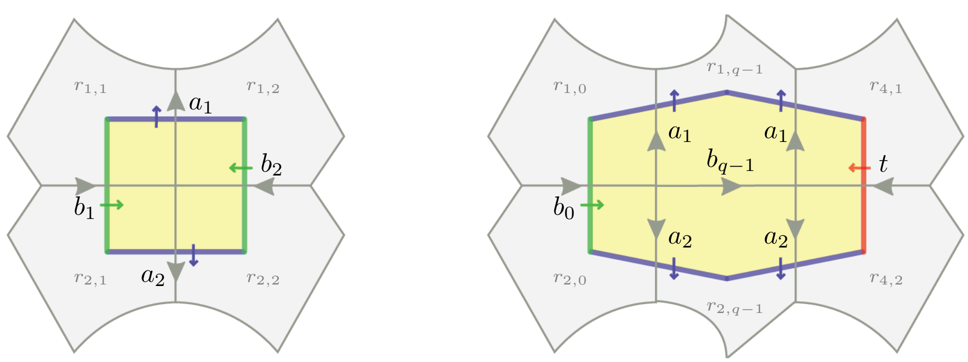


FIGURE 9.1. Examples of regions satisfying the conditions of Lemma 9.3(2)

Proof. For (1), first suppose that there is a 2-cell c in R. By Lemma 9.2, there is no t-loop in Δ, and so the two t-edges in ∂c are part of a t-subtrack that subdivides R into two regions R_1 and R_2. If (1) holds true for R_1 and R_2, then it holds true for R. Thus, via repeated such subdivisions, we reduce to the case where R contains no 2-cell. In that event, the subgraph $\mathcal{F}$ of $\Delta^{(1)}$ formed by the 1-cells in R is a forest: were it to contain an embedded circle, there would be a 2-cell within that circle and so in R. (In the examples of Figure 9.1, $\mathcal{F}$ is a single vertex in the left diagram and it is the single central edge labelled b_{q-1} in the right diagram.)

Assume there is no x-edge in ∂R. Suppose, for a contradiction, that there is an x-edge in R, and so in some connected component $\mathcal{F}_0$ of $\mathcal{F}$. Let v be the word one reads around $\mathcal{F}_0$. Let $\overline{v}$ be v with all letters other than $x_1^{\pm1}$ and $x_2^{\pm1}$ deleted.

By hypothesis, there are no y-edges in R. So v is a word on $a_1, a_2, b_0, \ldots, b_p, t, x_1, x_2$. Any x_1 or x_2 in v is the label of an edge e_x of a 2-cell and so is either part of a Rips subword from $\mathcal{X}$ in a defining relation, or is the lone x_j at the top (in the sense of Figure 4.1) of an $r_{3,i,j}$-cell c (for some $i \in \{0, \ldots, p\}$). In the latter event, no part of the b-track through c can be part of ∂R because then there would be an *outward*-oriented b-subtrack, contrary to hypothesis. It follows that ∂R contains the t-track of c (as c is not in R) and that $\mathcal{F}_0$ contains a portion of ∂c containing e_x so that x_j is part of a subword $X_*^{-1} b_i^{-1} x_j b_i X_*^{-1}$ of $v^{\pm1}$. So, after replacing $\overline{v}$ with a cyclic conjugate if necessary, $\overline{v}$ is a word on the X_*, $X_*^{-1} x_1 X_*$, and $X_*^{-1} x_2 X_*$.

Now, v freely reduces to the empty word since it is read around the tree $\mathcal{F}_0$. Therefore $\overline{v}$ also freely reduces to the empty word. Lemma 5.3 applies to $\overline{v}$. Folding up an edge-loop labelled by $\overline{v}$ to get the tree $\mathcal{F}_0$ equates to freely reducing $\overline{v}$. So the lemma tells us that parts of the boundary cycles of some pair of 2-cells is a common path in $\mathcal{F}_0$ labelled by a subword of some X_* of at least a quarter-length. These 2-cells are a back-to-back cancelling pair, contrary to the diagram being reduced. So we have the contradiction we seek.

To prove (2), we assume there are no x-edges in ∂R, and therefore none in R by (1).

For (2a), suppose τ is a t-subtrack in ∂R. It cannot intersect a t-edge that is part of a subword $Y_* t Y_*$ or $Y_* t^{-1} Y_* t Y_*$ in the boundary of a 2-cell, for then an adjacent y-edge would be in R, contrary to hypothesis. It also cannot intersect a t-edge that is part of a subword $X_* t X_*$ or $X_* t^{-1} X_* t X_*$ in the boundary of a 2-cell, for then an adjacent x-edge would be in R. The remaining possibility is that it intersects a t-edge at the top of an $r_{3,i}$- or $r_{4,i}$-cell. It cannot intersect the other t-edge in that cell, so ∂R has to switch from a t-subtrack to, respectively, a b_i- or a_i- subtrack within that cell. The former case cannot occur, as it would lead to an outward oriented b-track. In the latter case, the 2-cell on the other side of that top t-edge must also be an $r_{4,i}$-cell. As the diagram is reduced, we deduce that τ crosses from an $r_{4,1}$-cell to an $r_{4,2}$-cell across their common 'top' t-edge. Moreover, to avoid any y-edge being in R, ∂R must exit the $r_{4,1}$-cell across an a_1-edge and exit the $r_{4,2}$-cell across a_2-edge, and these a_1- and a_2-edges must have a common end-vertex in R and must both be oriented out of R.

For (2b), suppose β is a b-subtrack in ∂R. It is impossible that β enters and then exits a 2-cell: by hypothesis β is inward-oriented and so R would contain x- or y-edges from the bottom of the 2-cell (in the sense of Figure 4.1). So β crosses only a single b-edge, and when doing so it travels from one 2-cell to another. (It cannot start and end in the same 2-cell, as then two b-edges in the boundary of one 2-cell would be identified in Δ and that would imply that some subword of the boundary word represents 1 in G in such as way as to contradict the HNN-structure established in Proposition 7.1.) From our analysis of t-subtracks, we know that β cannot transition in ∂R

to a t-subtrack, and so it must transition to a-subtracks at each end. Indeed, it must transition to outward-oriented a-subtracks, since the x- or y-edges of a 2-cell in which a transition to an inward-oriented a-subtrack occurred would be in R. And β must connect an a_1-subtrack at one end and to an a_2-subtrack at the other, because otherwise the two 2-cells it passes through would be a cancelling pair, contrary to Δ being reduced.

For (2c), all that remains is to verify that ∂R is not an a-loop. It cannot be an inward-oriented a-loop, for then there would be x- or y-letters in R. Consider an inner-most outward-oriented a-loop α. The orientations on junctions in $\mathcal{G}_a$ force α to be an a_1- or a_2-loop, and the associated a_1- or a_2-annulus has inner boundary labelled by a non-empty word w on $b_0, \ldots, b_p$, which freely reduces to the empty word. The 2-cells in the annulus are $r_{1,i}$-cells ($i = 1, \ldots, p$) in the a_1 case and are $r_{2,i}$-cells ($i = 1, \ldots, p$) in the a_2 case. In either case cancellation of an inverse-pair of letters in w implies cancellation of a pair of 2-cells in Δ, contrary to the diagram being reduced.

We conclude that ∂R has at least one a_1-subtrack and at least one a_2-subtrack, and any a-subtrack transitioning to a b- or t-track is outward oriented. Were there an inward-oriented a-subtrack, it would have to be an a_1-subtrack α transitioning at either end to an outward oriented a_2-subtrack in distinct $r_{1,q-1}$-cells c and c'. Any 2-cell that α passed through between c and c' would lead to an x-or y-edge in R, so c and c' must be adjacent, which would be a contradiction because they are oppositely oriented. Thus (2d) follows. □

Here is the corresponding lemma for y-letters. It forgoes hypotheses excluding any particular type of edges from R, and it requires the a-subtracks, instead of b-subtracks, in ∂R to be inward-oriented.

Lemma 9.4. ***(Trapped y-noise)*** *Suppose R is a region in a reduced van Kampen diagram Δ over $\mathcal{P}$, bordered by b-subtracks, t-subtracks, inward-oriented a-subtracks, and paths in $\Delta^{(1)}$.*

(1) If there is a y-edge in R, then there is a y-edge in ∂R.

(2) If the b-subtracks in ∂R are inward oriented, then ∂R must include at least one x-edge or y-edge. In particular, in a reduced diagram there is no region R such that ∂R is comprised of inward-oriented a-subtracks, inward-oriented b-subtracks, and t-subtracks. (Figure 9.2 shows some examples of regions this precludes.)

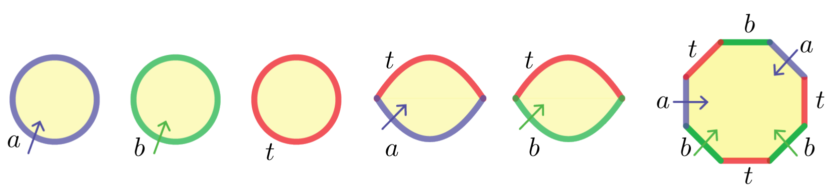


Figure 9.2. Examples of regions precluded by Lemma 9.4(2)

Proof. For (1), we follow the same approach as our proof of Lemma 9.3(1). As there, it suffices to prove the result in the case where there is no 2-cell in R. In that case, if there is a y-edge in R, then it appears in some connected component $\mathcal{F}_0$ of the forest of 1-cells in R, and around $\mathcal{F}_0$ we read a word v which freely reduces to the empty word. This v is a word on $a_1, a_2, b_0, \ldots, b_p, x_1, x_2, t$, and the Rips words $\mathcal{Y}$ (arising in the $Y_* t Y_*$ or $Y_* t^{-1} Y_* t Y_*$ per our presentation $\mathcal{P}$), and the $Y_*^{-1} a_i^{-1} y_j a_i Y_*^{-1}$ around $r_{4,i,j}$-cells—the key point here is that y_1 and y_2 do not appear on their own in this list and this is because if the y_j of $Y_*^{-1} a_i^{-1} y_j a_i Y_*^{-1}$ is in $v^{\pm 1}$, then the whole of that subword is in $v^{\pm 1}$ as an a-subtrack across that $r_{4,i,j}$-cell would be outwards-oriented, contrary to hypothesis. Let $\overline{v}$ be v with all letters other than $y_1^{\pm1}$ and $y_2^{\pm1}$ deleted. Then $\overline{v}$ is a word on the Y_*, $Y_*^{-1}y_1Y_*$, and $Y_*^{-1}y_2Y_*$ which freely reduces to the empty word. Lemma 5.3, translated to y-letters instead of x-letters, applies to $\overline{v}$, so as to imply that a pair of 2-cells cancel, contrary to the diagram being reduced.

For (2), assume, for a contradiction, that there is no x- or y-edge in ∂R. Then, by (1), there is no y-edge in R. This, together with the hypothesis that the b-subtracks in ∂R are inward oriented and the assumption that ∂R has no x-edges, means Lemma 9.3(2) applies, and part (2d) tells us that ∂R has non-trivial outward-oriented a-tracks, contradicting the hypothesis that a-subtracks in ∂R are inward-oriented. □

Remark 9.5. The analogue of Lemma 9.4(1) for x-edges fails. For an example, take the van Kampen diagram that demonstrates that $b_0^{-1}b_1^{-1}tb_1b_0$ equals a word on y_1, y_2, and t, which is comprised of one $r_{3,1}$-cell and a b_0-corridor made up of $r_{3,0,1}$- and $r_{3,0,2}$-cells and one $r_{3,0}$-cell.

Lemma 9.4(2) fails without the hypothesis that the b-tracks be inward-orientated. A "button" (Definition 9.8) provides an example.

Lemma 9.4(2) rules out a- and b-loops that are inward oriented. At this stage we can also rule out outward oriented a-and b-loops in some situations:

Lemma 9.6. *Let Δ be a reduced van Kampen diagram over $\mathcal{P}$.*

(1) If Δ has only $r_{4,*}$-cells, then Δ has no a-loops.
(2) If Δ has only $r_{2,*}$- and $r_{3,*}$-cells, then Δ has no b-loops

Proof. In both cases there are no $r_{1,*}$-cells. Thus the dual graphs $\mathcal{G}_a$ and $\mathcal{G}_b$ of Δ have no junctions, so every a-track is an a_i-track and every b-track is a b_j-track, for some i and j.

To prove (1), suppose for a contradiction that Δ has an a-loop. Then there is an innermost one α, which is an a_i-loop for $i = 1$ or 2, such that the region R enclosed α has no a-subtracks (as there are no junctions). As Δ has only $r_{4,*}$-cells, this means that the inner boundary of the annulus associated to α is a closed path in $\Delta^{(1)}$ that encloses no 2-cells, so traverses some edge e twice (in opposite directions). Lemma 9.4(2) implies that α is outward-oriented and this, together with the fact that α is an a_i-track for a fixed i, means that the possible labels y_1, y_2, t of e determine unique $r_{4,*}$-cells. It follows that there is an adjacent pair of oppositely oriented identical cells, contradicting the fact that Δ is reduced.

The proof of (2) is identical, noting that, for an innermost b_j-loop in a Δ as in (2), the possible labels x_1, x_2, t, a_2 of the edge e each determine a unique cell (given the orientation of b_j). □

We now define two types of diagrams containing bigons of subtracks which can occur in reduced diagrams over $\mathcal{P}$.

Definition 9.7. (Badge) A *badge* is a subdiagram consisting of a path with label t^n, where $n > 0$, with $2n + 2$ cells arranged around it as shown in Figure 9.3(left) for $n = 4$. Specifically, it has two $r_{i,j}$-cells that are connected by an a_i-corridor made up of n $r_{4,i}$-cells and a b_j-corridor made of n $r_{3,j}$-cells, such that the a_i-corridor and b_j-corridor are identified along their boundaries labelled t^n.

Definition 9.8. (Button) A *button* is a pair of 2-cells, specifically an $r_{1,p-1}$-cell and an $r_{1,p}$-cell, in a van Kampen diagram that are joined along the common a_1b_p subwords in their boundary word. Figure 9.3(center) shows a button. The mirror image of a button is also a button, so there are two buttons in the diagram in Figure 9.3(right).

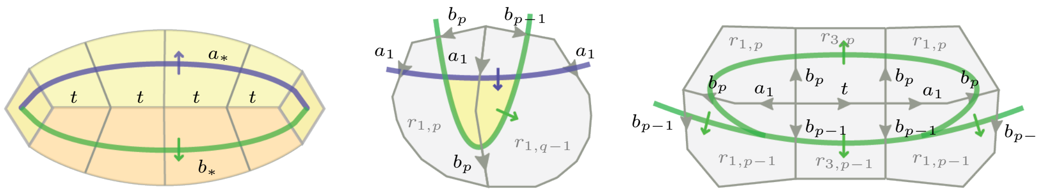


Figure 9.3. Left: a *badge*. Middle: a *button*. Right: a reduced diagram that includes two buttons and contains a loop that is an outward-oriented b-track.

Observe that a badge or button is dual to a bigon comprised of an a-subtrack and an outward oriented b-subtrack. The next lemma shows that such bigons always give rise to badges or buttons in the absence of y-edges. The second part puts a further restriction on certain bigons formed by an a_1-track and a b_i-track, which will be used in the proof of Corollary 9.10.

Lemma 9.9. (Bigons, badges, and buttons) *Let R be a region in a reduced van Kampen diagram Δ over $\mathcal{P}$, such that R does not contain any y-edges, and ∂R is a bigon comprised of an a-subtrack α and an outward oriented b-subtrack β. Then*

(1) The minimal subdiagram of Δ containing R contains either a badge or a button.
(2) If α is an a_1-subtrack and β is a b_i-subtrack, and R has no a_1-subtracks in its interior, then one of the intersections between α and β occurs in an $r_{1,i-1}$-cell.

Proof. If R is as in the statement of the lemma, we first prove that R contains a minimal region of the same type. Specifically, R contains a region S with boundary a bigon comprised of an a-track α_S and an outward oriented b-track β_S such that the interior of S contains no a- or b-subtracks.

To construct S, first observe that there can be no a-loop in R, as if there were one, it would enclose a region with no y-edges, contradicting Lemma 9.3(2c). Since R also has no teardrops (by Lemma 6.4), any a-subtrack α_1 in R is a path with distinct endpoints on ∂R. If α_1 has both endpoints on α, then (in the absence of a-loops and teardrops) we get a smooth path by replacing a subsegment of α with α_1, and this forms a smaller bigon with β. If one or both endpoints of α_1 are on β, then α_1 divides R into two regions, one of which has boundary a bigon comprised of an a-subtrack and a subtrack of β. Passing to a minimal instance, we obtain a region R' with boundary a bigon comprised of an a-track α' and an outward oriented b-track β' (a subtrack of β), such that R' has no a-subtracks in its interior.

Consider the minimal diagram containing R', and let D' be the subdiagram consisting of 2-cells not dual to α'. Then D' has only cells of type $r_{3,i}$ or $r_{3,i,j}$ (as any other cells would introduce a-subtrack in R'). So D' has no junctions and, by Lemma 9.6(2), has no b-loops. Suppose there is a b-subtrack $\beta_1 \neq \beta$ in R. Then β_1 has both endpoints on α' (as there are no junctions in D'). If β_1 is oriented into the bigon that it forms with α', then α' must be oriented outward by Lemma 9.4(2). As there are no y-edges in R, Lemma 9.3(2) applies, and implies that α' transitions from a_1 to a_2. This happens at some $r_{1,q-1}$-cell dual to α'. However, as α' is oriented outward, such a cell contributes part of an a_1-subtrack to the interior of R', a contradiction.

Thus any b-subtrack in R' has both endpoints on α', and is oriented out of the bigon it forms with α'. By passing to an innermost instance, we obtain a region S with boundary a bigon comprised of a subtrack α_S of α' and an outward oriented b-track β_S such that the interior of S contains no a- or b-subtracks.

To complete our proof of (1), we show that if R is minimal in that its interior contains no a- or b-subtracks, then the minimal subdiagram D containing R is either a badge or a button.

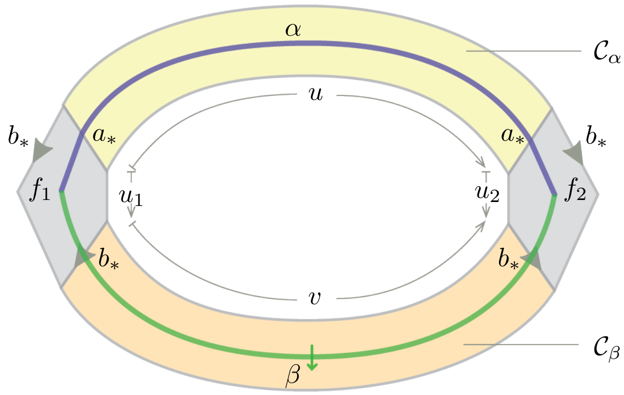


FIGURE 9.4. A bigon region per Lemma 9.9

Let C_α and C_β be the corridors dual to α and β respectively—see Figure 9.4. They intersect in distinct 2-cells f_1 and f_2 of type $r_{i,j}$ with $i = 1$ or 2. (If $f_1 = f_2$, then the orientation on β would force both corners of ∂S to be on the top half of some $r_{i,j}$-cell, and a terminal subpath of α would merge with an initial one to create a teardrop, which contradicts Lemma 6.4.) Further, the 2-cells of D are exactly the 2-cells of $C_\alpha \cup C_\beta$ (because a 2-cell strictly in the interior of R would result in interior a- or b subtracks).

The inner boundary of $C_\alpha \cup C_\beta$ has subpaths coming from f_1 and f_2 (labelled u_1 and u_2 respectively), from C_α (labelled u) and from C_β (labelled v), and these are oriented as shown in Figure 9.4. Next, we determine which letters can occur in these labels examining Figure 4.1 for cells which could occur in D under the given constraints.

Firstly, f_1 is an $r_{1,*}$- or $r_{2,*}$- cell, and given that β is outward-oriented, one sees that the only non-empty word that could arise as u_1 is b_i for some i (when f_1 is a $r_{1,i-1}$ cell and α is inward-oriented). However, as this would lead to b-subtracks inside R, we conclude that u_1 is empty. Likewise u_2 is empty. Thus $u = v$ as group elements.

Next, each cell of C_β apart from f_1 and f_2 is of type $r_{3,k}$, $r_{3,k,1}$ and $r_{3,k,2}$ (as any others would introduce a-subtracks in the interior of R). Since β is oriented outward, this means v is a word on x_1, x_2, t. Furthermore, the part of β between (and excluding) f_1 and f_2 has no junctions, and so it is a b_k-track for some fixed k. As x_1, x_2, t freely generate a free group in G (as a consequence of Proposition 7.4), and each of them appears in a unique $r_{3,k}$- or $r_{3,k,j}$-cell, Δ being reduced implies v is freely reduced.

If α is oriented outward, then each cell of C_α apart from f_1 and f_2 is of type $r_{4,i}$ (as any other cells would introduce y-edges or b-subtracks to the the interior of R). So u is a reduced word of the form t^n for some $n \in \mathbb{Z}$. Now, since v is reduced, we have that $v = u = t^n$ as words. Furthermore $n \neq 0$, for otherwise f_1 and f_2 would be identified along a pair of adjacent edges in each with label $a_i^{-1} b_k$ and as each such word appears in a unique cell, f_1 and f_2 would be oppositely oriented identical cells, contradicting the fact that Δ is reduced. Thus R is a badge.

If α is oriented inward, then C_α cannot have any 2-cells apart from f_1 and f_2, so u is empty. Then, as v is reduced, it is also empty, and f_1 and f_2 are distinct cells identified along a corner in each with label $a_i b_j$. Examining Figure 4.1 again, we see that this can only happen if they are a $r_{1,p-1}$-cell and an $r_{1,p}$-cell identified along their corners labelled $a_1 b_p$, so that R is a button. This completes the proof of (1).

Now assume R satisfies the additional hypotheses in (2) of this lemma (but is not necessarily minimal). In particular, the interior R has no a_1-subtracks, but could have a_2- or b_j-subtracks. We continue with the notation of Figure 9.4. The intersection of an a_1-track and a b_i-track can only occur in an $r_{1,i}$- or $r_{1,i-1}$-cell. Assume for a contradiction that f_1 and f_2 are both of the former type. Now, if α is oriented outwards, then u_1 and u_2 are empty and u is a word on $b_0, \ldots, b_p$ (here we do not have t, because an $r_{4,1}$-cell would produce a y-edge in R, a contradiction). If α is oriented inwards, then $u_1 = b_{i+1}^{-1}$ and $u_2 = b_{i+1}$ and u is a word on

$$b_p(X_* t^{-1} X_* t X_*)^{-1}, \quad b_q b_{q-1}(X_* t^{-1} X_* t X_*)^{-1}, \quad b_{i+1} b_i (X_* t^{-1} X_* t X_*)^{-1} \ \ (i \neq 0,\, q-1,\, p).$$

Now define $\overline{u}$ and $\overline{v}$ to be the images of these words in the quotient $Q = F(b_0, \ldots, b_p) \rtimes \mathbb{Z}$ of G from (3.1) resulting from killing $a_2, t, x_1, x_2, y_1, y_2$. Then $\overline{v}$ is empty and $\overline{u}$ is a word on $b_1 b_0, \ldots, b_p b_{p-1}, b_p$, which is a free basis for $F(b_0, \ldots, b_p)$. So $b_{i+1}^{-1} \overline{u} b_{i+1} = 1$ in Q, and so $\overline{u} = 1$. So there is a canceling pair in u, and this implies that there is a pair of adjacent oppositely oriented cells, contradicting the hypothesis that the diagram Δ is reduced. An analogous analysis rules out α_1 being outward-oriented. This proves (2). □

The next corollary summarizes the restrictions on loops in reduced diagrams obtained so far.

Corollary 9.10. (Loops) *Suppose Δ is a reduced diagram.*

(1) Δ has no t-loops and no inward-oriented a- or b-loops.
(2) Every a-loop in Δ encloses a y-edge.
(3) Δ has no b_i-loops, and if Δ has no buttons, then it has no b-loops.

Proof. Lemmas 9.2 and 9.4(2) establish (1).

Were there an a-loop enclosing no y-edges, it would satisfy the hypotheses of Lemma 9.3(2) but fail the conclusion in part (2c) of that lemma. This proves (2).

For (3), suppose β is a b-loop in Δ, as shown in Figure 9.5. Then β is oriented outward by (1). If R is the region enclosed by β, then R contains no y-edges by Lemma 9.4(1). Consequently, R contains no a-loops by (2) of this corollary. Because Δ has no teardrops by Lemma 6.4, any a_1-subtrack in R must intersect β in two distinct points, and divides R into two bigons.

Let Δ_0 be the minimal diagram containing R. There are no 2-cells of type $r_{4,*,*}$ or $r_{4,*}$ in Δ_0, because any such 2-cell would have to be inside β and would give rise to a y-edge there. So Lemma 9.6(2) tells us that Δ_0 contains at least one $r_{1,*}$-cell. Therefore R contains an a_1-subtrack. Let α be an a_1-subtrack in R that forms a bigon with a subtrack β_1 of β, and is *innermost* in that there is no a_1-subtrack in the region R_1 enclosed by α and β_1.

Now suppose β is a b_i-loop for some fixed i, and so β_1 is a b_i-subtrack. Then applying Lemma 9.9(2) to R_1, we see that one of the intersections between α and β_1 occurs in an $r_{1,i-1}$-cell. This is a contradiction, as β, being a b_i-track, cannot pass though an $r_{1,i-1}$-cell. Thus Δ has no b_i-loops.

Finally suppose that Δ has no buttons and that β is a b-loop. Then, by Lemma 9.9(1), the minimal subdiagram containing R_1 contains a badge. The a-subtrack of this badge is dual to at least one $r_{4,i}$-cell, and this cell is in the interior of R. This is a contradiction: as already noted, each $r_{4,i}$-cell has a y-edge, while R has none. This completes our proof of (3). □

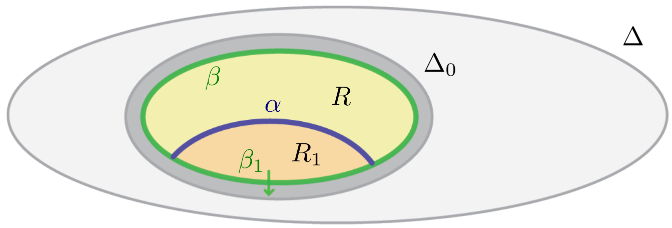


FIGURE 9.5. Our proof of Corollary 9.10(3), illustrated

Remark 9.11. Figure 9.3 shows how Corollary 9.10(3) can fail without the hypothesis absenting buttons. Corollary 9.10(2) cannot be upgraded to rule out all a-loops: a reduced diagram with an outward oriented a_1-track can be formed by circling an $r_{3,0}$-cell (which has y-edges) with an outward oriented a_1-annulus made up of two $r_{1,0}$-cells, two $r_{4,1}$-cells, and some $r_{4,1,j}$-cells.

Our next two lemmas concern the impact of the presence of Rips subwords in the sides of t-corridors or in generalizations defined in the following manner. The following expanded definition of a corridor $\mathcal{C}$ and the lemma that follows it are motivated by applications to our proof of Lemma 9.16.

Definition 9.12. (Generalized corridors) Let $\mathcal{C}$ be a set of r distinct 2-cells $C_1, C_2, \ldots, C_r$ in a reduced van Kampen diagram over our presentation $\mathcal{P}$ for G such that there are edges $e_0, \ldots, e_r$ with the property that for $i = 1, \ldots, r-1$, the edge e_i is in both ∂C_i and ∂C_{i+1}. Suppose the word read clockwise around C_i is $z_i f_i z_{i+1}^{-1} g_i$, where z_i labels edge e_i. Then the words along the top and bottom boundaries of $\mathcal{C}$ are $f_1 f_2 \cdots f_r$ and $g_1^{-1} g_2^{-1} \cdots g_r^{-1}$ respectively.

Lemma 9.13. (Rips words cause the sides of corridors to be near injective and adjacent corridors to have small overlap.) *There exists a constant $K \geq 1$ such that reduced van Kampen diagrams Δ have the following properties.*

Suppose $\mathcal{C}$ is a generalized corridor, μ is the path along one side of $\mathcal{C}$, and the word read along μ is $f := f_1 f_2 \cdots f_r$ (all per Definition 9.12). Refer to $f_1, \ldots, f_r$ as the syllables of f. A Rips subword *in a syllable f_i of f is an element of $(\mathcal{X} \cup \mathcal{Y})^{\pm 1}$ appearing as a subword. Suppose that if $1 \leq i \leq j \leq r$ are such that $f_i, \ldots, f_j$ do not have Rips subwords, then $f_i \cdots f_j$ is a reduced word on $\{a_1, a_2, b_0, \ldots, b_p\}^{\pm 1}$.*

Suppose $\overline{\mu} \subseteq \mu$ is an injective path from the initial vertex of μ to its terminal vertex. So the word $\overline{f}$ read along $\overline{\mu}$ can be obtained from f by a sequence Σ of free reductions (successive cancellations of adjacent inverse-pairs of letters). Then:

(1) (a) At least one letter of every Rips subword in a syllable survives in $\overline{f}$.
(b) $|f| \leq K|\overline{f}| + K$.
(c) If a subpath μ_0 of μ is a loop and encloses no 2-cells, then the subword f_0 of f read along μ_0 has length at most K.

Suppose μ' is the path along one side of another generalized corridor $\mathcal{C}'$ and $f' := f'_1 f'_2 \cdots f'_{r'}$ is the word read along it. Suppose that for all i, some element of $(\mathcal{X} \cup \mathcal{Y})^{\pm 1}$ is a subword of f'_i. Suppose $\mathcal{C}$ and $\mathcal{C}'$ have no 2-cells in common and that they start and end on $\partial\Delta$ (that is, $e_0, e_r, e'_0, e'_{r'}$ are in $\partial\Delta$). Suppose that

$$I \ := \ \mathcal{C} \cap \mathcal{C}' \ = \ \mu \cap \mu' \ \neq \ \emptyset.$$

(2) *Suppose μ_0 and μ'_0 are the shortest subpaths of μ and of μ', respectively, such that $I = \mu_0 \cap \mu'_0$. If $\mu_0 \cup \mu'_0$ encloses no 2-cells, then $|\mu_0|, |\mu'_0| \leq K$.*

Proof. For (1), we can interpret the sequence Σ as folding together adjacent pairs of edges in a $|f\overline{f}^{-1}|$-sided simple polygonal-path in the plane until we have the planar tree in Δ whose boundary circuit is $\mu\overline{f}^{-1}$. Because every cyclic conjugate of a defining relator (of Figure 4.1) is freely reduced, no cancellation of a pair of letters within a syllable of f occurs in the course of Σ.

Given $\sigma \in (\mathcal{X} \cup \mathcal{Y})^{\pm 1}$, let P_σ and S_σ denote its prefix and suffix, respectively, such that $\sigma = P_\sigma S_\sigma$ as words, and $|P_\sigma| = \lfloor |\sigma|/2 \rfloor$. Suppose of all the Rips subwords in the syllables of f, some subword σ of f_l is the first such that either P_σ and S_σ is fully cancelled away in the course of Σ. Assume it is S_σ that is first cancelled away. (The argument if it is P_σ will be essentially the same, and we omit it.) Then S_σ must cancel with a subword of f_m, where $m > l$ is minimal such that f_m has a Rips subword. But that is impossible: the $C'(1/4)$-condition for $\mathcal{X} \cup \mathcal{Y}$ and the fact that each of its elements has length at least 100, imply that some subword of σ^{-1} of at least a quarter of its length is a subword of f_m, and moreover the 2-cell C_l cancels with C_m in Δ, contrary to Δ being a reduced diagram. This proves (1a).

Now suppose that syllables $f_i, \ldots, f_j$ do not contain Rips subwords. Then (by hypothesis) $f_i \cdots f_j$ is a reduced word on $\{a_1, a_2, b_0, \ldots, b_p\}^{\pm 1}$. So the number of letters that can cancel away on freely reducing $f_{i-1} f_i \cdots f_j f_{j+1}$ is less than four times the length of the longest defining relation for our group. Together with (1a), this implies (1b) and (1c) for a suitable constant $K \geq 1$.

For (2), first we observe that I is a path because, by hypothesis, $\mu_0 \cup \mu'_0$ encloses no 2-cells. Let w_0 and w'_0 be the words read along μ_0 and μ'_0, respectively. Assume, without loss of generality, that μ_0 and μ'_0 are oriented in the same direction—which is to say that $w_0(w'_0)^{-1}$ is the word around $\mu_0 \cup \mu'_0$. Then free reduction takes w_0 and w'_0 to the word w read along I. (We are not claiming w is freely reduced—further free reduction may be possible.)

The proof can then be completed in a similar manner to part (1c). In short, if there is a Rips subword σ in w'_0, then there must be a subword of σ in w_0 also and these two words have large overlap in w, so as to imply that there are cancelling 2-cells in $\mathcal{C}$ and $\mathcal{C}'$. So μ'_0 contains no complete Rips subword and, because each of the syllables of μ' contains a Rips subword (by hypothesis), μ'_0 has length at most a constant. It then follows that μ_0, which also contains no complete Rips subword, also has length at most a constant: within w_0, any f_i that contains no Rips subword can only cancel with the neighbouring f_{i-1} or f_{i+1} if they contain a Rips subword (so at most some constant number of letters in total can cancel away) and the remaining letters must be in w', which has length at most $|\mu'_0|$. $\square$

Lemma 9.14. *Suppose μ is the path along one side of a t-corridor $\mathcal{C}$ in a reduced van Kampen diagram Δ. Then the first y-edge e of Δ traversed by μ is not traversed a second time by μ.*

Proof. Suppose, on the contrary, μ traverses e more than once, then (because Δ is planar and μ is the side of a corridor) it does so exactly twice—once in each direction—and the subpath $\overline{\mu}$ of μ starting with the first traverse of e and ending with the second traverse is a loop. (See Figure 9.6.)

With a view to applying Lemma 9.13(1) to $\mathcal{C}$, we check its hypotheses. As $\mathcal{C}$ is a t-corridor, our defining relations imply that the label of $\mu \cap C$ contains a Rips subword for every cell C of $\mathcal{C}$. There are no t-edges within the region $\overline{\Delta}$ enclosed by $\overline{\mu}$, for if there were, then there would be a

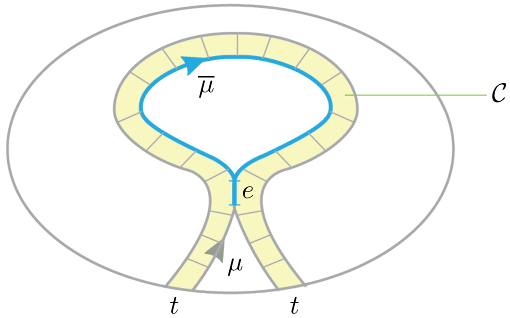


FIGURE 9.6. The t-corridor of our proof of Lemma 9.14

t-loop within $\overline{\Delta}$, contradicting Lemma 9.2. So $\overline{\mu}$ does not enclose any 2-cells. Thus Lemma 9.13(1a) applies, and tells us that the label $\overline{w}$ of $\overline{\mu}$ has no Rips subword from $(\mathcal{X} \cup \mathcal{Y})^{\pm 1}$ as a subword.

On the other hand, Corollary 7.2 implies that $\overline{w}$ cannot be a subword of the boundary word of a single 2-cell of $\mathcal{C}$. In particular, if C_e is the cell of $\mathcal{C}$ containing the initial point of $\overline{\mu}$ (and the edge e), then $\overline{\mu}$ extends beyond C_e, and intersects at least one other cell of $\mathcal{C}$. Thus if $t^{\pm 1} u t^{\mp 1} = v$ is the boundary label of C_e, where u labels $\mu \cap C_e$, then u has the form $u_1 y_* u_2$, where $y_* u_2$ is a prefix of $\overline{w}$. Moreover, as e is the first y-edge in μ, it follows that u_1 has no y-edges. Then, examining Figure 4.1, we see that u_2 necessarily contains the entirety of some Rips subword Y_* from $\mathcal{Y}^{\pm 1}$ as a subword. (This is true even if the first letter of $\overline{w}$ is the lone $y_j^{\pm 1}$ that arises in the $r_{4,i,j}$-cells.) This contradicts our earlier conclusion that $\overline{w}$ has no Rips subwords. $\square$

We will use our next lemma in our proof of Lemma 11.2(2). Here is the intuition. Imagine a diagram consisting of a sequence of side-by-side vertical corridors as in Figure 9.7. If there are no y-edges at the bottom of the diagram, then we can slice horizontally through it and discard the portion above the cut, so that the diagram that remains has no y-edges and the length of the cut is at most a constant times the length of the top.

Lemma 9.15. **(y-edges in side-by-side t-corridors)** *There exists a constant $C > 0$ with the following property. Suppose u and v are words that represent the same element of G and that v contains no y-letters. Suppose Δ is a reduced diagram for uv^{-1}. Let $*_0$ and $*_1$ be the vertices on $\partial\Delta$ where both u and v start and end (respectively). Assume that every t-corridor in Δ connects a $t^{\pm 1}$ in u to a $t^{\pm 1}$ in v.*

Then there is a word v' read along some injective path through $\Delta^{(1)}$ from $*_0$ to $*_1$ such that $|v'| \leq C|u|$ and the subdiagram Δ' (per Figure 9.7), which is a van Kampen diagram for $v(v')^{-1}$, contains no y-edges.

Proof. We denote the t-corridors of Δ by $\tau_1, \ldots, \tau_m$, for some m, where τ_i connects the ith $t^{\pm 1}$ in v to the ith $t^{\pm 1}$ in u. Every t-corridor is of this form, by hypothesis. Observe that $m \leq |u|$.

For all i, let $\mathcal{S}_i^-$ and $\mathcal{S}_i^+$ be the paths from v to u along the two sides of τ_i, with $\mathcal{S}_i^-$ emanating from the starting vertex of the $t^{\pm 1}$ of τ_i in v and $\mathcal{S}_i^+$ from its ending vertex. Assuming there is a y-edge on $\mathcal{S}_i^{\pm 1}$, let $e_i^{\pm 1}$ be the lowest—which is to say that $e_i^{\pm 1}$ is the first y-edge that $\mathcal{S}_i^{\pm 1}$ traverses. If there are y-edges in one side of a 2-cell in a t-corridor, then there are y-edges in the other side of that cell. So e_i^- and e_i^+ (if defined) are in the boundary of the same 2-cell C_i of τ_i. Moreover,

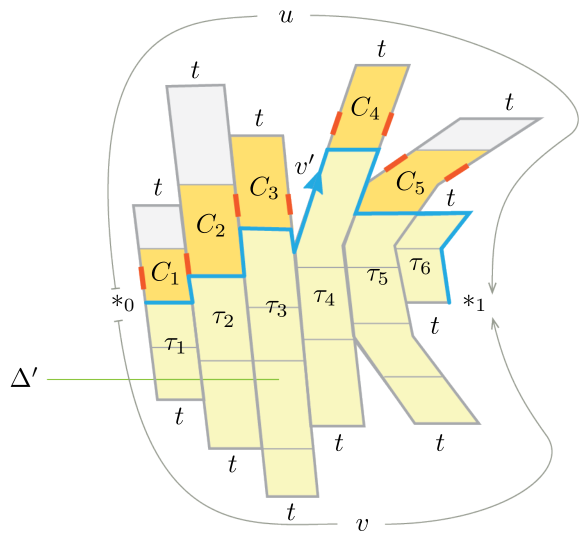


FIGURE 9.7. Lemma 9.15, illustrated

as Lemma 9.14 guarantees that $\mathcal{S}_i^{\pm 1}$ does not traverse $e_i^{\pm 1}$ a second time and, because v has no y-edges, $e_i^{\pm 1}$ is either in u or part of the neighboring t-corridor. It follows that for all i,

- either both e_i^+ and e_{i+1}^- exist, they agree, and they are not in u,
- or both exist and are in u,
- or only one exits and is in u,
- or neither exists.

Take C to b the maximum length of a defining relator in $\mathcal{P}$. Then there is an injective path through $\Delta^{(1)}$ from $*_0$ to $*_1$ that follows portions of u and portions of the boundary circuits of the at most $|u|$ 2-cells C_i, such that the word v' along this path satisfies the required conditions. (This path is shown in blue in Figure 9.7.) $\square$

Our final lemma is illustrated by Figures 9.8 and 9.9. (The path ρ is in the graph dual to $\Delta^{(1)}$.) In short, it says, in the notation of Figure 9.8, that the diagram cannot flare out towards v. Its application in Lemma 11.2(3) will be that certain regions can be sliced off a reduced diagram without much increasing the length of that diagram's boundary. Thereby we will simplify diagrams that demonstrate distortion.

Lemma 9.16. (The lengths of compound-tracks between points on the boundary) *There exists a constant $C \geq 1$ with the following property. Suppose a region R in a reduced diagram Δ is bounded by a portion μ of $\partial\Delta$ and a compound track ρ that is a concatenation of a-subtracks, inward-oriented b-subtracks, and t-subtracks. Let D be the minimal subdiagram of Δ containing R. (That is, D is the union of R and the generalized corridor $\mathcal{C}$ through which ρ passes.) So D is a van Kampen diagram for vu^{-1} for some words v and u such that v is read around $\partial\Delta$ starting and ending with the edges where μ and ρ meet. Suppose either*

(1) the a-subtracks in ρ are oriented into R, or
(2) D contains no y-edges.

Then $|u|$ and the number of edges $|\rho|$ of Δ that ρ crosses are both at most $C|v|$.

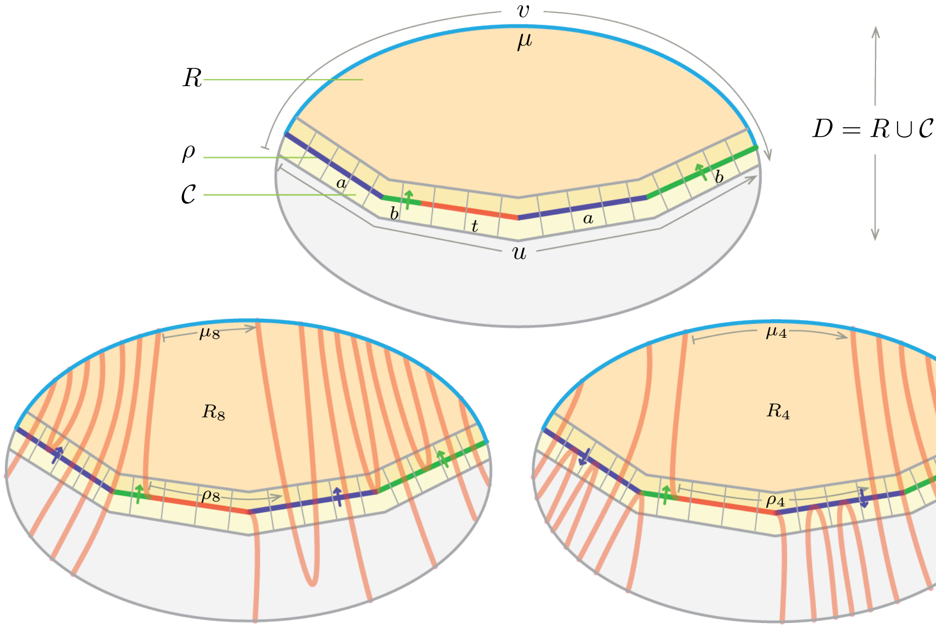


FIGURE 9.8. Top: a region R enclosed by a portion μ of $\partial\Delta$ and a compound track ρ comprised of a- and t-subtracks and inward-oriented b-subtracks per Lemma 9.16. The lower diagrams depict the t-tracks incident with ρ when (left) the a-subtracks are inward-oriented, and (right) when R and $\mathcal{C}$ contain no y-edges. Note that each R_i could have t-tracks with both endpoints on μ_i—these are are not pictured here, but are shown in the detail in Figure 9.9.

Proof. We will establish the claimed bounds by examining the t-tracks through R. By Lemma 9.2, there are no t-loops in R or indeed anywhere in Δ, because Δ is reduced. Next we will argue that there is no t-subtrack τ in R which is non-trivial (i.e., not a single point) and which starts and ends on ρ and otherwise is in the interior of R. If there were, then a subpath of τ together with a subpath of ρ would bound a region $R' \subseteq R$ that cannot exist in a reduced diagram: under hypothesis (1), R' would be contrary to Lemma 9.4(2), and under hypothesis (2), Lemma 9.3(2) applies to R' and its conclusion (2a) tells us there is an $r_{4,1}$-cell and an $r_{4,2}$-cell in D, and therefore a y-edge in D, contrary to assumption.

The tracks $\tau_1, \ldots, \tau_m$ of R which have one endpoint on μ and the other on ρ divide R into subregions $R_0, R_1, \ldots, R_m$ as illustrated in Figure 9.8, with the lower left diagram depicting hypothesis (1) and lower right, hypothesis (2). Under either hypothesis (1) or (2), the previous paragraph implies that every t-subtrack entering the interior of R_i has both endpoints on μ. In more detail, μ and ρ can be expressed as concatenations of subpaths $\mu_0, \mu_1, \ldots, \mu_m$ and $\rho_0, \rho_1, \ldots, \rho_m$, respectively, so that for each i, the region R_i is bounded by μ_i, ρ_i, τ_i and τ_{i+1} (with τ_0 and τ_{m+1} being trivial paths).

Guided by the locations of the letters t^{ϵ_i} read along the edges where the τ_i meet μ, express v as

$$v \;=\; t^{\epsilon_0} v_0 t^{\epsilon_1} v_1 t^{\epsilon_2} v_2 \cdots t^{\epsilon_m} v_m t^{\epsilon_{m+1}}$$

where each $\epsilon_i = \pm 1$ and each v_i is a subword of v (which may contain further $t^{\pm 1}$).

Fix $i \in \{0, \ldots, m\}$. Let ν_i denote the concatenation of τ_i, ρ_i and τ_{i+1}, so that R_i is bounded by μ_i and ν_i. Let $C_1, \ldots, C_r$ denote the 2-cells traversed by ν_i, as shown in Figure 9.9 (with $i = 4$ and $r = 17$). Together they form a generalized corridor $\mathcal{C}$ in the sense of Definition 9.12. Let Δ_i be the maximal subdiagram that is a subset of R, includes the portion of $\partial\Delta$ labelled by v_i, and does not intersect τ_i, ρ or τ_{i+1}. Let $f = f_1 \ldots f_r$ be the word along the side of $\mathcal{C}$ that is in R_i. Then Δ_i is a van Kampen diagram for $f v_i^{-1}$. We refer to $f_1, \ldots, f_r$ as the *syllables* of f. (It may be that f is not reduced and Δ_i is not homeomorphic to a 2-disc.)

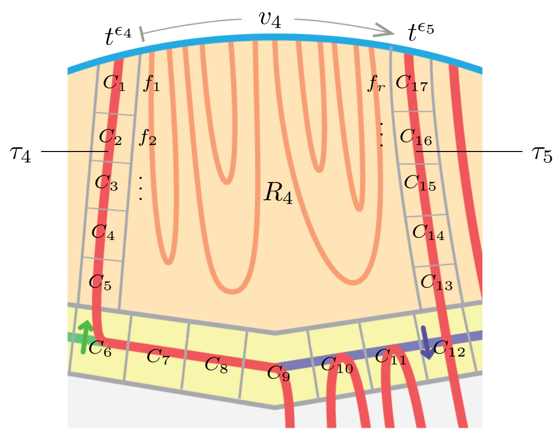


FIGURE 9.9. The region R_4 illustrated per our proof of Lemma 9.16.

We will show that there exists a constant $L \geq 1$ such that, if $|\nu_i|$ denotes the number of edges of Δ crossed by ν_i, then

$$|\nu_i| \;\leq\; L|v_i| + L. \tag{9.1}$$

We will argue that $\mathcal{C}$ satisfies the hypotheses of Lemma 9.13. The label of C_j, read clockwise, is of the form $\alpha f_j \beta^{-1} \hat{f}_j$, with $\alpha, \beta \in \left\{a_1^{\pm 1}, a_2^{\pm 1}, b_1, \ldots, b_p, t^{\pm 1}\right\}$ being the letters labeling edges dual to which ν_i enters and leaves C_j, respectively. (The hypothesis that the b-subtracks that are part of ρ are oriented into R precludes α or β being among $b_1^{-1}, \ldots, b_p^{-1}$.)

Suppose f_j does not have a Rips subword. Inspecting the defining relators for G (Figure 4.1), we find that one of α and β is in $\left\{a_1^{-1}, a_2^{-1}\right\}$ and the other is in $\left\{a_1^{-1}, a_2^{-1}, t\right\}$, and this can only occur when there is an a-subtrack in ρ that is oriented out of R, contrary to hypothesis (1), which means that hypothesis (2) must apply. But then the only way one of α and β can be t is if C_j is an $r_{4,i}$-cell and α and β label the top and right edges (or vice versa) in the sense of Figure 4.1, which is excluded by (2) because $r_{4,i}$-cells have y-edges. So $\alpha, \beta \in \left\{a_1^{-1}, a_2^{-1}\right\}$ and C_j is an $r_{1,*}$- or $r_{2,*}$-cell, with $* \neq 0$ lest we contradict (2). If C_j is an $r_{1,*}$-cell, then $f_j \in \{b_1, \ldots, b_p, b_{q-1}a_1\}^{\pm 1}$. If C_j is an $r_{2,*}$-cell, then $f_j \in \{b_1, \ldots, b_p\}^{\pm 1}$.

Next suppose f_{j+1} also does not contain Rips word. If one of C_j and C_{j+1} is an $r_{1,*}$-cell and the other is an $r_{2,*}$-cell, then one of them must be an $r_{1,q-1}$-cell and they meet along an edge labelled a_2^{-1}. In this event, there is no cancellation between f_j and f_{j+1}, because $f_j f_{j+1}$ is $(b_l^{\pm 1} a_1^{-1} b_{q-1}^{-1})^{\pm 1}$ for some l. If, on the other hand, C_j and C_{j+1} are both $r_{1,*}$-cells or both $r_{2,*}$-cells, then there can be no cancellation between f_j and f_{j+1} lest C_j and C_{j+1} be a cancelling pair of 2-cells, contrary to Δ being a reduced diagram. Thus if consecutive syllables $f_j, \ldots, f_l$ (for $j \leq l$) do not contain Rips words, then $f_j, \cdots, f_l \in \{b_1, \ldots, b_p, b_{q-1}a_1\}^{\pm 1}$ and $f_j \cdots f_l$ is a freely reduced word. So $\mathcal{C}$ satisfies the hypotheses of Lemma 9.13.

Let $\Delta_{\mathcal{C}}$ be the minimal subdiagram of Δ containing $\mathcal{C}$ and let $\overline{\Delta_i}$ be the maximal subdiagram of Δ_i that contains the path labelled v_i and does not intersect the interior of $\Delta_{\mathcal{C}}$. Let $\overline{f}$ be the word such that $\overline{\Delta_i}$ is a van Kampen diagram for $\overline{f}v_i^{-1}$. There are no 2-cells in $\Delta_i \setminus \overline{\Delta_i}$ because there would be a t-track through such a 2-cell and we know that all t-tracks in Δ_i connect a pair of edges in v_i. So $\overline{f}$ can be obtained from f by freely reducing f (perhaps only partially: $\overline{f}$ need not be freely reduced), so as to remove all the letters which label any 1-dimensional *spikes* of Δ_i that protrude into $\mathcal{C}$. By Lemma 9.13(1b), there is a constant $K \geq 1$ such that

$$|f| \ \leq \ K|\overline{f}| + K. \tag{9.2}$$

Next, suppose $\mathcal{C}'$ is a t-corridor that joins a pair of t-letters in v_i. Then $\mathcal{C}$ and $\mathcal{C}'$ have no 2-cells in common: were there such a 2-cell, the t-track through $\mathcal{C}'$ would intercept ν_i (see Figure 6.3). Moreover, there can be no 2-cell in any subdiagram of Δ_i whose boundary is made up of a path along one side of $\mathcal{C}$ and a path along one side of $\mathcal{C}'$: there would be a t-subtrack through such a 2-cell, and it would either be part of a t-loop (contrary to Lemma 9.2) or would join two points on ρ_i (which we argued at the start of this proof cannot happen). So Lemma 9.13(2) applies and tells us that the overlap between $\mathcal{C}$ and $\mathcal{C}'$ has length at most the constant K.

Each edge of the $\overline{f}$-portion of $\partial\Delta_i$ is either in the v_i-portion of $\partial\Delta_i$ or is the side of such a t-corridor $\mathcal{C}'$. At most $|v_i|/2$ t-corridors join a pair of t-edges in v_i. We conclude that there is a constant $K' \geq 1$ such that

$$|\overline{f}| \ \leq \ K'|v_i|. \tag{9.3}$$

The existence of a constant $L \geq 1$ such that (9.1) holds now comes from combining $|\nu_i| \leq |f|$, (9.2), and (9.3).

Finally, using $|\rho_i| \ \leq |\nu_i|$ and summing (9.1) over all $0 \leq i \leq m$, we get that

$$|\rho| \ \leq \ \sum_{i=0}^{m} |\rho_i| \ \leq \ L|v| + L(m+1) \ \leq \ 2L|v|.$$

So $|\rho|$ and $|u|$ are both at most $C|v|$ for a suitable constant $C \geq 1$ derived from L and the maximum length of a defining relation. $\square$

While we will only call on the lemma above in its full generality, we note that in the case when ρ is a t-track, it gives:

Corollary 9.17. *The vertex groups of the HNN-structure $G = F \ast_t$ are undistorted in G.*

10. Intersection patterns for a pair of paths across a disc

Towards further understanding the intersection patterns of tracks, we consider here how a pair of transversely oriented paths in a disc may intersect if there are no "sink-regions." The results in this section are formulated so as to be combinatorial, bypassing issues such as paths intersecting each other infinitely many times. We could, equivalently, have made the paths in this section injective combinatorial paths in the 1-skeleton of a finite 2-complex homeomorphic to a 2-disc.

Definition 10.1. (Sinks and sources) Let σ and τ be piecewise-linear paths in a 2-disc D, each of which is made up of finitely many straight-line segments and has a transverse orientation. Suppose that σ and τ meet ∂D at exactly four points—their end points—and that their intersections are transverse. A region R in D such that ∂R is a union of subpaths of σ and τ is called a *sink region* if the orientation on each subpath in ∂R points inward and a *source region* if the orientation on each subpath in ∂R points outward. Note that by definition, the boundary of a sink or source region does not include any part of ∂D.

Lemma 10.2. *Let σ and τ be paths in a 2-disc D as per Definition 10.1. If there is no sink region in D, then, up to a homeomorphism of D, we have one of the cases displayed in Figure 10.1. (The cases are arranged into four families according to the possible relative orientations of σ and τ where they meet $S^1 = \partial D$. Cases (2) and (3) include the possibility that σ and τ do not intersect.)*

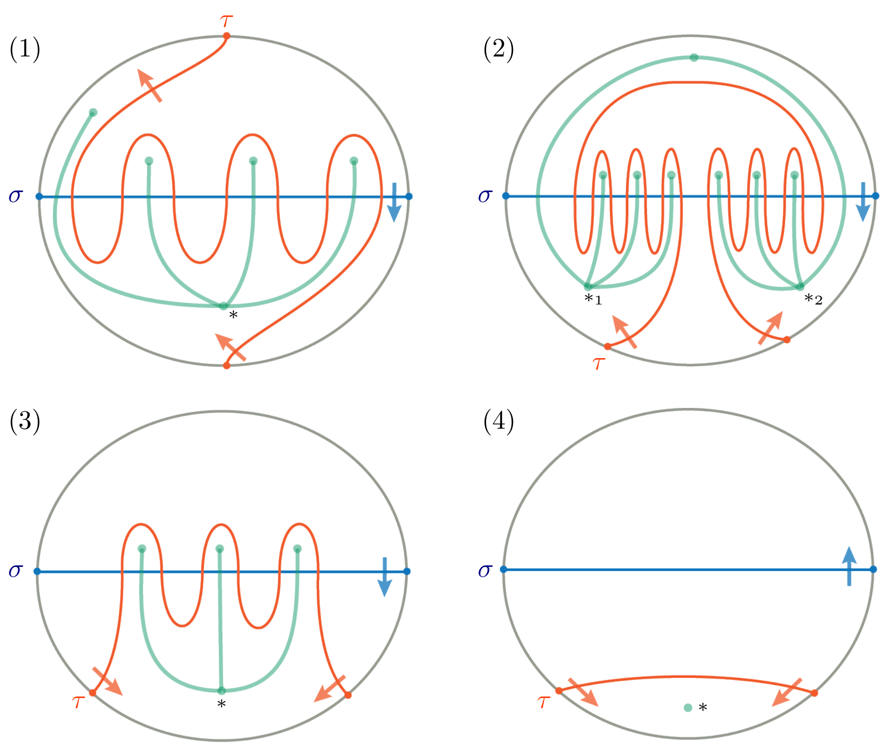


FIGURE 10.1. The intersections patterns of two transversely oriented chords σ and τ across a disc per Lemma 10.2, if there are no sink regions. There are four cases depending on the relative positions of the end points of σ and τ and on their orientations. In (1) σ and τ intersect $2n-1$ times for some $n \geq 1$, in (2) they intersect either 0 times or $(2m-1)+(2n-1)$ times for some $m, n \geq 1$, in (3) they intersect $2n$ times for some $n \geq 0$, and in (4) they do not intersect.

Proof. Consider the planar graph $\mathcal{G}$ whose vertices are the points of intersection of σ and τ and the four end points, and whose edges are the subpaths of σ, τ, and ∂D that connect them (call these σ-, τ-, and ∂D-*edges*, respectively). The path τ subdivides D into two subdiscs (ditto the path σ). Let $\mathcal{T}$ be the planar graph (in fact, *tree*) that has

- vertices dual to every *face* of $\mathcal{G}$ (i.e, connected component of $D \setminus \mathcal{G}$) that the orientation of τ points into, and
- edges dual to all σ-edges.

Figure 10.2(left) shows an example—there is no loss of generality in taking σ to be a diameter of the disc.

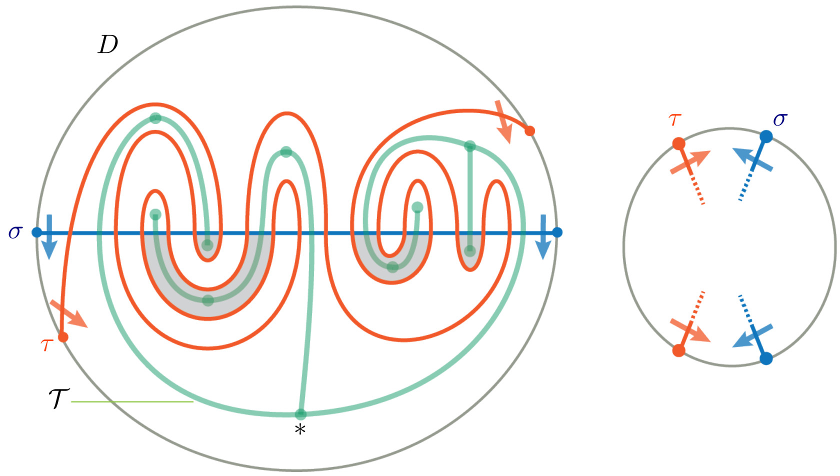


FIGURE 10.2. Left: our proof of Lemma 10.2, illustrated. Right: orientations per Corollary 10.3.

Case (1) of Figure 10.1 concerns when the end points of σ and τ alternate around ∂D. Cases (2)–(4) subdivide the eventuality where they do not alternate to three mutually exclusive possibilities for the orientations of σ and τ where they meet ∂D, namely, oriented towards each other, in the same direction, or away from each other.

Depending on whether or not σ and τ intersect, there are either four or three faces in $\mathcal{G}$ that have ∂D-edges in their boundaries. Call these *boundary faces*. A face f of $\mathcal{G}$ either has all the σ-edges in its boundary oriented into or all out of f, depending on which side of σ the face f is on. The same is true of the τ-edges in ∂f. In case (1), let f be the unique boundary face that has all σ- and τ-edges in ∂f oriented into f. In cases (3) and (4), let f be the unique boundary face that has all τ-edges in ∂f oriented into f. Now, the vertex $*$ dual to f is a vertex of $\mathcal{T}$. In cases (1) and (3), every other vertex of $\mathcal{T}$ that is an even distance (in $\mathcal{T}$) from $*$ is dual to a face that is a sink region. (In the example of Figure 10.2 there are four such vertices, all a distance 2 from $*$. The four faces that they are dual to are shown shaded.) In case (4) every vertex of $\mathcal{T}$ that is an odd distance from $*$ is dual to a sink region. As our hypotheses prohibit sink regions, $\mathcal{T}$ is restricted accordingly. Thus σ and τ cannot intersect in case (4), and in cases, (1) and (3), if σ and τ intersect, they must do so as shown in Figure 10.1, where n is the valence of $*$.

In the instance of case (2) if σ and τ do intersect, there are two boundary faces f_1 and f_2 into which all σ- and τ-edges in their boundaries are inward-oriented. Let $*_1$ and $*_2$ be their dual vertices. It follows that $*_1$ and $*_2$ are an even distance apart in $\mathcal{T}$ and any there can be no other vertices in $\mathcal{T}$ that are an even distance from either. Thus $\mathcal{T}$ is the tree shown in Figure 10.1(2), with m and n being the valences of $*_1$ and $*_2$, and moreover, no other arrangement of $\mathcal{T}$ along σ is possible. $\square$

Corollary 10.3. *Suppose σ and τ are paths in a 2-disc D as per Definition 10.1, but we prohibit source regions instead of sink regions. If the order and relative orientations of σ and τ close to ∂D are as shown in Figure 10.2 (right), then σ and τ do not intersect.*

Proof. This is case (4) of Lemma 10.2, but with the orientations reversed. □

Our final lemma is the observation which says, roughly, that a pair of oriented paths through a disc that intersect transversely, can be "combined" to obtain a new transversely oriented such path, so that the original paths both lie to one side of the new path. This is illustrated in Figure 10.3, under the simplifying assumption that the intersections between the paths are transverse. The lemma allows subpaths as intersections, so it can be applied to (compound) tracks.

Lemma 10.4. *Suppose for $i = 1, 2$, an injective piecewise-linear path σ_i in a 2-disc D is made up of finitely many straight-line segments, and that σ_i meets ∂D at exactly 2 points, specifically its endpoints. Suppose σ_1 and σ_2 have transverse orientations. So, for $i = 1, 2$, there are subsets D_i^+ and D_i^- of D, each homeomorphic to a 2-disc, such that $D = D_i^+ \cup D_i^-$, and σ_i traverses the intersection of D_i^+ and D_i^- with σ_i oriented into D_i^+ and out of D_i^-. Assume σ_1 and σ_2 intersect in the interior of D. We allow the intersection of σ_1 and σ_2 to include (finitely many) straight line segments, provided their orientations agree on the common segments.*

Suppose there is a point $p \in \partial D$ that is in $D_1^+ \cap D_2^+$ and is not on σ_1 or σ_2. Let C_0^+ be the maximal connected open subset of D that contains p and does not intersect σ_1 or σ_2. Let C^+ be the closure of C_0^+ and C^- be $D \smallsetminus C_0^+$. Then C^+ and C^- are homeomorphic to 2-discs. Furthermore,

(1) C^+ contains p,
(2) $D_1^- \cup D_2^- \subseteq C^-$. In particular, σ_1 and σ_2 are in C^-, and
(3) an injective piecewise-linear path τ traverses $C^+ \cap C^-$, connecting two different points on ∂D. It is a concatenation of subpaths of σ_1 and σ_2, all oriented into C^+, and so has a well-defined orientation (into C^+).

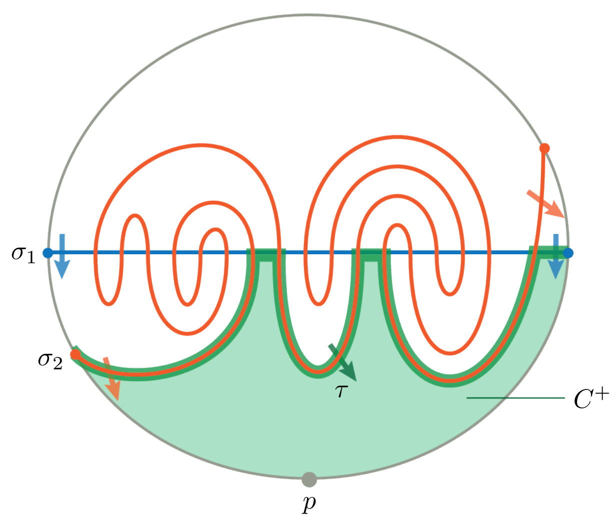


FIGURE 10.3. Lemma 10.4, illustrated.

11. Tracks in distortion diagrams

In Section 9 we established constraints on *reduced* van Kampen diagrams over our presentation $\mathcal{P}$ for G. Here, we will show that diagrams pertinent to the distortion of H in G are further constrained. The rigidity we will prove here and in Section 12 will allow us to calculate upper bounds on distortion in Section 13.

Definition 11.1. (Distortion diagrams, sides) A *distortion diagram* Δ is a reduced van Kampen diagram for $w\chi^{-1}$ over $\mathcal{P}$, where χ is a word on t, y_1, y_2 and w is a word on our generating set for G. Where no confusion should result, we refer to the portions of the boundary circuit $\partial\Delta$ that are labelled by w and by χ simply as w and χ. When an a- or b-track ρ connects two edges in $\partial\Delta$ those edges must both be in w, as there are no a- or b-letters in χ. So, as shown in Figure 11.1, the track ρ subdivides Δ into two subsets whose intersection is ρ. The subset that contains χ is the χ*-side* of ρ, and the other subset is the w*-side*.

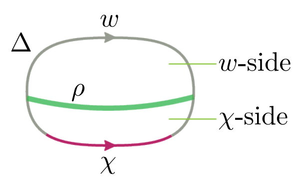
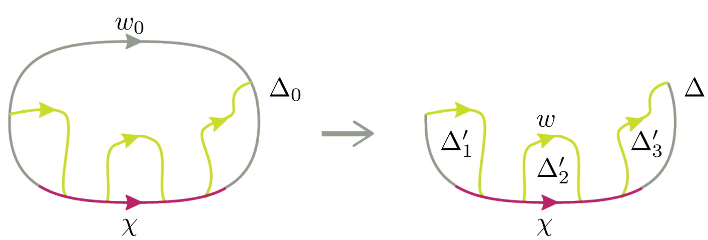


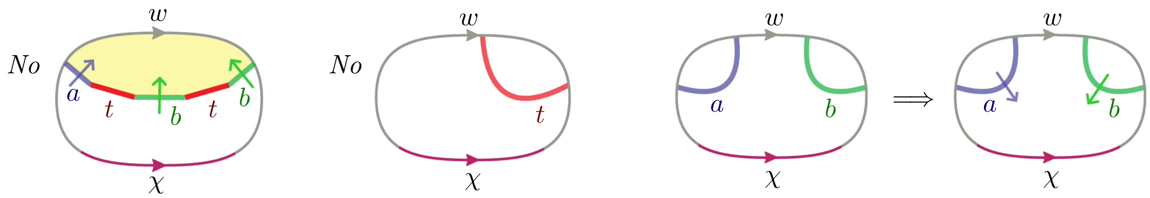
Figure 11.1. An a- or b-track ρ in a distortion diagram

Lemma 11.2. *(a- and b-tracks in distortion diagrams.) There exists $C > 0$ satisfying the following. Suppose w_0 is a word on the generators of G that equals in G a reduced word χ on t, y_1, y_2, and suppose Δ_0 is a distortion diagram for $w_0\chi^{-1}$. Assume that Δ_0 is homeomorphic to a 2-disc. Then there is a subdiagram Δ of Δ_0 that is a van Kampen diagram for $w\chi^{-1}$, where w is a word of length at most $C|w_0|$ and the following properties are satisfied.*

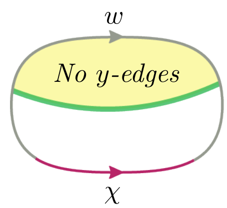
(0) The portions of $\partial\Delta$ labelled by w and by χ are both injective paths, so that Δ is a concatenation of paths and distortion diagrams $\Delta'_1, \ldots, \Delta'_r$, each homeomorphic to a 2-disc and each demonstrating that some subword of w equals some subword of χ (as shown on the right below).



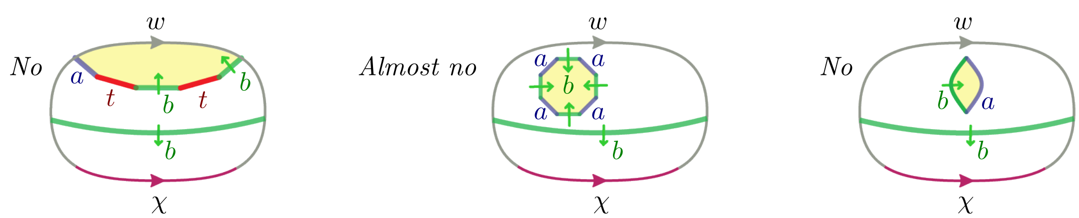
(1) No compound track in Δ between a pair of edges in w is made up of a-subtracks oriented towards w, b-subtracks oriented towards w, and t-tracks (oriented either way). In particular, no t-corridor in Δ connects two t-letters in w and every a- or b-track that connects a pair of edges in $\partial\Delta$ is oriented towards χ.



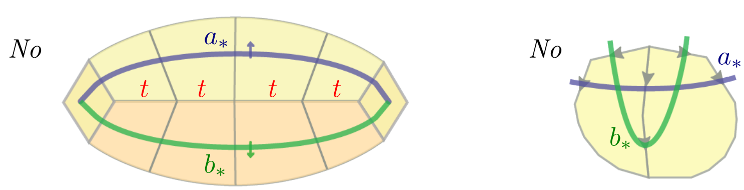
(2) There are no y-edges in the w-side of any b-track β that connects two edges in $\partial\Delta$.



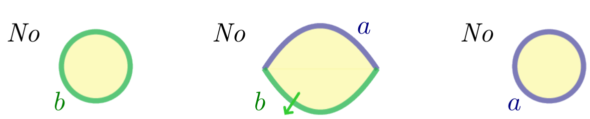
(3) Suppose a region R is a subset of the w-side of a b-track connecting two points in w.
   (a) ∂R cannot be comprised of a (non-trivial) subpath of the boundary circuit $\partial\Delta$, a-subtracks, inward oriented b-subtracks, and t-subtracks.
   (b) If ∂R is comprised of a-subtracks and inward-oriented b-subtracks, then it satisfies the constraints 2b–2d of Lemma 9.3. In particular, ∂R cannot be a bigon comprised of an a_1-subtrack and an inward oriented b-subtrack.



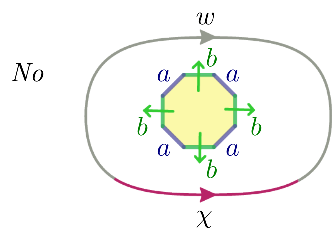
(4) Δ contains no badge and no button (Definitions 9.7 and 9.8).



(5) Δ has no a- or b-loops and no bigons comprised of an a-subtrack and an outward oriented b-subtrack.



(6) More generally, no region of Δ has boundary made up of consistently oriented *(meaning all inward- or all outward-oriented) a-subtracks and outward-oriented b-subtracks.*

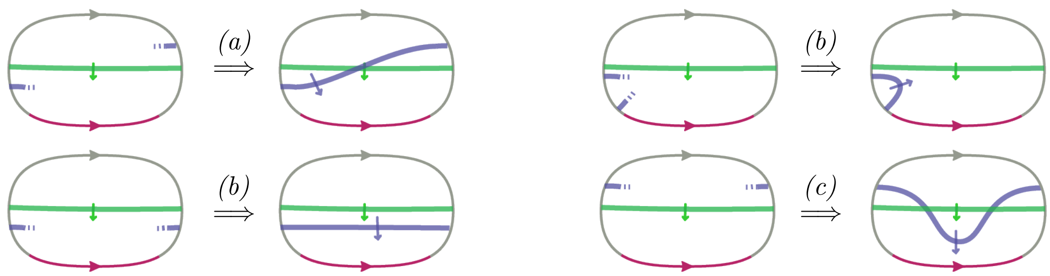


(7) Suppose α is an a_1-track and β is a b-track in Δ.

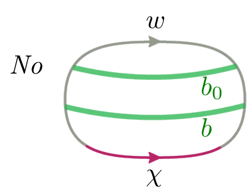
(a) If α has one endpoint on either side of β then α and β intersect exactly once.

(b) If both endpoints of α are on the χ-side of β, then α and β do not intersect.

(c) If both endpoints of α are on the w-side of β, then α and β intersect exactly twice.



(8) There can be no b_0-track $\beta_0 \neq \beta$ in the w-side of a b-track β.



Proof. We will sever parts of Δ_0 to obtain subdiagrams Δ_1, then Δ_2, and then Δ_3, that establish, respectively, (1), then (2), and then (3). Then we will sever parts of Δ_3 to get Δ such that the portion of $\partial\Delta$ labelled by w is an injective path, and we will argue that Δ satisfies *all* of (0)–(3). Then we will verify that Δ also satisfies (4)–(8).

For (1), define a *bad* path in Δ_0 to be a compound track connecting a pair of edges in w_0 comprised of a-and b-subtracks oriented towards w_0, and t-tracks (oriented either way). Let Δ_1 be the maximal subdiagram of Δ_0 that contains χ and intersects no bad path. Let w_1 be the word such that Δ_1 is a van Kampen diagram for $w_1\chi^{-1}$. If bad paths σ_1 and σ_2 intersect, then we may apply Lemma 10.4 with p a point on χ and σ_1 and σ_2 oriented *towards* χ, to obtain a new path τ which is a concatenation of subpaths of σ_1 and σ_2 (and therefore is again a bad path), such that both σ_1 and σ_2 are contained in the w_0-side of τ. Therefore there is a collection of bad paths $\tau_1, \ldots, \tau_m$ that are disjoint and are such that Δ_1 is the result of removing from Δ_0 the subdiagrams bounded by the corridors of 2-cells through which τ_i passes and by subwords of w_0. Now Lemma 9.16(1) tells us that there exists a constant $C_1 > 0$ such that $|w_1| \leq C_1|w_0|$.

For (2), we first establish that there exist disjoint b-tracks $\beta_1, \ldots, \beta_k$, each a path between two points in $\partial\Delta_1$, such that every b-track between two points in $\partial\Delta_1$ is on the w_1-side of β_i for some i. To see this, note that following (1), all b-tracks between pairs of points in $\partial\Delta_1$ are oriented towards

χ, and if two such b-tracks σ_1 and σ_2 intersect, then applying Lemma 10.4 with p a point on χ, we obtain a path τ connecting a pair of points on $\partial\Delta_1$, such that both σ_1 and σ_2 are on the w_1 side of τ, and τ is a concatenation of subtracks of σ_1 and σ_2, each oriented into the component of $\Delta_1 \setminus \tau$ containing χ. Since a concatenation of consistently oriented b-subtracks is again a b-subtrack, τ is again a b-track. The existence of $\beta_1, \ldots, \beta_k$ as above follows.

Thus, in constructing Δ_2 by severing parts of Δ_1, it suffices to guarantee that (2) holds for $\beta = \beta_i$ for each $1 \leq i \leq k$. Our argument in this case is illustrated by Figure 11.2.

By Lemma 9.4(1), there is no y-edge in any region R_i enclosed by a subpath of β and a t-subtrack on the w_1-side of β (such as regions R_1, R_2, and R_3 in Figure 11.2), as ∂R_i has no edges in this case. Define Δ'_β to be the maximal subdiagram of Δ_1 that is contained in the w_1-side of β and intersects no t-subtracks that start and end on β. Then Δ'_β is a van Kampen diagram for uv^{-1}, where u is a subword of w_1 and v is the word along the remainder of $\partial\Delta'_\beta$, as shown in Figure 11.2.

We will apply Lemma 9.15 to Δ'_β. Let us check the hypotheses. To see that there are no y-letters in v, observe that v is comprised of subpaths that run along the corridor associated to β, on the side that β is oriented away from, and subpaths that run along the sides of t-corridors. The defining relations of G (see Figure 4.1) imply that the first type of subpath cannot have any y-edges, and if there were a y-edge in a subpath of the second type, then then there would be one on the other side of the t-corridor also, and so in one of the regions R_i, a contradiction.

Next, we observe that all t-corridors in Δ'_β connect a t-edge in u to a t-edge in v. This is because there are no t-loops by Lemma 9.2; were there a t-track connecting a pair of edges in u, it would be a part (or whole) of a bad path in Δ_0, and would have been cut off in the construction of Δ_1; and no t-corridor joins pair of t-edges in v by construction.

Lemma 9.15 now implies that there is a constant $C_2 > 0$ (depending only on $\mathcal{P}$) and a word v' labeling a path in $\Delta'^{(1)}_\beta$ with the same endpoints as u and v with $|v'| \leq C_2|u|$ such that the subdiagram enclosed by v and v' has no y-edges. We now cut Δ'_β along v', discarding the subdiagram bounded by u and v'. As $\beta_1, \ldots, \beta_k$ are disjoint and non-nested, we do this independently for each $\beta = \beta_i$, resulting in a subdiagram Δ_2 of Δ_1 for a relation $w_2\chi^{-1}$, where w_2 is obtained from w_1 by replacing a disjoint collection of subwords with words whose lengths are greater by at most a factor of C_2. It follows that $|w_2| \leq C_2|w_1|$, and by construction, there are no y-edges on the w_2 side of β_i for any i. In particular, (2) holds for Δ_2.

Now suppose Δ_2 has a bad path σ—i.e., suppose that (1) fails for Δ_2. Since Δ_1 had none, σ must have at least one end on along a path labelled by one of the v', and this path is on the w side of some β which is oriented towards χ. If σ intersects β at least twice, then, since β is oriented towards χ, a subtrack of β and a subpath of σ together bound a region R that is precluded by Lemma 9.4 (see Figure 9.2). If σ crosses β exactly once, then a subpath of β, together with the part of σ on the χ side of β form a bad path (in the sense of (1)) in Δ_2, which is not possible. Thus any bad path σ in Δ_2 lies on the w_2 side of β. Such paths will be removed next, in the construction of Δ_3.

For (3a), define a region R to be *bad* if it is of the form (3a) excludes: that is, R is a subset of the w_2-side of a b-track β connecting two edges in w and ∂R is comprised of a non-trivial subpath of the boundary circuit $\partial\Delta_2$ and a compound track consisting of a-subtracks, inward oriented b-subtracks, and t-subtracks. We may assume that β is one of the tracks $\beta_1, \ldots, \beta_k$ identified above, which persist in Δ_2. Here are two key observations:

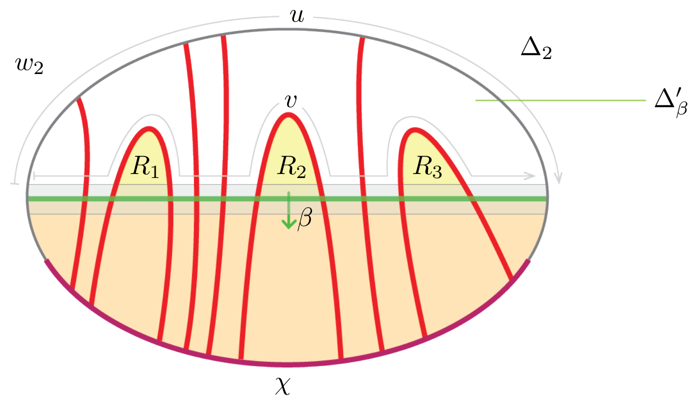


FIGURE 11.2. Subdiagrams and t-tracks per our proof of Lemma 11.2(2)

i. If two bad regions R_1 and R_2 have intersecting interiors, they are on the w_2-side of a common b-track, say β_i. Then, applying Lemma 10.4 to the compound tracks in ∂R_1 and ∂R_2, we get a new bad region R_3 containing $R_1 \cup R_2$ that is again on the w_2-side of β_i.

ii. Suppose R is a bad region on the w-side of a b-track β. Then the minimal subdiagram D of Δ_2 containing R contains no y-edges. To see this, note that no subpath of β can contribute to ∂R, as β is oriented towards χ, and so no 2-cell through which β passes can be in D. Thus D is a subset of the w-side of β and has no y-edges by (2).

Define Δ_3 to be the maximal subdiagram of Δ_2 that includes χ and does not intersect any bad region. On account of (i), Δ_3 is obtained from Δ_2 by severing a finitely many subdiagrams D per Lemma 9.16 by, in the notation of that lemma, cutting along the paths labelled u_1. Moreover, any two of these D have disjoint interiors and the associated words u_0 label paths in $\partial \Delta_2$ that are non-overlapping (but can share endpoints). By (ii), hypothesis (2) of Lemma 9.16 holds and we can apply that lemma to each of these D. Let w_3 be the word such that Δ_3 is a van Kampen diagram for $w_3\chi^{-1}$. The inequality in Lemma 9.16 then tells us that there exists a constant $C_3 > 0$ such that $|w_3| \leq C_3|w_2|$. Finally, Δ_3 satisfies conditions (1)–(3): as shown above, the only paths that could fail (1) were removed in the construction of Δ_3; (2) is immediately inherited from Δ_2; (3a) is satisfied by construction; and, in light of (2), Lemma 9.3 implies (3b).

If the portion of $\partial \Delta_3$ labelled by w_3 is not an injective path, then some subword labels a subdiagram which is only attached to the rest of Δ_3 at a single vertex. We sever all subdiagrams that so arise, so as to produce a van Kampen diagram Δ for a word $w\chi^{-1}$, with $|w| \leq |w_3|$, such that conditions (1)–(3) hold, and the portion of $\partial\Delta$ labelled by w is an injective path. By hypothesis, χ is a reduced word on t, y_1, y_2, which freely generate a free subgroup of G by Corollary 7.5, so χ also labels an injective path in $\partial\Delta$. So Δ is a concatenation of paths and distortion diagrams $\Delta', \ldots, \Delta'_r$, each homeomorphic to a 2-disc and each demonstrating that some subword of w equals some subword of χ. This establishes (0). Further, if we let $C = C_1C_2C_3C_3$, then our inequalities combine to give $|w| \leq C|w_0|$, as required.

For the remainder of the proof, we assume, for convenience, that Δ is homeomorphic to a 2-disc. The proofs of (4)–(8) in the general case follow easily. (In cases (a) and (c) of (7) the hypothesis forces α and β to be in the same component. In case (b), the result is automatic if they are in different components.)

(4). Suppose there is a badge or button $\mathcal{B}$ in Δ. Per Definition 6.3, let $\mathcal{G}_b$ be the graph whose edges are the duals of the b-edges in Δ. Let $\mathcal{C}$ be the connected component of $\mathcal{G}_b$ that includes the b-track through $\mathcal{B}$. Let i be minimal such that $\mathcal{C}$ includes the dual of a b_i-edge. A b-track that enters a 2-cell across a b_i-edge can exit across another b_i-edge unless that 2-cell is an $r_{1,i-1}$-cell. So the minimality of i ensures that $\mathcal{C}$ contains a b_i-track β. By Corollary 9.10(3), β is not a loop, and so it connects two b_i-edges in w, and is oriented towards χ by (1). So no b-tracks branch off β on its χ-side and, in particular, the b-tracks through $\mathcal{B}$ are on its w-side. (They can have subpaths in common with β.) By (2), there are no y-edges on the w-side of β. This ensures that $\mathcal{B}$ is not a badge, as if it were, it would have an $r_{4,i}$-cell contributing a y-edge to the w-side of β.

Any a_1-track intersecting the w-side of β intersects β exactly once—it is not a loop on the w-side of β (by (2) and Corollary 9.10(2)), it is dual to at most one edge in $\partial\Delta$ (by (3a)), and it intersects β at most once, for if it formed a bigon with β, then Lemma 9.9(2) would apply to an innermost such instance α, and one of the intersections of α and β would have to occur in an $r_{1,i-1}$-cell, contradicting the minimality of i.

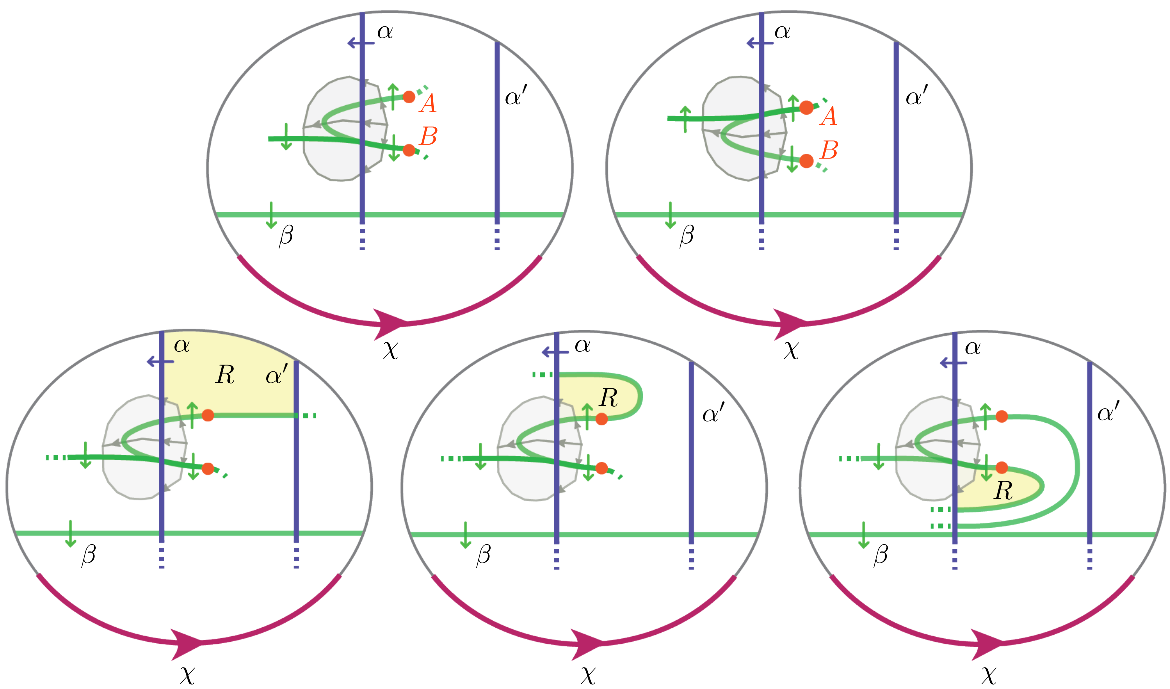


FIGURE 11.3. Cases in our proof of Lemma 11.2(4)

Let α be the a_1-track through $\mathcal{B}$, which we now know to be a button. Figure 11.3 (top-left and top-right) shows the two possible placements of $\mathcal{B}$ along α, once we assume, without loss of generality, that α is oriented towards the left (in the sense of the figure). Let A and B be the points shown (in either case). Let α' be the first a_1-track one meets on following β to the right (in the sense of the figure) from its intersection with α. (If there is no such α' a simpler version, which we omit, of the following analysis will apply.) Then $\mathcal{G}_b$ can have no junction in the (closed) region bounded by α (on the left), α' (on the right), β (below), and a portion of $\partial\Delta$ (above), as this region has no a_1-tracks. Thus there are three possible continuations for the b-track at A through this region: (i) it continues to α' or to $\partial\Delta$ (as shown lower left in Figure 11.3); (ii) it returns to

α *above* the button (as shown lower middle); and (iii) it returns to α *below* the button (as shown lower right). In case (iii), it must return to α *below* the button also (as otherwise there would be a junction). In all cases (i)–(iii) there is a region R (shown shaded in the figure) with boundary made up of an inward-oriented b-subtrack, a_1-subtracks, and (in case (i)) a portion of the w-part of $\partial\Delta$, contrary to (3) of this lemma.

(5). In light of (4), Lemma 9.9(1) and Corollary 9.10(3) preclude bigons comprised of an a-subtrack and an outward oriented b-subtrack and b-loops respectively. Corollary 9.10(1) precludes inward oriented a-loops. Suppose, for a contradiction, that there exists a non-trivial a-loop α. Then the region R enclosed by α cannot contain a b-subtrack, as such a subtrack would give rise to a teardrop, a b-loop, or a bigon comprised of an outward oriented b-subtrack and an a-subtrack, all of which have been ruled out. It follows that the minimal subdiagram containing R contains only cells of type $r_{4,*}$ (as any other cells with a-letters would introduce b-subtracks), which contradicts Lemma 9.6(1).

(6). Suppose, for a contradiction, that R is a region of Δ whose boundary is comprised of a-subtracks and outward-oriented b-subtracks. We may assume that no a- or b-track intersects the interior of R, because such a track would subdivide R into two regions, at least one of which would satisfy the hypotheses of (6).

By (5), ∂R cannot be an a- or b-loop or a bigon comprised of an a-track and an outward-oriented b-track. Any two adjacent b-subtracks in the circuit ∂R are together a single b-subtrack. As the a-subtracks in ∂R are consistently oriented, the same is true for a-subtracks. So, ∂R is a concatenation of non-trivial paths $\overline{\alpha}_1$, $\overline{\beta}_1$, $\ldots$, $\overline{\alpha}_m$, $\overline{\beta}_m$ where $m \geq 2$ and each $\overline{\alpha}_i$ is a subtrack of some a-track α_i, each $\overline{\beta}_i$ is a subtrack of some b-track β_i and the $\overline{\beta}_i$ are all oriented out of R.

As $\overline{\beta}_1$ is oriented out of R and its continuation β_1 is oriented toward χ (by (1)), R is in the w-side of β_1. Now, because β_2 is also oriented toward χ, and because the interior of R has no b-subtracks, β_2 must merge with β_1 either to the left or right of R, as shown in Figure 11.4. Then some subtrack of β_2 bounds a region R' either with $\overline{\alpha}_2$ (per Figure 11.4, left) or with the concatenation $\overline{\alpha}_3\overline{\beta}_3 \cdots \overline{\alpha}_m\overline{\beta}_m\overline{\alpha}_1$ (per Figure 11.4, right). In the latter case, the extension α_1 of $\overline{\alpha}_1$ cannot enter R (as R contains no a-subtracks) so must meet the part of β_2 in $\partial R'$ (after possibly passing through some other $\overline{\alpha}_i$'s for $3 \leq i \leq m$). In either case we get a bigon B bounded by an a-subtrack and an outward-oriented b-subtrack, contrary to (5).

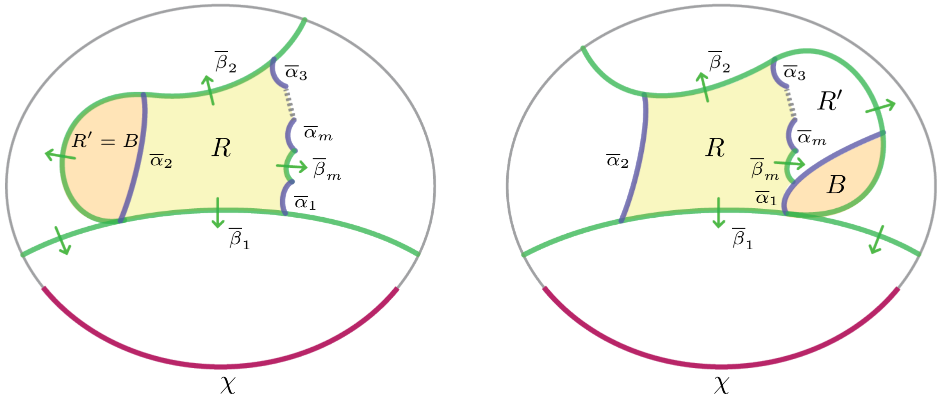


FIGURE 11.4. Illustrating our proof of Lemma 11.2(6)

(7). We will use Lemma 10.2 with $\{\tau, \sigma\} = \{\alpha, \beta\}$. Lemma 9.4(2) tells us that there is no region in Δ that is bounded by inward-oriented a-and b-subtracks, which establishes the no-sink-regions hypothesis of Lemma 10.2.

The case (7a) corresponds to case (1) of Lemma 10.2 with $\tau = \alpha$ and $\sigma = \beta$. By (1) of the present lemma, α and β are oriented towards χ. So (7b) corresponds to either case (2) or case (3) of Lemma 10.2 with $\tau = \alpha$ and $\sigma = \beta$, and (7c) concerns case (3) with $\tau = \beta$ and $\sigma = \alpha$. With just one exception, (5) of the present lemma (specifically the part concerning bigons) rules out all the intersection patterns catalogued in Lemma 10.2 apart from those listed in the conclusion of (7). That one exception occurs in (7c), where we need to further exclude the possibility that α and β do not cross, which we do by invoking (3) of this lemma.

(8). Suppose there is a b_0-track β_0 in the w-side of a b-track β. If C is a 2-cell dual to β_0, then C has a y-edge, so cannot be on the w-side of β by (2). Thus C is dual to β as well, and is an $r_{*,0}$-cell. Then β agrees with β_0 on C. (This is clear if $*$ is 2 or 3. If C has type $r_{4,1}$, it follows from the fact that β and β_0 are oriented towards χ by (1).) Consequently, $\beta_0 = \beta$. ☐

12. (a_2, b_q)-TRACKS

A key idea leading to the "p/q" in the subgroup distortion function of Theorem 1.1 is that the generation of b_p letters within distortion diagrams is offset by generation of letters b_q that must "appear" in w either as b_q-letters or in the guise of a_2-letters. The reason for this is that b_q letters feature in (a_2, b_q)-*tracks*, which are the subject of this section and will be crucial to our proof of Lemma 13.12.

Definition 12.1. ((a_2, b_q)-tracks) An (a_2, b_q)-*track* in a van Kampen diagram Δ over our presentation $\mathcal{P}$ for G is a maximal path that is a concatenation of edges dual to consistently oriented a_2-edges and b_q-edges in Δ, such that an (a_2, b_q)-track entering a 2-cell of the form shown rightmost in Figure 12.1 across an a_2-edge leaves across the consistently oriented b_q-edge. The two (a_2, b_q)-tracks in the 2-cell shown rightmost in Figure 12.1 touch, but we do not consider them to intersect. Examples are shown in Figures 3.1 and 3.3.

Lemma 12.2. *(a_2, b_q)-tracks in a van Kampen diagram Δ have the following properties:*

(1) (a_2, b_q)-tracks inherit orientations from the orientations of their constituent subtracks.
(2) Every a_2-edge and b_q-edge in Δ is dual to an edge in exactly one (a_2, b_q)-track.
(3) An (a_2, b_q)-track cannot intersect itself or another (a_2, b_q)-track.
(4) The set of a_2- and b_q-edges in $\partial\Delta$ are paired off according to whether there is an (a_2, b_q)-track whose first and last edges are dual to them.
(5) If Δ is a distortion diagram as constructed in Lemma 11.2, then an (a_2, b_q)-track in Δ cannot be a loop.

Proof. (1) holds because constituent subtracks are consistently oriented edges by construction.

With the sole exception of $r_{2,q}$ (shown rightmost in Figure 12.1), all our defining relators contain either none of the letters a_2, a_2^{-1}, b_q, and b_q^{-1}, or contain exactly one of a_2 and b_q, and exactly one of a_2^{-1} and b_q^{-1}. So (2)–(4) follow. For (5), suppose there is a (a_2, b_q)-loop in a distortion diagram Δ. As the orientations of its constituent subtracks are consistent, it is either inward- or outward-oriented. The former is impossible by Lemma 9.4(2) and the latter by Lemma 11.2(6). ☐

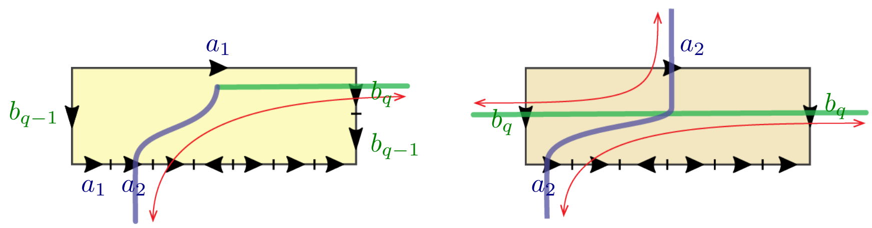


FIGURE 12.1. *How (a_2, b_q)-tracks progress through $r_{1,q-1}$- and $r_{2,q}$-cells*

Given Δ as per Lemma 11.2, its b_0-tracks $\beta_1, \ldots, \beta_m$ must be arranged consecutively around $\partial\Delta$ as per Figure 12.2 (since they cannot nest by Lemma 11.2(8)). In short, our next lemma states that the intersections of an (a_2, b_q)-track with the b_0-tracks in Δ progress in order around the diagram. We will use it in our proof of Proposition 13.1 at the end of Section 13.

Lemma 12.3. *Suppose Δ is a distortion diagram for $w\chi^{-1}$ as per Lemma 11.2. Let Q_0 and P_{m+1} be the initial and terminal vertices of the w portion of $\partial\Delta$. For distinct points P and Q on w, write $P < Q$ when one reaches P first when following w from Q_0 to P_{m+1}. Suppose, as shown in Figure 12.2, $P_1 < Q_1 < \cdots < P_m < Q_m$ are $2m$ successive points on the w-portion of $\partial\Delta$ and, for $i = 1, \ldots, m$, β_i is a b_0-track from P_i to Q_i oriented towards χ. Let R be the maximal region of Δ that is bounded by $\beta_1, \ldots, \beta_m$ and the intervening subpaths of $\partial\Delta$.*

Suppose τ is an (a_2, b_q)-track in Δ starting at some P and ending at some Q in $\partial\Delta$, with $P < Q$. Let Σ be the set of points where τ meets ∂R. The order in which τ visits the points of Σ as it progresses from P to Q is the same as the order in which they occur on the boundary circuit ∂R starting from Q_0 and following it around to P_{m+1}.

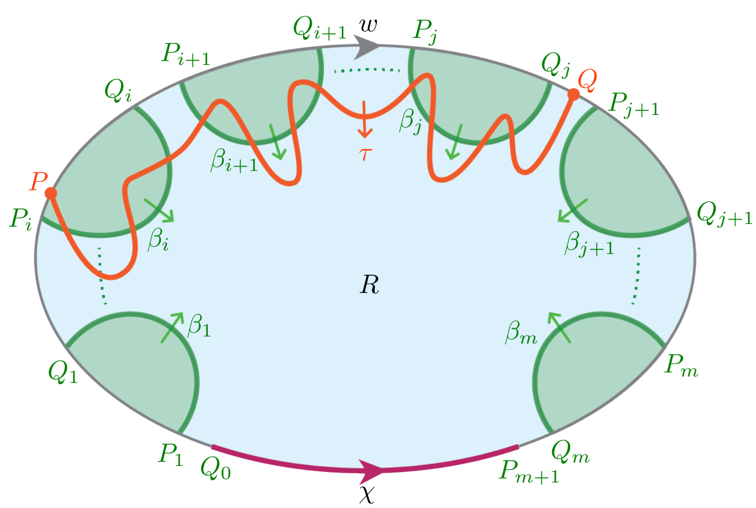


FIGURE 12.2. Illustrating Lemma 12.3

Proof. As its constituent a_2- and b_q-subtracks are, by construction, consistently oriented, τ is a compound track which is oriented either towards or away from χ. The latter eventuality is precluded by Lemma 11.2(1).

The lemma will be proved by applying either Lemma 10.2 or Corollary 10.3 to pairs consisting of τ (or a subpath thereof) and β_l, for each l.

Let $i, j \in \{0, \ldots, m+1\}$ be such that $Q_{i-1} < P < Q_i$ and $P_j < Q < P_{j+1}$. By Lemma 11.2(6), for all ℓ, there is no source-region bounded by subtracks of τ and β_ℓ. If $\ell < i$ or $\ell > j$, then the orientations of β_ℓ and τ near $\partial\Delta$ are as shown in Figure 10.2(right), so β_ℓ and τ cannot intersect by Corollary 10.3.

Consider traveling along τ from P to Q. If τ intersects β_k for some k, then τ cannot intersect any β_ℓ with $\ell < k$. This is because were there such an ℓ, there would be a subpath $\hat{\tau}$ of τ that connects a pair of points on $\beta_k \cup \{Q\}$ and intersects β_ℓ. However, in the disc obtained from Δ by excising the w-side of β_k, the orientations on $\hat{\tau}$ and β_ℓ are as shown in Figure 10.2(right), so this intersection is contrary to Corollary 10.3.

So τ intersects none of $\beta_1, \ldots, \beta_{i-1}, \beta_{j+1}, \ldots, \beta_m$ and, proceeding from P, it intersects $\beta_i, \beta_{i+1}, \ldots, \beta_j$ in order (intersecting each some number of times, possibly zero). If $P_i < P < Q_i$, then how τ intersects β_i is described by case (1) of Lemma 10.2. The other possibility is that $P < P_i$, which is handled by case (3). Case (3) likewise describes how τ intersects $\beta_{i+1}, \ldots, \beta_{j-1}$, and case (1) or (3) how τ intersects β_j. These observations combine to prove the result. $\square$

13. The upper bound on distortion

Modulo calculations we will postpone to Section 14, we will prove here:

Proposition 13.1. *For χ, w and Δ as per Lemma 11.2, there exists a constant $K > 1$, depending only on our presentation $\mathcal{P}$ for G, such that*

$$|\chi| \;\leq\; K^{|w|^{p/q}}. \tag{13.1}$$

As a corollary, we obtain the desired upper bound on distortion:

Corollary 13.2. $\mathrm{Dist}_H^G(n) \preceq \exp(n^{p/q})$.

Proof of Corollary 13.2, assuming Proposition 13.1. Suppose $n \geq 0$. Let χ be a reduced word on the generators of H which realizes the distortion function of H, i.e.:

$$\mathrm{Dist}_H^G(n) \;=\; |\chi|. \tag{13.2}$$

More precisely, χ is a maximal length reduced word on the generators of H that equals, in G, some word w_0 of length at most n. We can assume w_0 has no subwords representing the identity in G.

Let Δ_0 be a reduced van Kampen diagram for $w_0\chi^{-1}$. If Δ_0 is homeomorphic to a 2-disc, then Lemma 11.2 and hence Proposition 13.1 apply, yielding w such that $|\chi| \;\leq\; K^{|w|^{p/q}}$ and $|w| \leq C|w_0|$. This, combined with (13.2) and $|w_0| \leq n$ gives the result.

Now suppose that Δ_0 is not a 2-disc. Our choice of w_0 guarantees that no two vertices along the part of $\partial\Delta_0$ labelled w_0 are identified. The same holds for χ, as it is reduced. It follows that w_0 and χ are concatenations of subwords $w_1, w_2, \ldots, w_r$ and $\chi_1, \chi_2, \ldots, \chi_r$ respectively, such that for each i, either $w_i = \chi_i$ and the paths with these labels along $\partial\Delta_0$ are identified, or there is a (reduced) subdiagram Δ_i of Δ_0 homeomorphic to a 2-disc whose boundary reads $w_i\chi_i^{-1}$. In either case, we have $\chi_i \;\leq\; K^{|w_i|^{p/q}}$, and the bound we require follows from the superadditivity of the function $n \mapsto \exp(n^{p/q})$. $\square$

Let χ, w, and Δ be as per Lemma 11.2. To prove Proposition 13.1, we will decompose Δ into the subdiagrams we now define.

Definition 13.3. (**Decomposing a distortion diagram into b-blocks and an a-block.**) Given a b-track β in Δ, define Δ_β to be the minimal subdiagram of Δ containing the w-side of β (see Definition 11.1). So Δ_β is comprised of all the 2-cells of Δ that either have β passing through them or are in the w-side of β. Say that β is *outermost* when there is no b-track β' such that $\Delta_{\beta'}$ properly contains Δ_β. The Δ_β such that β is outermost are the *b-blocks* of Δ.

Let $\mathcal{B}_1, \ldots, \mathcal{B}_r$ be the b-blocks of Δ as per Figure 13.1 (when $r = 3$). Define the *a-block* $\mathcal{A}$ of Δ to be the maximal subdiagram of Δ that contains χ and intersects no b-tracks. So $\mathcal{A}$ is obtained from Δ by severing $\mathcal{B}_1, \ldots, \mathcal{B}_r$.

Corollary 13.4. *For $\mathcal{A}$ and $\mathcal{B}_1, \ldots, \mathcal{B}_r$ as defined above—*

(1) $\mathcal{A}$ is a subdiagram of Δ whose 2-cells are of type $r_{4,*,*}$ and $r_{4,*}$ (per Figure 4.1).
(2) $\mathcal{B}_1, \ldots, \mathcal{B}_r$ are subdiagrams of Δ whose 2-cells are of type $r_{1,*}$, $r_{2,*}$, $r_{3,*}$, and $r_{3,*,*}$.
(3) For all i, there exists j_i such that the outermost b-track β_i of $\mathcal{B}_i$ is a b_{j_i}-track. It is oriented towards χ and the cells of Δ that it traverses comprise a b_{i_j}-corridor in $\mathcal{B}_i$ whose top boundary (the boundary the b_{j_i}-edges are oriented towards) follows $\mathcal{A} \cap \mathcal{B}_i$. If $j_i = 0$, then this is the only b_0-corridor in $\mathcal{B}_i$.

Proof. Lemma 11.2(2) implies that the b-blocks contain no $r_{4,*}$- or $r_{4,*,*}$-cells. Statements (1) and (2) are then consequences of the definitions of the a- and b-blocks. Lemma 11.2(1) tells us that every b-track is oriented towards χ. Part (3) then follows, except we also invoke Lemma 11.2(8) for its final claim. $\square$

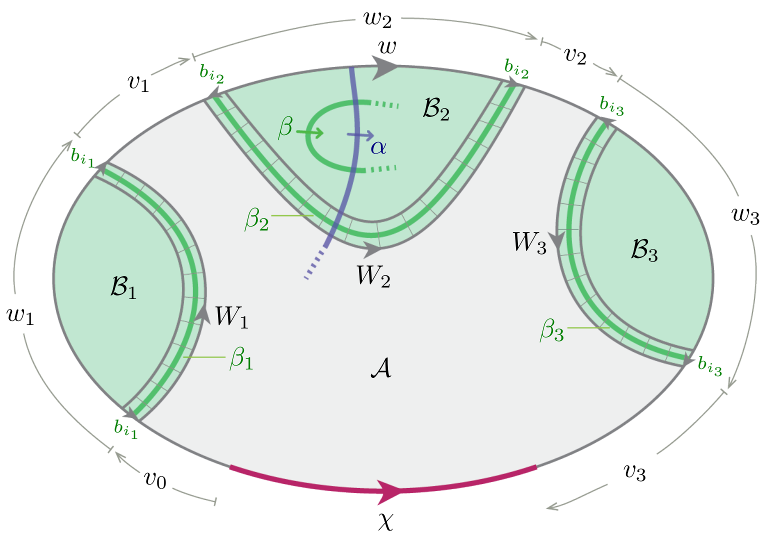


FIGURE 13.1. The a-block and b-blocks in Δ. The a_1-track α and b-track β illustrate a case in the proof of Lemma 13.6.

Express w as the concatenation of words

$$w \;=\; v_0 w_1 v_1 w_2 \cdots w_r v_r$$

where, for all i, w_i is the word along $\partial\mathcal{B}_i \cap \partial\Delta$ as shown in Figure 13.1 and the v_i are the (possibly empty) intervening subwords. Per Corollary 13.4(3), each w_i has first letter $b_{j_i}^{-1}$, final letter b_{j_i}, and the intervening subword is a word on $b_1^{\pm1}, \ldots, b_p^{\pm1}$. For all i, let W_i be the word along the other side of $\mathcal{B}_i$, so that $\mathcal{B}_i$ is a van Kampen diagram for $w_i W_i^{-1}$. Let

$$W \;=\; v_0 W_1 v_1 W_2 \cdots W_r v_r. \tag{13.3}$$

So $\mathcal{A}$ is a van Kampen diagram for $W\chi^{-1}$.

In the following lemmas we analyze the structure of a b-block $\mathcal{B}_i$ in Δ. When β_i is a b_0-track, this will lead (in Lemma 13.13) to an upper bound on the length of W_i.

Lemma 13.5. *Let $\mathcal{B}_i$ be a b-block of Δ, and let w_i and W_i be as above. Then every a_1-track in $\mathcal{B}_i$ runs from an $a_1^{\pm1}$ in w_i to an $a_1^{\pm1}$ in W_i.*

Proof. Let α be an a_1-track of Δ intersecting $\mathcal{B}_i$. It cannot be a loop by Lemma 11.2(5). It must have at least one endpoint in the w-side of β_i by Lemma 11.2(7a). If it has one endpoint on each side of β_i, then it intersects β_i exactly once by Lemma 11.2(7b), and so corresponds to a single a_1-track of $\mathcal{B}_i$ running from w_i to W_i. If it has both endpoints in the w-side of β_i, then, by Lemma 11.2(7c), it intersects β_i exactly twice, giving rise to two a_1-tracks in $\mathcal{B}_i$ both running from w_i to W_i. ☐

Lemma 13.6. *Suppose β_i is a b_0-track. Let $\mathcal{C}$ be an a_1-corridor of $\mathcal{B}_i$. The bottom boundary of $\mathcal{C}$ is labelled (in the direction from w_i to W_i) by a word λb_0, where λ is a positive word on $b_1, \ldots, b_p$.*

Proof. By Lemma 13.5, $\mathcal{C}$ has one end in w_i and the other in W_i. By Corollary 13.4(2),(3), the cells of $\mathcal{C}$ are of type $r_{1,*}$ (per Figure 4.1), and only the cell where $\mathcal{C}$ meets W_i has an edge labelled b_0, so that the bottom boundary of $\mathcal{C}$ (in the direction from w_i to W_i) is labelled by a word λb_0 where λ is a word on $b_1^{\pm1}, \ldots, b_p^{\pm1}$. We will argue that λ is a *positive* word. Suppose, for a contradiction, that λ includes a letter b_j^{-1} for some j. Let β be any b-track that has an edge dual to the edge of $\partial\mathcal{C}$ labelled by that b_j^{-1}. Let α be the a_1-track dual to $\mathcal{C}$. By Lemma 11.2(1), β is oriented towards χ, and so β intersects α at least one more time. So α and β form a bigon. This leads to a contradiction: that bigon violates (3b) or (5) of Lemma 11.2, depending on whether β is oriented into or out of the bigon, respectively. (The (3b) case is illustrated in Figure 13.1.) We conclude that λ is a positive word on $b_1, \ldots, b_p$. ☐

Our next lemma is illustrated by Figure 13.2.

Lemma 13.7. *Given $\mathcal{B}_i, \mathcal{C}$, and λ as in Lemma 13.6, the side of $\mathcal{C}$ labelled by λb_0 divides $\mathcal{B}_i$ into two subdiagrams. Of these two subdiagrams, let Λ_0 be that which does not contain $\mathcal{C}$. Its boundary word is $\tilde{\mu} b_0 \nu (\lambda b_0)^{-1}$, where ν and $\tilde{\mu}^{-1}$ are, respectively, some prefix of (W_i or W_i^{-1}) and of (w_i or w_i^{-1}). (Which of these pairs it is depends on the orientation of $\mathcal{C}$. Figure 13.2 shows the case where they are prefixes of W_i and w_i.) Let Λ_1 be the maximal subdiagram of Λ_0 that contains portions of $\partial\Lambda_0$ coming from λb_0 and $\hat{W}_i$, but intersects no b-track in Λ_1 that connects a pair of edges in the $\tilde{\mu}$ portion of $\partial\Lambda_0$. (See Figure 13.7.) Let $\hat{\mu}$ be the word such that $\hat{\mu} b_0 \nu (\lambda b_0)^{-1}$ is the word read around $\partial\Lambda_1$. Then:*

1. *The a_1-tracks in Λ_1 all arise from removing initial subtracks from a_1-tracks in Λ_0. In particular, each runs from an $a_1^{\mp1}$ in $\hat{\mu}$ to an $a_1^{\pm1}$ in ν, and the number of $a_1^{\pm1}$-letters in $\hat{\mu}$ is at most the number in $\tilde{\mu}$, and therefore at most $|w_i|$.*
2. *In $\hat{\mu}$ there are no letters $b_0^{\pm1}, b_1^{-1}, \ldots, b_p^{-1}$ and*
3. *There are at most $|\tilde{\mu}|$ letters $b_1, \ldots, b_p$ in $\hat{\mu}$.*

(4) The word read along the bottom boundary (in the direction from $\hat{\mu}$ to ν) of a corridor dual to an a_1-track in Λ_1 is a positive word on $b_0, b_1, \ldots, b_p$. Moreover, it has only one b_0, namely its final letter.

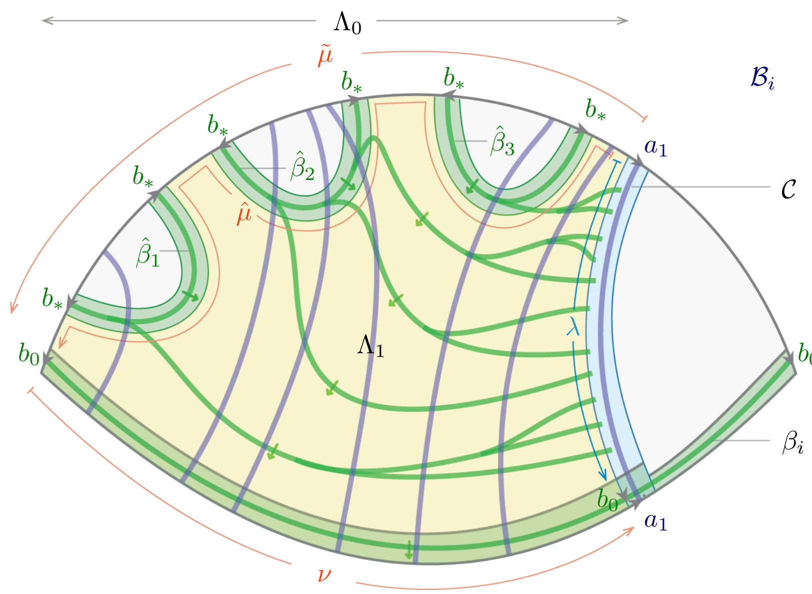


Figure 13.2. Illustrating our proof of Lemma 13.7

Proof. There are no letters $b_0^{\pm 1}$ in $\tilde{\mu}$ by construction. If there is a b_r^{-1} in $\tilde{\mu}$ for some $1 \leq r \leq p$, then it is connected by a b-track to some letter b_r labeling an edge in $\partial\Lambda_0$—in fact, that b_r must be in $\tilde{\mu}$, because there are no b_r^{-1} letters in λb_0 (by Lemma 13.6) or in ν (such are the 2-cells in b_0-corridors). By Lemma 11.2(1), all such b-tracks are oriented towards χ in Δ, and so towards ν in Λ_0. So there are such b-tracks $\hat{\beta}_1, \ldots, \hat{\beta}_k$ (in Figure 13.2 they are shown with $k = 3$) in Λ_0 that we might call *outermost* in that

- the w-sides of any two of them are disjoint,
- every such b-track is in the w-side of one of $\hat{\beta}_1, \ldots, \hat{\beta}_k$.

Then Λ_1 is obtained from Λ_0 by cutting along the top boundaries of the corridors $C_{\hat{\beta}_1}, \ldots, C_{\hat{\beta}_k}$ dual to $\hat{\beta}_1, \ldots, \hat{\beta}_k$.

So (1) then follows from Lemma 13.5 and the observation that, by Lemma 11.2(5), no a_1-track can cross one of the $\hat{\beta}_j$ twice.

For (2) and (3), we examine the b-letters in $\hat{\mu}$. Those that arise as letters in $\tilde{\mu}$ include no $b_0^{\pm 1}, b_1^{-1}, \ldots, b_p^{-1}$ by construction. Each of the other $b_l^{\pm 1}$ in $\hat{\mu}$ arises on the top boundary of one of the $C_{\hat{\beta}_j}$ at some 2-cell of type $r_{1,l}$ (per Figure 4.1) where some other b-track branches off $\hat{\beta}_j$. There are no b_0-edges in Λ_1 except in the b_0-corridor abutting ν—for otherwise there would be an additional b_0-corridor and therefore a $b_0^{\pm 1}$ in $\tilde{\mu}$ or λ, which is not so. So $1 \leq l \leq p-1$. In fact, the letter cannot be a b_l^{-1} because then there would be a b-track that initially follows $\hat{\beta}_j$ until branching off into Λ_1 and eventually terminates back on $\tilde{\mu}$ (not on λ because λ is a positive word), so as to contradict $\hat{\beta}_1, \ldots, \hat{\beta}_k$ being outermost. This proves (2). Then, for (3), observe that each 2-cell of type $r_{1,*}$ in $C_{\hat{\beta}_j}$ has a different a_1-track passing through it which, in light of (1), connects to an a_1-edge in $\tilde{\mu}$ between the between the endpoints of $\hat{\beta}_j$.

Finally, Lemma 13.6 implies (4). □

We will use the conclusions of Lemma 13.7 to further analyze λ via calculations in

$$(13.4) \qquad Q \;=\; \langle a_1, b_0, \ldots, b_p \mid a_1^{-1} b_i a_1 = \varphi(b_i) \;\; \forall i \;\rangle \;\; \text{with} \;\; \varphi(b_j) \;=\; \begin{cases} b_{j+1} b_j & \text{for } j < p \\ b_j & \text{for } j = p, \end{cases}$$

which is a free-by-cyclic quotient of G via the map $G \twoheadrightarrow Q$ killing a_2, t, x_1, x_2, y_1, and y_2.

Our next simplifying step, in Lemma 13.10, will dispense with the positive a_1-letters from $\hat{\mu}$. But first, we need two technical results concerning Q:

Lemma 13.8. *Suppose u and v are positive words on $b_0, \ldots, b_p$. Take $\varphi^{-1}(u)$ to denote the reduced word on $b_0, \ldots, b_p$ representing that element of Q. Then $\varphi^{-1}(u)v$ is reduced—that is, there is no cancellation between $\varphi^{-1}(u)$ and v. In particular, if w is a positive word on $b_0, \ldots, b_p$ which equals $\varphi^{-1}(u)v$ in Q, then v is a suffix of w.*

Proof. We downwards induct on the minimal index i such that u includes a letter b_i. If $i = p$, the result holds because u is a power of b_p and $\varphi^{-1}(u) = u$. For the induction step, write u as the concatenation $u_0 u_1$, where u_0 ends in b_i, and u_1 contains no b_i.

It can be checked that for $j = 0, \ldots, p$,

$$\varphi^{-1}(b_j) \;=\; \begin{cases} b_{j+1}^{-1} \cdots b_{p-3}^{-1} b_{p-1}^{-1} b_p \cdots b_{j+2} b_j & \text{when } p - j \text{ is even,} \\ \\ b_{j+1}^{-1} \cdots b_{p-2}^{-1} b_p^{-1} b_{p-1} \cdots b_{j+2} b_j & \text{when when } p - j \text{ is odd,} \end{cases}$$

which is a reduced word on $b_j, b_{j+1}^{\pm 1}, \ldots, b_p^{\pm 1}$ whose one and only b_j is its final letter.

So $\varphi^{-1}(u_0)$ has one i-letter, its last, and $\varphi^{-1}(u_1)$ has no b_i letters. Thus $\varphi^{-1}(u) = \varphi^{-1}(u_0)\varphi^{-1}(u_1)$ as words—there is no cancellation between the two factors. By the induction hypothesis, there is no cancellation between $\varphi^{-1}(u_1)$ and v, so the result follows. □

Lemma 13.9. *If u and $\varphi^{-1}(u)$ are both positive words on $b_0, \ldots, b_p$, then $|\varphi^{-1}(u)| \leq |u|$.*

Proof. For $0 \leq j \leq p$, let n_j and m_j be the number of b_j-letters in u and $\varphi^{-1}(u)$, respectively. Then in view of the form of φ^{-1} given in the proof of Lemma 13.8, we have

$$\begin{array}{lllll} 0 \;\leq\; m_0 & = & n_0, & & \text{and so} \\ 0 \;\leq\; m_1 & = & n_1 - n_0 & \leq \; n_1, & \text{and so} \\ 0 \;\leq\; m_2 & = & n_2 - n_1 + n_0 & \leq \; n_2, & \text{and so on,} \end{array}$$

from which the result follows. □

Lemma 13.10. *Given λ as in Lemmas 13.6 and 13.7, there exists a word μ on $a_1^{-1}, b_1, \ldots, b_p$ (so containing no $a_1, b_1^{-1}, \ldots, b_p^{-1}$) such that $|\mu| \leq 2|w_i|$, and an integer $0 \leq l \leq |w_i|$ such that in Q,*

$$\mu b_0 a_1^l \;=\; \lambda b_0.$$

Proof. Suppose that $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_l)$, $\mathbf{u} = (u_0, \ldots, u_l)$, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_l)$, where each λ_j is a positive word on $b_1, \ldots, b_p$, each u_j is a prefix of λ_j, each $\epsilon_i = \pm 1$, and $u_0 = \lambda_0$. Say that $\sigma^{-1} b_0 \tau = \lambda b_0$ *in* Q *via* $(\boldsymbol{\lambda}, \mathbf{u}, \boldsymbol{\epsilon})$ when

$$\begin{aligned} \sigma &= u_0^{-1} a_1^{\epsilon_1} u_1^{-1} a_1^{\epsilon_2} \cdots u_{l-1}^{-1} a_1^{\epsilon_l} u_l^{-1} \\ \tau &= a_1^{\epsilon_1} a_1^{\epsilon_2} \cdots a_1^{\epsilon_l} \\ \lambda &= \lambda_l \end{aligned}$$

as words, and for all $0 \leq j \leq l$,

$$\lambda_j b_0 \;=\; (u_j a_1^{-\epsilon_j} u_{j-1} \cdots a_1^{-\epsilon_2} u_1 a_1^{-\epsilon_1} u_0)\; b_0\; (a_1^{\epsilon_1} a_1^{\epsilon_2} \cdots a_1^{\epsilon_j}) \tag{13.5}$$

in Q, as illustrated in Figure 13.3.

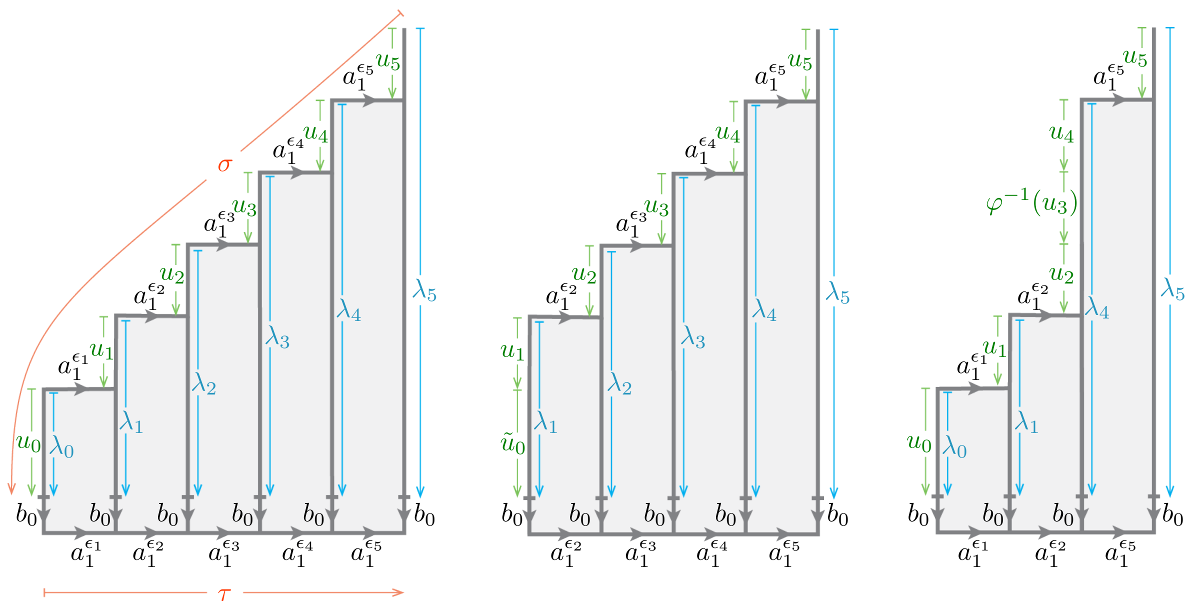


FIGURE 13.3. Illustrating our proof of Lemma 13.10 (with $l = 5$). Left: a diagram for $\sigma^{-1} b_0 \tau = \lambda b_0$ in Q via $(\boldsymbol{\lambda}, \mathbf{u}, \boldsymbol{\epsilon})$. Centre: the result of applying move I. Right: the result of applying move II (with $j = 4$).

Let $\lambda_0, \ldots, \lambda_{l-1}$ be the positive words on $b_1, \ldots, b_p$ such that $\lambda_0 b_0, \ldots, \lambda_{l-1} b_0$ are the words along the bottom boundaries (read in the direction from $\hat{\mu}$ to ν) of the a_1-corridors in Λ_1. Let $\lambda_l = \lambda$. Per Lemma 13.7, $\hat{\mu} b_0 \nu = \lambda b_0$ in G and, given how the a_1-corridors in Λ_1 pair off the $a_1^{\pm 1}$ in ν with the $a_1^{\pm 1}$ in $\hat{\mu}$, if we define σ and τ to be $\hat{\mu}^{-1}$ and ν with all letters a_2, t, x_1, x_2, y_1, and y_2 deleted, then they have the forms displayed above. Accordingly, they define $\mathbf{u}$ and $\boldsymbol{\epsilon}$ so that $\sigma^{-1} b_0 \tau = \lambda b_0$ in Q via $(\boldsymbol{\lambda}, \mathbf{u}, \boldsymbol{\epsilon})$. Moreover, $l \leq |w_i|$ and $|\mathbf{u}| := \sum_{j=0}^{l} |u_i| \leq 2|w_i|$, the last inequality coming from summing the bounds from Lemma 13.7 (1) and (3).

We will simplify $(\boldsymbol{\lambda}, \mathbf{u}, \boldsymbol{\epsilon})$ in two ways:

I. Suppose that $\epsilon_1 = -1$. Then (13.5) in the case $j = 1$ gives that in Q,

$$\lambda_1 b_0 \;=\; u_1 a_1 u_0\; b_0\; a_1^{-1} \;=\; u_1 \varphi^{-1}(u_0 b_0).$$

Now, u_1 is a prefix of λ_1 and so $\varphi^{-1}(u_0 b_0)$ is a suffix of $\lambda_1 b_0$, and so is a positive word. Therefore Lemma 13.9 applies and tells us that $|\varphi^{-1}(u_0 b_0)| \leq |u_0 b_0|$. Define $\tilde{u}_0$ to be the word obtained from $\varphi^{-1}(u_0 b_0)$ by removing its final letter b_0. Then $|\tilde{u}_0| \leq |u_0|$ and $\lambda_1 = u_1 \tilde{u}_0$. Define $\hat{\boldsymbol{\lambda}}$ to be $\boldsymbol{\lambda}$ with λ_0 discarded, define $\hat{\mathbf{u}}$ to be $\mathbf{u}$ with u_0 discarded and u_1 replaced by $u_1 \tilde{u}_0$, and define $\hat{\boldsymbol{\epsilon}}$ to be $\boldsymbol{\epsilon}$ with ϵ_1 discarded. Then $\sigma^{-1} b_0 \tau = \lambda b_0$ in Q via $(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\epsilon}})$, the lengths of the three sequences have all decreased by 1. And because $|\tilde{u}_0| \leq |u_0|$, we get $|\hat{\mathbf{u}}| \leq |\mathbf{u}|$.

II. Suppose $\epsilon_{j-1} = 1$ and $\epsilon_j = -1$ for some $2 \leq j \leq l$. Using (13.5) to relate $\lambda_{j-2} b_0$ and $\lambda_j b_0$, we get

$$\lambda_j b_0 \ = \ u_j a_1 u_{j-1} a_1^{-1} \ \lambda_{j-2} b_0 \ a_1 a_1^{-1} \ = \ u_j \varphi^{-1}(u_{j-1}) \ \lambda_{j-2} b_0$$

in Q. Now, u_j is a prefix of λ_j and $\lambda_j b_0$ is a positive word, so $\varphi^{-1}(u_{j-1}) \ \lambda_{j-2} b_0$ is equal in Q to a positive word, and then by Lemma 13.8, $\varphi^{-1}(u_{j-1})$ is a prefix of that positive word. Given that both $\varphi^{-1}(u_{j-1})$ and u_{j-1} are positive words, Lemma 13.9 tells us that $|\varphi^{-1}(u_{j-1})| \leq |u_{j-1}|$. Now define $\hat{\boldsymbol{\lambda}}$ to be $\boldsymbol{\lambda}$ with λ_{j-1} and λ_j discarded, define $\hat{\mathbf{u}}$ to be $\mathbf{u}$ with u_{j-2} and u_{j-1} discarded and u_j replaced with $u_j \varphi^{-1}(u_{j-1}) u_{j-2}$, and define $\hat{\boldsymbol{\epsilon}}$ to be $\boldsymbol{\epsilon}$ with ϵ_{j-1} and ϵ_j discarded. Then $\sigma^{-1} b_0 \tau = \lambda b_0$ in Q via $(\hat{\boldsymbol{\lambda}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\epsilon}})$, the lengths of the three sequences have all decreased by 2, and $|\hat{\mathbf{u}}| \leq |\mathbf{u}|$.

Repeat I and II until we have $(\boldsymbol{\lambda}, \mathbf{u}, \boldsymbol{\epsilon})$ via which $\sigma^{-1} b_0 \tau = \lambda b_0$ in Q with $\boldsymbol{\epsilon} = (1, \cdots, 1)$. Throughout, the bounds $l \leq |w_i|$ and $|\mathbf{u}| \leq 2|w_i|$ are maintained. The resulting $\mu = \sigma^{-1}$ and $\tau = a_1^l$ have the required properties. $\square$

A calculation in Q now bounds the length of λ. We state the result in the following lemma, deferring the proof to Section 14.

Lemma 13.11. *There exists $C_0 > 1$ with the following property. Suppose there are words μ on $a_1^{-1}, b_1, \ldots, b_p$ (so containing no $a_1, b_1^{-1}, \ldots, b_p^{-1}$) and λ on $b_1, \ldots, b_p$ (so containing only positive letters), and a number $l \geq 1$ such that in Q*

$$\mu b_0 a_1^l \ = \ \lambda b_0. \tag{13.6}$$

Then, if $|\cdot|_q$ counts the number of b_q in a given word, we have:

$$|\lambda| \ \leq \ C_0(|\mu| + |\lambda|_q)^{p/q}.$$

In the situation of Corollary 13.4, this leads to an upper bound on the lengths of the a_1-corridors in $\mathcal{B}_i$ for all i such that β_i is a b_0-corridor.

Lemma 13.12. *There exists $C_1 > 1$ such that if $\mathcal{C}$ is as in Lemma 13.6 and ξb_0 and λb_0 are the words read along the top and bottom boundaries (respectively) of $\mathcal{C}$, then*

$$\max\{|\lambda|, |\xi|\} \ \leq \ C_1 |w|^{p/q}.$$

Proof. First consider the word λb_0 along the bottom boundary of $\mathcal{C}$. Use Lemma 13.7 and 13.10 to obtain a word $\mu = \mu(b_1, \ldots, b_p, a_1^{-1})$ and a number $l \geq 1$ such that Lemma 13.11 applies. Then $|\lambda| \leq C_0(|\mu| + |\lambda|_q)^{p/q}$. By Lemma 13.10, we have $|\mu| \leq 2|w_i| \leq 2|w|$.

We estimate $|\lambda|_q$ using (a_2, b_q)-tracks (see Definition 12.1). The dual of every edge labelled b_q in λ is part of an (a_2, b_q)-track of Δ with endpoints on w (by parts (2) and (5) of Lemma 12.2). Suppose some (a_2, b_q)-track γ crosses $\mathcal{C}$ twice. Then the edges of λ dual to γ are necessarily labelled by $b_q^{\pm 1}$, as λ has no a_2, and since γ is oriented (Lemma 12.2(1)) at least one of these must be b_q^{-1}. This contradicts the fact, established in Lemma 13.6, that λ is a positive word. Thus any (a_2, b_q)-track crosses λ at most once. It follows that $|\lambda|_q \le |w|$. Thus

$$|\lambda| \ \le\ C_0(|\mu| + |\lambda|_q)^{p/q} \ \le\ C_0(4|w|)^{p/q} \ \le\ C_0'|w|^{p/q}, \tag{13.7}$$

for a suitable constant C_0'.

Now if ξb_0 is the top boundary of an a_1-corridor, then we have a relation $\xi b_0 = a_1^{-1}(\lambda b_0)a_1$, where λ is a positive word on $b_1, \ldots, b_p$. Inspecting the $r_{1,*}$-defining relations (of Figure 4.1), we see that $|\xi| \le C_0''|\lambda|$ for a suitable constant $C_0'' \ge 1$. Combining this with (13.7), we obtain $\max\{|\lambda|, |\xi|\} \le C_1|w|^{p/q}$ for a suitable constant $C_1 > 1$. □

Our next lemma is illustrated by Figure 13.4. We can now derive:

Lemma 13.13. *There exists a constant $C_2 > 1$ such that for all i such that β_i is a b_0-track,*

$$|W_i| \ \le\ C_2^{|w|^{p/q}}. \tag{13.8}$$

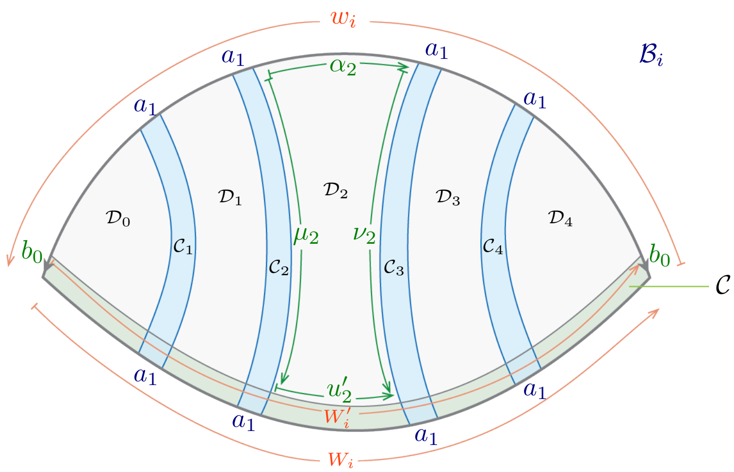


FIGURE 13.4. Illustrating our proof of Lemma 13.13 (with $l = 4$)

Proof. Let $\mathcal{C}$ be the (unique) b_0-corridor in $\mathcal{B}_i$ and let W_i' be its bottom boundary, so we have the relation $b_0^{-1}W_i'b_0 = W_i$. Then there exists a constant $K_0 \ge 1$ such that

$$|W_i| \ \le\ K_0|W_i'|. \tag{13.9}$$

Let $\mathcal{C}_1, \ldots \mathcal{C}_l$ be the a_1-corridors of $\mathcal{B}_i$ and let $\mathcal{D}_0, \ldots, \mathcal{D}_l$ be the (closures of the) components of $\mathcal{B}_i \setminus (\mathcal{C} \cup C_1 \cup \cdots \cup \mathcal{C}_l)$. Then, for all j, $\mathcal{D}_j$ is a van Kampen diagram for the relation $\mu_j^{-1}\alpha_j\nu_j = u_j'$, where α_j is a subpath of w_i, the paths μ_j and ν_j (which are possibly empty) run along the a_1-corridors bounding $\mathcal{D}_j$, and u_j' is a subpath of W_i'. As $\mathcal{D}_j$ has no a_1- or b_0-corridors, this relation holds in (in the notation of Figure 4.1)

$$\langle\ a_2, t, x_1, x_2,\ b_1, \ldots, b_p\ |\ \ \{r_{2,i},\ r_{3,j}\}_{1 \le i,j \le p},\ r_{4,2,1},\ r_{4,2,2},\ r_{4,2}\ \rangle,$$

which is a multiple HNN-extension of $F(a_2, t, x_1, x_2)$ with stable letters $b_1, \ldots, b_p$. It follows that $|u_j'| \le |\alpha_j|K_1^M$, where $K_1 \ge 1$ is a constant, and $M = \max(|\mu_j|, |\nu_j|)$, which is at most $C_1|w|^{p/q}$ by

Lemma 13.12. Then, because the number of a_1-corridors is l, we have

$$|W_i'| \ \le \ l + \sum_{i=0}^{l+1} |u_j'| \ \le \ l + \sum_{i=0}^{l+1} |\alpha_j| K_1^M \ \le \ \left(l + \sum_{i=0}^{l+1} |\alpha_j| \right) K_1^M \ \le \ |w| K_1^{C_1 |w|^{p/q}}.$$

This and (13.9) together establish (13.8) for a suitable constant $C_2 > 1$. $\square$

We can now complete:

Proof of Proposition 13.1. Recall that Δ is a van Kampen diagram for $w\chi^{-1}$ and $\mathcal{A}$ is a subdiagram for $W\chi^{-1}$, where W is as defined in (13.3) and all the 2-cells of $\mathcal{A}$ are $r_{4,*}$- or $r_{4,*,*}$-cells (per Figure 4.1). Now, $\mathcal{A}$ is a tree-like arrangement of 2-disc components connected by 1-dimensional portions (trees). As $r_{4,*}$- and $r_{4,*,*}$-cell have no x-edges on their boundaries, any x-edges in $\mathcal{A}$ are in 1-dimensional portions. Let $\widehat{\mathcal{A}}$ be the subdiagram of $\mathcal{A}$ consisting of the path χ and all its 2-disc components that share at least one edge with χ. Then $\widehat{\mathcal{A}}$ is a van Kampen diagram for $\widehat{W}\chi^{-1}$, where $\widehat{W}$ is a word obtained from W by deleting some of its letters. Then $\widehat{W}$ contains no x-letters: its letters are either along the path χ or are on the boundaries of 2-cells, neither of which have x-edges.

If β_i is not a b_0-track, then W_i is a word on $a_1X_*tX_*$, $a_1X_*tX_*$, $a_1a_2X_*tX_*$, $X_*t^{-1}X_*tX_*$ and X_*tX_*. And (because Δ is reduced and thanks to the $C'(1/4)$ small-cancellation condition of Section 4 for the set $\mathcal{X}$ of the X_*), if a subword of the freely reduced form of W_i contains no x-letters, then it has length at most 2. It follows that W_i can contribute at most two letters to $\widehat{W}$.

Therefore, in the notation of (13.3), $|\widehat{W}|$ is at most $\sum_{i=0}^{r} |v_i|$, plus twice the number of W_i such that β_i is not b_0-track, plus the lengths of the remaining W_i. So, using Lemma 13.13 and that there are at most $|w|$ subwords in W_i in W, for a suitable constant $C_3 > 1$, we get

$$|\widehat{W}| \ \le \ |w| + 2|w| + |w| C_2^{|w|^{p/q}} \ \le \ C_3^{|w|^{p/q}}. \tag{13.10}$$

Next we claim that there exists a constant $C_4 > 1$ such that

$$|\chi| \ \le \ |\widehat{W}|\, C_4^{|w|}. \tag{13.11}$$

Since the 2-cells in $\widehat{\mathcal{A}}$ are all of type $r_{4,*}$ or $r_{4,*,*}$ (per Figure 4.1), $\widehat{\mathcal{A}}$ is a union of non-intersecting a_1- and a_2-corridors. Each a_1-corridor of $\widehat{\mathcal{A}}$ is part of an a_1-corridor of Δ whose ends are in w, and Lemma 11.2(7) implies that no two a_1-corridors of $\widehat{\mathcal{A}}$ are part of the same a_1-corridor in Δ. On the other hand, several a_2-corridors of $\widehat{\mathcal{A}}$ could be part of the same (a_2, b_q)-corridor of Δ. However, by Lemma 12.3, if a pair of a_2-corridors of $\mathcal{A}$ nest (meaning one is entirely in the W-side of the other), then they cannot be part of the same (a_2, b_q)-corridor of Δ. It follows that the same is true of $\widehat{\mathcal{A}}$: no pair of a_2-corridors of $\widehat{\mathcal{A}}$ have the property that one is entirely in the $\widehat{W}$-side of the other. Distinct (a_2, b_q)-corridors end on distinct pairs of edges of w.

Thanks to these observations, we can strip away successive portions of $\widehat{\mathcal{A}}$ by at most $|w|$ moves, each of which either

- removes an a_1-corridor, or
- removes *all* the a_2-corridors of $\widehat{\mathcal{A}}$ that are part of the same (a_2, b_q)-corridor of Δ.

The result is a sequence of diagrams which demonstrate that each word in a sequence of words equals χ in G. Moreover, this sequence of words starts with $\widehat{W}$ and ends with a word freely equal to χ, and the length of each word is longer than the last by at most a constant factor. This proves (13.11) for a suitable constant $C_4 > 1$.

Finally, (13.10) and (13.11) combine to yield

$$|\chi| \;\le\; |\widehat{W}|\, C_4^{|w|} \;\le\; C_3^{|w|^{p/q}} C_4^{|w|} \;\le\; K^{|w|^{p/q}}$$

for a suitably chosen constant $K > 1$. □

14. Why p/q?

This section is devoted to a proof of Lemma 13.11, which we used in our proof of Proposition 13.1. The lemma concerns the group

$$Q \;=\; \langle a_1, b_0, \ldots, b_p \mid a_1^{-1} b_i a_1 = \varphi(b_i) \;\; \forall i \;\rangle \;\;\text{with}\;\; \varphi(b_j) \;=\; \begin{cases} b_{j+1} b_j & \text{for } j < p \\ b_j & \text{for } j = p. \end{cases}$$

We begin with two preparatory lemmas. We use the convention that the binomial coefficient $\binom{n}{r}$ equals 0 for all $r \notin \{0, \ldots, n\}$.

Lemma 14.1. *Consider the relation $a_1^{-m} b_i a_1^m = \lambda$ in Q, where $m \ge 0$, $0 \le i \le p$, and λ is a word in $b_0, \ldots, b_p$. Then*

(1) For $0 \le j \le p - i$, there are $\binom{m}{j}$ instances of b_{i+j} in λ. Also, λ has no b_k for $k < i$.
(2) If $m > 2p$, then $|\lambda| \le (p+1)\binom{m}{p-i}$.
(3) If $m \le 2p$, then $|\lambda| \le (p+1)(2p)^p$

Proof. For (1), induct on m or refer to [BR09]. For (2), note that if $0 \le i \le p$ and $m > 2p$, then $p - i \le p < m/2$, and so $\binom{m}{j} \le \binom{m}{p-i}$ for all $j \le p - i$. Then from (1), we have

$$|\lambda| \;=\; \sum_{j=0}^{p-i} \binom{m}{j} \;\le\; \sum_{j=0}^{p-i} \binom{m}{p-i} \;\le\; (p-i+1)\binom{m}{p-i} \;\le\; (p+1)\binom{m}{p-i}.$$

For (3), we use the fact that $\binom{m}{j} \le m^j$ for any $j \le m$, and

$$|\lambda| \;=\; \sum_{j=0}^{p-i} \binom{m}{j} \;\le\; \sum_{j=0}^{p-i} m^j \;\le\; (p-i+1)m^{p-i} \;\le\; (p+1)(2p)^p.$$

□

Lemma 14.2. *Let $K = (2p)^{p^2}$. For all $m, k, l \in \mathbb{Z}$ such that $m > 2p$ and $1 \le k, l \le p$,*

(1) $\binom{m}{k} \le K \binom{m}{l}^{k/l}$
(2) If $l < k$, then $\binom{m}{k} \le K \binom{m}{l}\binom{m}{k-l}$

Proof. Let $m > 2p$. Now, if t satisfies $1 \le t \le p$, then $m > 2t$, or equivalently $-t > -m/2$. Consequently, $m - t + 1 > m - m/2 + 1 > m/2$, which gives the "$>$" in:

$$m^t \ge \binom{m}{t} \;=\; \frac{m(m-1)\ldots(m-t+1)}{t!} \;>\; \left(\frac{m}{2}\right)^t \frac{1}{t!} \;\ge\; \frac{m^t}{2^p p!} \;\ge\; \frac{m^t}{(2p)^p}. \tag{14.1}$$

Now, $\binom{m}{k} \le m^k$, (14.1), and $k < p$, respectively, imply the first, second, and third of the following inequalities:

$$\binom{m}{k}^l \;\le\; m^{kl} \;\le\; (2p)^{pk} \binom{m}{l}^k \;\le\; (2p)^{p^2} \binom{m}{l}^k.$$

Then (1) follows since $(2p)^{p^2/l} \le (2p)^{p^2} = K$.

For (2), now apply (14.1) to $t = l$ and $t = k - l$, and note that $2p \le p^2$ (since $1 \le l < k \le p$ implies that $p \ge 2$):

$$\binom{m}{k} \;\le\; m^k = m^l m^{k-l} \;\le\; (2p)^p \binom{m}{l} (2p)^p \binom{m}{k-l} \;=\; (2p)^{2p} \binom{m}{l}\binom{m}{k-l} \;\le\; K\binom{m}{l}\binom{m}{k-l}.$$

□

For a word π, we write $|\pi|_b$ and $|\pi|_q$ to denote the number of b-letters and the number of b_q-letters (respectively) in π.

Suppose μ is a word on $a_1^{-1}, b_1, \ldots, b_p$ (no $a_1, b_1^{-1}, \ldots, b_p^{-1}$ letters), λ is a positive word on $b_1, \ldots, b_p$, and $l \ge 1$ is an integer such that in Q

$$\mu b_0 a_1^l \;=\; \lambda b_0. \tag{14.2}$$

Lemma 13.11 asserts that

$$|\lambda| \;\le\; C_0(|\mu| + |\lambda|_q)^{p/q} \tag{14.3}$$

for a suitable constant $C_0 > 1$.

Here is the idea behind this. When we shuffle the $a_1^{\pm 1}$ letters through $\mu b_0 a_1^l$, in order to collect them together and cancel them away and obtain λb_0, the effect is to apply φ to the intervening b-letters. Lemma 14.1(1) indicates how the number of b-letters then grows: as a function of l, the number of b_i-letters in λ is at most a polynomial of degree i. Whether this rate of growth is achieved depends on μ. What (14.3) states is how the total number of b-letters produced is contingent on the length of μ *and* the number of b_q-letters produced.

Proof of Lemma 13.11. Let $C_0 = (p+1)(2p)^{2p^2}$. We induct on $|\mu|_b$.

Base case. In the base case, $|\mu|_b = 0$, and so $\mu = a_1^{-l}$ and (14.2) is $a_1^{-l} b_0 a_1^l = \lambda b_0$. Then $|\lambda|_q = \binom{l}{q}$ by Lemma 14.1(1), and so

$$|\mu| + |\lambda|_q \ge \binom{l}{q}. \tag{14.4}$$

If $l > 2p$, then Lemmas 14.1(2) and 14.2(1) apply so as to give the first and second (respectively) of the following inequalities; the definition of C_0 and (14.4) give the third:

$$|\lambda| \;\le\; (p+1)\binom{l}{p} \;\le\; (p+1)(2p)^{p^2}\binom{l}{q}^{p/q} \;\le\; C_0(|\mu| + |\lambda|_q)^{p/q}.$$

If, on the other hand, $l \le 2p$, then, by Lemma 14.1(3), we have that

$$|\lambda| \;\le\; (p+1)(2p)^p \;\le\; C_0 \;\le\; C_0(|\mu| + |\lambda|_q)^{p/q},$$

with the final inequality true because $l \ge 1$. This completes our proof of the base case.

Inductive step. Suppose we have $\hat{\mu} b_0 a_1^l = \hat{\lambda} b_0$ as per (14.2) with $|\hat{\mu}|_b = k + 1$. We will show that $|\hat{\lambda}|^q \le C_0^q \hat{n}^p$, where

$$\hat{n} \;=\; |\hat{\mu}| + |\hat{\lambda}|_q. \tag{14.5}$$

Suppose b_i is the first b-letter in $\hat{\mu}$. Then $\hat{\mu} = a_1^{-m} b_i \beta$ for some integer m such that $0 \le m \le l$, and word β that contains $l - m$ instances of a_1^{-1} and satisfies $|\beta|_b = k$. The exponent sums of the

a_1-letters in $a_1^{-m}b_ia_1^m$ and $a_1^{-m}\beta a_1^l$ are both 0, so there exist positive words γ and λ, respectively, on $b_1,\ldots,b_p$ representing them in Q. Then in Q,

$$\hat{\lambda}b_0 \;=\; \hat{\mu}b_0a_1^l \;=\; (a_1^{-m}b_ia_1^m)(a_1^{-m}\beta b_0a_1^l) \;=\; \gamma\lambda b_0.$$

Thus $|\hat{\lambda}| = |\lambda| + |\gamma|$. We will bound $|\hat{\lambda}|^q$ by combining bounds on $|\lambda|$ and $|\gamma|$.

Setting $\mu = a_1^{-m}\beta$, we have $\mu b_0 a_1^l = \lambda b_0$ in Q, where μ satisfies the hypotheses of the present lemma and $|\mu|_b = k$. By the induction hypothesis, $|\lambda| \le C_0 n^{p/q}$, where $n = |\mu| + |\lambda|_q$.

Before bounding $|\gamma|$ we make some observations about n and $\hat{n}$. Firstly, the presence of b_0 in the relation $a_1^{-m}\beta b_0 a_1^l = \lambda$, together with Lemma 14.1(1) implies that $|\lambda|_q \ge \binom{m}{q}$, and so

$$n \;\ge\; \binom{m}{q}. \tag{14.6}$$

Note that $|\hat{\mu}| = |\beta| + 1 + m = |\mu| + 1$, leading to:

$$\hat{n} \;=\; |\hat{\mu}| + |\hat{\lambda}|_q \;=\; |\mu| + 1 + |\lambda|_q + |\gamma|_q \;=\; n + 1 + |\gamma|_q. \tag{14.7}$$

Then, since $|\hat{\lambda}| = |\lambda| + |\gamma|$, we have

$$\begin{aligned}
|\hat{\lambda}|^q \;\le\; (|\lambda| + |\gamma|)^q \;&=\; \sum_{j=0}^{q}\binom{q}{j}|\lambda|^{q-j}|\gamma|^j \\
&\le\; \sum_{j=0}^{q}\binom{q}{j}\left(C_0 n^{p/q}\right)^{q-j}|\gamma|^j \qquad \text{(by the induction hypothesis)} \\
&\le\; \sum_{j=0}^{q}\binom{q}{j}C_0^{q-j}n^{p-\frac{pj}{q}}|\gamma|^j. \qquad (14.8)
\end{aligned}$$

Similarly to the base case, we treat the cases $m \le 2p$ and $m > 2p$ separately. When $m > 2p$, our estimate depends on whether $i \ge q$, in which case no new b_q letters are created in γ, or $i < q$, in which case new b_q letters are created in γ. Thus, we have three cases as follows.

Case 1: $m \le 2p$. In this case, $|\gamma| \le C_0$ by Lemma 14.1(3). Moreover, since $p > q$, we have $n^{p-\frac{pj}{q}} \le n^{p-j}$ and $\binom{q}{j} \le \binom{p}{j}$ for each j. Continuing from (14.8), we get

$$|\hat{\lambda}|^q \;\le\; \sum_{j=0}^{q}\binom{q}{j}C_0^{q-j}n^{p-j}C_0^j \;\le\; C_0^q\sum_{j=0}^{q}\binom{p}{j}n^{p-j} \;\le\; C_0^q(n+1)^p$$

Finally, since $\hat{n} \ge n+1$ by (14.7), we obtain $|\hat{\lambda}|^q \le C_0^q\hat{n}^p$, as desired.

Case 2: $m > 2p$ and $q \le i \le p$. We have that for $K = (2p)^{p^2}$:

$$\begin{aligned}
|\gamma| \;&\le\; (p+1)\binom{m}{p-i} && \text{by Lemma 14.1(2)} \\
&\le\; (p+1)\binom{m}{p-q} && \text{as } p-i \le p-q \le p \text{ and } m > 2p \\
&\le\; (p+1)K\binom{m}{q}^{\frac{p-q}{q}} && \text{by Lemma 14.2(1), as } m > 2p \text{ and } q, p-q \le p \\
&\le\; C_0 n^{\frac{p}{q}-1} && \text{by (14.6).}
\end{aligned}$$

Then, continuing from (14.8), and using that $\hat{n} \geq n+1$ by (14.7) and that $\binom{q}{j} \leq \binom{p}{j}$ for each j, we get

$$|\hat{\lambda}|^q \;\leq\; \sum_{j=0}^{q} \binom{q}{j} C_0^{q-j} n^{p-\frac{pj}{q}} (C_0 n^{\frac{p}{q}-1})^j \;\leq\; C_0^q \sum_{j=0}^{q} \binom{p}{j} n^{p-j} \;\leq\; C_0^q (n+1)^p \;\leq\; C_0^q \hat{n}^p.$$

Case 3: $m > 2p$ and $1 \leq i < q$. In this case, $|\gamma|_q = \binom{m}{q-i}$ by Lemma 14.1(1) and

$$\begin{aligned}
|\gamma| \;&\leq\; (p+1)\binom{m}{p-i} && \text{by Lemma 14.1(2)} \\
&\leq\; (p+1)K\binom{m}{p-q}\binom{m}{q-i} && \text{by Lemma 14.2(2), where } K=(2p)^{p^2} \\
&\leq\; (p+1)K^2\binom{m}{q}^{\frac{p-q}{q}}\binom{m}{q-i} && \text{by Lemma 14.2(1), as } m>2p \text{ and } 1 \leq q, p-q \leq p \\
&\leq\; C_0 n^{\frac{p-q}{q}} |\gamma|_q && \text{by (14.6)}, K=(2p)^{p^2}, \text{ and } |\gamma|_q = \binom{m}{q-i}.
\end{aligned}$$

Then, continuing from (14.8), we have

$$\begin{aligned}
|\hat{\lambda}|^q \;&\leq\; \sum_{j=0}^{q} \binom{q}{j} C_0^{q-j} n^{p-\frac{pj}{q}} \left(C_0 n^{\frac{p-q}{q}} |\gamma|_q\right)^j \\
&\leq\; C_0^q \sum_{j=0}^{q} \binom{p}{j} n^{p-j} |\gamma|_q^j \\
&\leq\; C_0^q \left(n + |\gamma|_q\right)^p \\
&\leq\; C_0^q \hat{n}^p,
\end{aligned}$$

where the last inequality follows from (14.7).

This concludes the proof of inductive step, as $|\hat{\lambda}| \leq C_0 \hat{n}^{p/q}$ in all three cases. $\square$

15. Iterated exponential functions

Recall that $\exp^k$ denotes the k-fold iterated exponential-function—that is, $\exp^1(x) = \exp(x)$ and $\exp^i(x) = \exp(\exp^{i-1}(x))$ for integers $i > 1$. Here we will leverage our examples $H \leq G$ from Section 4 to construct free subgroups of hyperbolic groups whose distortion functions are $\simeq$-equivalent to $n \mapsto \exp^k(n^{p/q})$, where $p > q \geq 1$ and $k > 1$ are integers, proving Theorem 1.1. We will take iterated amalgamated products of G with certain hyperbolic free-by-free groups constructed by Brady and Tran [BT21]. We begin by reviewing the parts of their construction we need. We write F_m to denote the free group on m generators.

Theorem 15.1. [BT21, Theorem 5.2] *Given $m \geq 1$, there exists $l > m$ and a group $F_l \rtimes F_m$ that is* CAT(0) *and hyperbolic.*

Definition 15.2. Let G_1 be a finitely generated group and let $F_{m_1} < G_1$ be a free subgroup of rank m_1. Take $m_1 < m_2 < \cdots$ so that $F_{m_{i+1}} \rtimes F_{m_i}$ is the group of Theorem 15.1 (with $m_{i+1} = l$ and $m_i = m$). For $i > 1$, define G_i by

$$G_i \;=\; (F_{m_i} \rtimes F_{m_{i-1}}) *_{F_{m_{i-1}}} G_{i-1}.$$

Proposition 15.3. [BT21, Proposition 4.4] *In the notation of Definition 15.2, if* $\mathrm{Dist}_{F_{m_1}}^{G_1} \simeq f$ *for some non-decreasing superadditive function* f*, then for all integers* $k \geq 1$*,*

$$\mathrm{Dist}_{F_{m_k}}^{G_k}(n) \;\simeq\; \exp^{k-1}(f(n)).$$

To complete the proof of Theorem 1.1, we will take G_1 and F_{m_1} to be our groups G and $H \cong F_3$, respectively, from Section 4. We will then use the following two results to conclude that G_k is hyperbolic when $k > 1$.

Theorem 15.4. ***(Hyperbolicity of amalgams)*** *If a finitely generated group* C *is a subgroup of two hyperbolic groups* A *and* B*, and* C *is quasi-convex and malnormal in* A*, then*

$$\Gamma \;=\; A *_C B$$

is hyperbolic. (We make no assumption on how C *sits in* B*.)*

Proof. Since C is finitely generated and is quasi-convex and malnormal in the hyperbolic group A, [Bow12, Theorem 7.11] tells us that A is hyperbolic relative to C. We then get that Γ is hyperbolic relative to B by [Dah03, Theorem 0.1(2)]. A group that is hyperbolic relative to a hyperbolic subgroup is itself hyperbolic by [Osi06, Corollary 2.41]. So Γ is hyperbolic. $\square$

Lemma 15.5. *If* A *and* B *are finitely generated free groups and* $G = A \rtimes B$ *is a hyperbolic group, then* B *is quasiconvex and malnormal in* G*.*

Proof. For quasiconvexity, observe that B is a retract of G, so it is in fact convex in G (with respect to standard generating sets).

To see that B is malnormal, recall that the group G can be identified with the Cartesian product $A \times B$ endowed with the multiplication $(a, b)(c, d) = (a\varphi_b(c), bd)$, where $\varphi_b(x) = bxb^{-1}$ for all $x \in A$. Note that $(c, d)^{-1} = (\varphi_{d^{-1}}(c^{-1}), d^{-1})$ for all $(c, d) \in G$. We identify B with $\{1\} \times B$.

Now if B is not malnormal, then there exists some $(c, d) \in G \setminus B$ such that $(c, d)^{-1}B(c, d) \cap B$ is non-trivial. Thus, there exists $b \in B$ with $b \neq 1$, such that

$$(c, d)^{-1}(1, b)(c, d) \;=\; (\varphi_{d^{-1}}(c^{-1}), d^{-1})(\varphi_b(c), bd) \;=\; (\varphi_{d^{-1}}(c^{-1})\varphi_{d^{-1}}(\varphi_b(c)), d^{-1}bd) \in B.$$

In particular, we must have $1 = \varphi_{d^{-1}}(c^{-1})\varphi_{d^{-1}}(\varphi_b(c)) = \varphi_{d^{-1}}(c^{-1}\varphi_b(c))$, and since $\varphi_{d^{-1}}$ is an automorphism, we have $c^{-1}\varphi_b(c) = 1$, or equivalently $c^{-1}bcb^{-1} = 1$. Observe that $c \neq 1$ as $(c, d) \in G \setminus B$. So b and c are commuting elements of infinite order (since A and B are free and inject into G) in a hyperbolic group, a contradiction. We conclude that B is malnormal. $\square$

Proof of Theorem 1.1. Given $p > q \geq 1$, let G and H be the groups we constructed in Section 4 and proved in Sections 8–14 to have $\mathrm{Dist}_H^G(n) \simeq \exp(n^{p/q})$. Define $G_1 = G$ and $m_1 = 3$, so that $F_{m_1} = F_3 \cong H$. Define the groups G_k for $k > 1$ as in Definition 15.2. Then, since G_1 is hyperbolic, and $F_{m_{i+1}} \rtimes F_{m_i}$ is hyperbolic for each i, we inductively conclude that each G_i is hyperbolic, using Theorem 15.4 and Lemma 15.5. Finally, since $\exp(n^{p/q})$ is a non-decreasing superadditive function, Proposition 15.3 implies

$$\mathrm{Dist}_{F_{m_k}}^{G_k}(n) \;\simeq\; \exp^k(n^{p/q}),$$

as desired. $\square$

Remark 15.6. ($\mathrm{CAT}(0)$ **and** $\mathrm{CAT}(-1)$ **structures for the groups** G_k) For all $p > q \geq 1$, our group G of Section 4 satisfies a uniform $C'(1/6)$ condition, so can be given a CAT(0) structure by [Wis04a] or even a CAT(−1) structure by [Bro, Gro01, Mar17]. The $F_l \rtimes F_m$ groups constructed by Brady–Tran have a piecewise Euclidean CAT(0) structure and furthermore, F_m is *ultra-convex* in $F_l \rtimes F_m$—a property they use to show that if the Gromov link condition holds in the complex associated to a group Γ, then it continues to hold for an amalgamated product of the form $(F_l \rtimes F_m) *_{F_m} \Gamma$. See [BT21, Lemma 5.10] for the precise statement. Moreover, the strategy used in [Bro, Gro01] to obtain CAT(−1) structures by changing each Euclidean polygon to a hyperbolic one by slightly shrinking each angle can be applied to the Brady–Tran groups to obtain CAT(−1) groups for the form $F_l \rtimes F_m$. Thus, we expect that by choosing CAT(0) or CAT(−1) structures on the building blocks and using the ultra-convexity as in [BT21], the groups G_k in Definition 15.2 can be shown to be CAT(0) or CAT(−1) for all k.

16. Distortion of hyperbolic subgroups of hyperbolic groups

Here we use ideas originating in I. Kapovich's [Kap99] to prove Theorem 1.2, which, in particular, extends our main result (Theorem 1.1) in that it allows the distorted subgroup H to be any non-elementary torsion-free hyperbolic group rather than F_3.

For each of the functions f listed in Theorem 1.2, there are constructions in the literature consisting of a hyperbolic group K and a finite-rank free group $F \leq K$ such that $\mathrm{Dist}_F^K \simeq f$: see [Mit98a, BBD07] for (1) when $p = q$, this article for (1) when $p > q$, and [BDR13] for (2). We will prove the theorem by amalgamating H with K along a subgroup of H that is isomorphic to F and is supplied by the following lemma.

Lemma 16.1. *Suppose H is a non-elementary torsion-free hyperbolic group. For all $k \geq 2$, H contains a malnormal quasiconvex free subgroup F of rank k.*

Proof. Kapovich showed that such an H has a malnormal quasiconvex rank-2 free subgroup $F(x, y)$ [Kap99, Theorem C]. There are malnormal rank-3 free subgroups in $F(x, y)$—for example

$$\langle x^{10}, y^{10}, (xy)^{10} \rangle$$

is malnormal by the criterion of [KM02, Theorem 10.9], which can be interpreted as being that there is no reduced word which read from two different vertices in the Stallings graph of the subgroup makes a loop. Likewise, for all $k \geq 2$, for sufficiently large n, the subgroup

$$\Big\langle\, x^n,\ y^n,\ (xy)^n,\ (x^2y^2)^n,\ \ldots,\ (x^{k-2}y^{k-2})^n \,\Big\rangle$$

of $F(x, y)$ is malnormal and rank-k. The result then follows from the following three facts. If $A \leq B \leq C$ are groups such that A is malnormal in B and B is malnormal in C, then A is malnormal in C. Quasiconvexity is similarly transitive. Finitely generated subgroups of F_2 are quasiconvex. □

Now, given H and f as in Theorem 1.2, let $F \leq K$ be as above so that K is hyperbolic, F is finite-rank free, and $\mathrm{Dist}_F^K \simeq f$. By Lemma 16.1, H has a quasiconvex malnormal subgroup which is isomorphic to F. We will also refer this subgroup of H as F, so that we can define

$$G \;=\; H *_F K. \tag{16.1}$$

The last ingredient we require for Theorem 1.2 is:

Theorem 16.2. *Let $\Gamma = A *_C B$, where A, B, and C are finitely generated groups and let f be a superadditive function such that $n \leq f(n)$ for all n.*

(1) If $\mathrm{Dist}_C^A \preceq f$ and $\mathrm{Dist}_C^B \preceq f$, then $\mathrm{Dist}_A^\Gamma \preceq f$ and $\mathrm{Dist}_B^\Gamma \preceq f$.
(2) If $\mathrm{Dist}_C^A(n) \simeq n$ and $\mathrm{Dist}_C^B \simeq f$, then $\mathrm{Dist}_A^\Gamma \simeq f$.

Proof of Theorem 1.2 assuming Theorem 16.2. Given H and f as in the theorem, let G be the group defined in (16.1). Since F is malnormal and quasiconvex in H, Theorem 15.4 tells us that G is hyperbolic. Now $\mathrm{Dist}_F^K \simeq f$ by construction, and note that every function f listed in the statement of Theorem 1.2 is superadditive and superlinear. Since F is quasiconvex in H and H is hyperbolic, we have $\mathrm{Dist}_F^H(n) \simeq n \preceq f(n)$, and Theorem 16.2(2) implies that $\mathrm{Dist}_H^G \simeq f$. $\square$

Proof of Theorem 16.2. We begin with some setup. For $X = A, B, C$, let S_X be a generating set for X, and let K_X be a $K(X, 1)$ with 1-skeleton a rose on $|S_X|$ petals. We assume that $S_C \subset S_A$ and $S_C \subset S_B$. Then Γ is generated by $S_\Gamma = S_A \cup S_B$. Let K be the standard graph of spaces with fundamental group Γ, i.e.,

$$K \;=\; (K_A \sqcup (K_C \times [0,1]) \sqcup K_B)/\sim$$

where $\sim$ identifies $K_C \times \{0\}$ and $K_C \times \{1\}$ with the images of the maps induced by the inclusion of C in A and B respectively. For convenience, we subdivide the cell structure so that $K_C^{(1)} \times \{1/2\} \subset K^{(1)}$.

Let c be the unique vertex of K_C, and let $p = \{c\} \times \{1/2\} \in K_C \times [0,1] \subset K$. We identify Γ with $\pi_1(K, p)$. More precisely, identify S_C with the set of petals of $K_C^{(1)} \times \{1/2\}$ and $S_A \setminus S_C$ with the collection of loops $\delta\alpha\bar{\delta}$, where δ and $\bar{\delta}$ are the interval $\{c\} \times [0, 1/2] \subset K_C \times [0,1]$ oriented towards and away from K_A respectively, and α is a petal of $K_A^{(1)}$ representing an element of $S_A \setminus S_C$. Identify $S_B \setminus S_C$ with the analogous set of loops, replacing $\{c\} \times [0, 1/2]$ with $\{c\} \times [1/2, 1]$. Let $\mathcal{S}_\Gamma$ be the set of the loops defined in this paragraph. Each element of $\mathcal{S}_\Gamma$ is contained in $K^{(1)}$.

The associated Bass–Serre tree is obtained by collapsing each lift of K_A or K_B in $\widetilde{K}$ to a vertex (called the A- and B- vertices, respectively) and each lift of $K_C \times [0,1]$ to an edge. We subdivide each edge by adding a midpoint, obtained by collapsing a lift of $K_C \times \{1/2\}$; we call each such midpoint a C-vertex. Let T denote this subdivided tree, and let $\psi : \widetilde{K} \to T$ denote the collapsing map. Given an A- or B- vertex v of T, define s_v to be the star of v in T. Since T is subdivided, every vertex of s_v besides v is a C-vertex.

If $\gamma \in \mathcal{S}_\Gamma$ corresponds to $g \in S_\Gamma$, then each lift $\tilde{\gamma}$ of γ in $\widetilde{K}^{(1)}$ is considered to be labelled by g. By construction, the image of $\psi \circ \tilde{\gamma}$ is a C-vertex if $g \in S_C$, and otherwise it is contained in a star s_v for an A- or B-vertex v. More generally, if w is any word over S_Γ, then for each lift $\tilde{p}$ of p, there is a path ξ_w starting at $\tilde{p}$ with label w in $\widetilde{K}^{(1)}$. (We abuse notation by suppressing $\tilde{p}$.)

Now if, in addition, $w = 1$ in Γ, then ξ_w is a loop based at some (any) $\tilde{p}$ and $\psi \circ \xi_w$ is a loop based at $\psi(\tilde{p})$ in T. The image of $\psi \circ \xi_w$ is a subtree of T, which we denote τ_w. We measure the complexity of w by $n(w)$, which counts the number of A- or B-vertex stars intersecting τ_w:

$$n(w) \;=\; \#\{v \mid v \text{ is an } A\text{- or } B\text{-vertex and } s_v \cap \tau_w \neq \emptyset\}.$$

Note that $n(w)$ is finite since τ_w is compact, and $n(w) \geq 1$ as $\psi(\tilde{p}) \in \tau_w$.

We are now ready to prove (1). In this proof, a *geodesic word in X* or *over S_X* will mean a word of minimal length over S_X representing an element of X, where $X = A, B$, or Γ. Let u be a geodesic word in either A or B and let w be a geodesic word in Γ with $u^{-1}w = 1$ in Γ. We wish to show that $|u| \leq f(|w|)$. The proof is by induction on $n(u^{-1}w)$.

If $n(u^{-1}w) = 1$, then $\tau_{u^{-1}w}$ is contained in some s_v, where v is an A- or B-vertex, depending on whether u is in A or B. We assume without loss of generality that v is an A-vertex. By construction, $s_v = \psi(Y)$, where $Y \subset \widetilde{K}$ consists of some lift of K_A, and all the lifts of $K_C \times [0, 1/2]$ intersecting it. Now $\xi_{u^{-1}w}$ is contained $Y^{(1)}$ and it follows that its label $u^{-1}w$ is a word over S_A. Thus u and w are both geodesics over S_A representing the same element of A, so $|u| = |w|$. This proves the base step of the induction.

For the induction step, assume that $|u'| \leq f(|w'|)$ whenever $u'^{-1}w' = 1$ with u' a geodesic in A or B and w' a geodesic in Γ and $n(u'^{-1}w') < n(u^{-1}w)$. Again, assume without loss of generality that u is a geodesic in A. Write $\xi_{u^{-1}w}$ as a concatenation $\xi_{u^{-1}}\xi_w$. Then $\psi(\xi_{u^{-1}}) \subset s_a$ for some A-vertex a (since u is a geodesic over S_A). Now, by considering $\psi^{-1}(\tau_{u^{-1}w} \setminus s_a^\circ)$, where s_a° denotes the interior of s_a, we obtain a concatenation $\xi_w = \xi_{x_0}\xi_{y_1}\xi_{x_1}\cdots\xi_{y_k}\xi_{x_k}$ (so $w = x_0y_1x_1\cdots y_kx_k$, as words), such that for each i, we have that $\psi(\xi_{x_i}) \subset s_a$ (so x_i is a word over S_A) and that $\psi \circ \xi_{y_i}$ is a loop in $\tau_{u^{-1}w} \setminus s_a^\circ$ based at a C-vertex p_i of s_a.

By construction, each ξ_{y_i} has its endpoints in some lift of $K_C \times \{1/2\}$, and so y_i represents an element of C, and therefore of B. Let z_i be a geodesic word over S_B with $z_i = y_i$ in Γ, and let ξ_{z_i} be the path in $\widetilde{K}$ with the same endpoints as ξ_{y_i}. Then $\psi(\xi_{z_i}) \subset s_{b_i}$ where b_i is the unique B-vertex adjacent to p_i. Now consider $\xi_{z_i^{-1}y_i} = \overline{\xi}_{z_i}\xi_{y_i}$ and note that $\psi(\xi_{y_i})$ intersects s_{b_i}, since the endpoints of ξ_{y_i} map to p_i. It follows that $\tau_{z_i^{-1}y_i}$ intersects the same number of A- and B-vertex stars as $\psi(\xi_{y_i})$, and, by construction, this number is less than $n(u^{-1}w)$ (since $\tau_{u^{-1}w}$ intersects the additional vertex star s_a). So $n(z_i^{-1}y_i) < n(u^{-1}w)$. Since y_i is a geodesic (being a subword of a geodesic) over S_Γ, we may apply the induction hypothesis to conclude that $|z_i| \leq f(|y_i|)$. Moreover, in Γ we have $u = w = x_0z_1x_1\cdots z_kx_k$ (as elements). So the facts that u is a geodesic and that $n \leq f(n)$ combined with the superadditivity of f yield:

$$|u| \;\leq\; \sum_{i=0}^{k}|x_i| + \sum_{i=1}^{k}|z_i| \;\leq\; \sum_{i=0}^{k}|x_i| + \sum_{i=1}^{k}f(|y_i|) \;\leq\; f\left(\sum_{i=0}^{k}|x_i| + \sum_{i=1}^{k}|y_i|\right) \;=\; f(|w|).$$

This completes the induction step and proves (1). The bound $\mathrm{Dist}_A^\Gamma(n) \preceq f(n)$ of (2) immediately follows.

For the reverse bound in (2), by the definition of Dist_C^B, there exist for each $n \geq 1$, geodesic words u_n and w_n over S_C and S_B, respectively, with $u_n = w_n$ in Γ, such that $|w_n| \leq n$ and $|u_n| = \mathrm{Dist}_C^B(n)$. Since u_n is an element of C, it is also an element of A. Let v_n be a geodesic word over S_A representing u_n. Since C is undistorted in A, there exists a constant $K \geq 1$ such that $|u_n| \leq K|v_n|$. Then, for each n, we have found a geodesic word v_n in A which represents the same element as the word w_n over S_Γ of length at most n, and $|v_n| \geq \frac{1}{K}|u_n| = \frac{1}{K}\mathrm{Dist}_C^B(n)$. It follows that $\mathrm{Dist}_A^\Gamma(n) \succeq \mathrm{Dist}_C^B(n)$. Combined with the hypothesis $\mathrm{Dist}_C^B(n) \simeq f(n)$, this gives $\mathrm{Dist}_A^\Gamma(n) \succeq f(n)$, which completes our proof of (2). $\square$

17. Height

An infinite subgroup H of a group G has *infinite height* when, for all n, there exist $g_1, \ldots, g_n \in G$ such that $\bigcap_{i=1}^n {g_i}^{-1}Hg_i$ is infinite and $Hg_i \neq Hg_j$ for all $i \neq j$. Otherwise it has *finite height*. New constructions of non-quasiconvex subgroups of hyperbolic groups are natural test cases for this longstanding question attributed to Swarup in [Mit98b]: if a finitely presented subgroup H of a hyperbolic group G has *finite height*, is H necessarily quasiconvex in G?

Therefore we note here that our examples do not speak to Swarup's question. This is because:

Proposition 17.1. *If H is the non-quasiconvex subgroup of the hyperbolic group G we construct to prove Theorem 1.1 or, more generally, to prove Theorem 1.2 in case (1) with $p > q$, then H has infinite height.*

Proof. Consider $\Gamma = F(t, x_1, x_2, y_1, y_2) \ast_{a_1,a_2}$ with the HNN-structure from Proposition 7.4, the defining relators being those specified by the $r_{4,\ast}$-cells of Figure 4.1.

We first show that $F = F(t, y_1, y_2)$ has infinite height in Γ. It is evident from the defining relators for Γ that $a_1^{-1} F a_1 \subset F$. So, for $i = 0, 1, \ldots$, we define $g_i = a_1^i$, and conclude that $g_{i+1}^{-1} F g_{i+1} \subset g_i^{-1} F g_i$. Then, for all $n \geq 0$, we have $\bigcap_{i=1}^n g_i^{-1} F g_i = g_n^{-1} F g_n$, which is a non-trivial subgroup of the free group F and so is infinite. And $F g_i \neq F g_j$ for all $i \neq j$ because, by virtue of the HNN-structure of Γ, we find that $a_1^k \in F$ only when $k = 0$. So F has infinite height in Γ.

For the G of Section 4 constructed to prove Theorem 1.1 when $k = 1$, we have $H = F(t, y_1, y_2) = F < \Gamma < G$ as a consequence of the HNN structure discussed in Section 7. When $k > 1$, the same is true because of the graph of groups structure from Definition 15.2. Since H has infinite height in Γ it has infinite height in G as well.

For the groups G we constructed to prove Theorem 1.2(1) when $p > q$, we have $G = H \ast_F K$, where K is one of the groups we constructed to prove Theorem 1.1. So $F < H$ and $F < \Gamma < K$. Moreover, the amalgamated product structure implies that $a_1^k \in H$ only when $k = 0$, so, using the same group elements g_i as before, $H g_i \neq H g_j$ when $i \neq j$. And, for all $n \geq 0$,

$$g_n^{-1} F g_n \; = \; \bigcap_{i=1}^{n} g_i^{-1} F g_i \; \subset \; \bigcap_{i=1}^{n} g_i^{-1} H g_i.$$

As $g_n^{-1} F g_n$ is infinite, we conclude that H has infinite height. □

References

[Ago13] I. Agol. The virtual Haken conjecture. *Doc. Math.*, 18:1045–1087, 2013. With an appendix by Agol, Daniel Groves, and Jason Manning.

[AO02] G. Arzhantseva and D. Osin. Solvable groups with polynomial Dehn functions. *Trans. Amer. Math. Soc.*, 354(8):3329–3348, 2002.

[BB00] N. Brady and M. R. Bridson. There is only one gap in the isoperimetric spectrum. *Geom. Funct. Anal.*, 10(5):1053–1070, 2000.

[BBD07] J. Barnard, N. Brady, and P. Dani. Super-exponential distortion of subgroups of CAT(-1) groups. *Algebr. Geom. Topol.*, 7:301–308, 2007.

[BBFS09] N. Brady, M. R. Bridson, M. Forester, and K. Shankar. Perron-Frobenius eigenvalues, snowflake groups, and isoperimetric spectra. *Geometry and Topology*, 13:141–187, 2009.

[BDR13] N. Brady, W. Dison, and T. R. Riley. Hyperbolic hydra. *Groups Geom. Dyn.*, 7(4):961–976, 2013.

[BF92] M. Bestvina and M. Feighn. A combination theorem for negatively curved groups. *J. Differential Geom.*, 35(1):85–101, 1992.

[BF96] M. Bestvina and M. Feighn. Addendum and correction to: "A combination theorem for negatively curved groups" [J. Differential Geom. **35** (1992), no. 1, 85–101; MR1152226 (93d:53053)]. *J. Differential Geom.*, 43(4):783–788, 1996.

[BH99] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-positive Curvature*. Number 319 in Grundlehren der mathematischen Wissenschaften. Springer Verlag, 1999.

[Bie81] R. Bieri. *Homological dimension of discrete groups*. Queen Mary College Mathematical Notes. Queen Mary College Department of Pure Mathematics, London, second edition, 1981.

[Bow12] B. H. Bowditch. Relatively hyperbolic groups. *Internat. J. Algebra Comput.*, 22(3):1250016, 66, 2012.

[BR09] M. R. Bridson and T. R. Riley. Extrinsic versus intrinsic diameter for Riemannian filling-discs and van Kampen diagrams. *J. Diff. Geom.*, 82(1):115–154, 2009.

[BR13] O. Baker and T. R. Riley. Cannon–Thurston maps do not always exist. *Forum Math. Sigma*, 1:e3 (11 pages), 2013.

[Bri00] P. Brinkmann. Hyperbolic automorphisms of free groups. *Geom. Funct. Anal.*, 10(5):1071–1089, 2000.

[Bro] S. Brown. CAT(-1) metrics on small cancellation groups. arXiv:1607.02580.

[BT21] N. Brady and H. C. Tran. Superexponential Dehn functions inside CAT(0) groups, 2021. `arxiv.org/abs/2102.13572`, to appear in Isreal J. Math.

[Dah03] F. Dahmani. Combination of convergence groups. *Geom. Topol.*, 7:933–963, 2003.

[Far94] B. Farb. The extrinsic geometry of subgroups and the generalised word problem. *Proc. London Math. Soc. (3)*, 68(3):577–593, 1994.

[Ger99] S. M. Gersten. Introduction to hyperbolic and automatic groups. In *Summer School in Group Theory in Banff, 1996*, volume 17 of *CRM Proc. Lecture Notes*, pages 45–70. Amer. Math. Soc., Providence, RI, 1999.

[Gro87] M. Gromov. Hyperbolic groups. In S. M. Gersten, editor, *Essays in group theory*, volume 8 of *MSRI publications*, pages 75–263. Springer–Verlag, 1987.

[Gro93] M. Gromov. Asymptotic invariants of infinite groups. In G. Niblo and M. Roller, editors, *Geometric group theory II*, number 182 in LMS lecture notes. Camb. Univ. Press, 1993.

[Gro01] M. Gromov. CAT(κ)-spaces: construction and concentration. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 280(Geom. i Topol. 7):100–140, 299–300, 2001.

[Kap99] I. Kapovich. A non-quasiconvexity embedding theorem for hyperbolic groups. *Math. Proc. Cambridge Philos. Soc.*, 127(3):461–486, 1999.

[Kap01] I. Kapovich. The combination theorem and quasiconvexity. *Intern. J. Algebra Comput.*, 11(2):185–216, 2001.

[KM02] I. Kapovich and A. Myasnikov. Stallings foldings and subgroups of free groups. *J. Algebra*, 248(2):608–668, 2002.

[Mar17] A. Martin. Complexes of groups and geometric small cancelation over graphs of groups. *Bull. Soc. Math. France*, 145(2):193–223, 2017.

[Mit98a] M. Mitra. Cannon–Thurston maps for trees of hyperbolic metric spaces. *J. Diff. Geom.*, 48(1):135–164, 1998.

[Mit98b] M. Mitra. Coarse extrinsic geometry: a survey. In *The Epstein birthday schrift*, volume 1 of *Geom. Topol. Monogr.*, pages 341–364 (electronic). Geom. Topol. Publ., Coventry, 1998.

[Ol′97] A. Yu. Ol′shanskii. On the distortion of subgroups of finitely presented groups. *Mat. Sb.*, 188(11):51–98, 1997.

[OS01] A. Yu. Ol′shanskii and M. V. Sapir. Length and area functions on groups and quasi-isometric Higman embeddings. *Internat. J. Algebra Comput.*, 11(2):137–170, 2001.

[Osi06] D. V. Osin. Relatively hyperbolic groups: intrinsic geometry, algebraic properties, and algorithmic problems. *Mem. Amer. Math. Soc.*, 179(843):vi+100, 2006.

[Pit92] Ch. Pittet. *Géométrie des groupes, inégalités isopérimétriques de dimension 2 et distorsions*. PhD thesis, Université de Genève, 1992.

[Pit93] Ch. Pittet. Surface groups and quasi-convexity. In *Geometric group theory, Vol. 1 (Sussex, 1991)*, volume 181 of *London Math. Soc. Lecture Note Ser.*, pages 169–175. Cambridge Univ. Press, Cambridge, 1993.

[Rip82] E. Rips. Subgroups of small cancellation groups. *Bull. London Math. Soc.*, 14(1):45–47, 1982.

[Sap18] M. V. Sapir. The isoperimetric spectrum of finitely presented groups. *J. Comb. Algebra*, 2(4):435–441, 2018. arXiv:1808.05840.

[SBR02] M. V. Sapir, J.-C. Birget, and E. Rips. Isoperimetric and isodiametric functions of groups. *Ann. of Math. (2)*, 156(2):345–466, 2002.

[Sho91] H. Short. Quasiconvexity and a theorem of Howson's. In *Group theory from a geometrical viewpoint (Trieste, 1990)*, pages 168–176. World Sci. Publ., River Edge, NJ, 1991.

[Wis01] D. T. Wise. The residual finiteness of positive one-relator groups. *Comment. Math. Helv.*, 76(2):314–338, 2001.

[Wis03] D. T. Wise. A residually finite version of Rips's construction. *Bull. London Math. Soc.*, 35(1):23–29, 2003.

[Wis04a] D. T. Wise. Cubulating small cancellation groups. *Geom. Funct. Anal.*, 14(1):150–214, 2004.

[Wis04b] D. T. Wise. Sectional curvature, compact cores, and local quasiconvexity. *Geom. Funct. Anal.*, 14(2):433–468, 2004.

Pallavi Dani
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803–4918, USA
pdani@math.lsu.edu, https://www.math.lsu.edu/~pdani/

Timothy Riley
Department of Mathematics, 310 Malott Hall, Cornell University, Ithaca, NY 14853, USA
tim.riley@math.cornell.edu, http://www.math.cornell.edu/~riley/