THE QUADRATIC ISOPERIMETRIC INEQUALITY FOR MAPPING TORI OF FREE GROUP AUTOMORPHISMS II: THE GENERAL CASE

MARTIN R. BRIDSON AND DANIEL GROVES

ABSTRACT. If F is a finitely generated free group and ϕ is an automorphism of F then $F \rtimes_\phi \mathbb{Z}$ satisfies a quadratic isoperimetric inequality.

1. INTRODUCTION

This is the third and final paper in a series whose purpose is to prove the following theorem.

Theorem A. *If F is a finitely generated free group and ϕ is an automorphism of F then $F \rtimes_\phi \mathbb{Z}$ satisfies a quadratic isoperimetric inequality.*

For an account of the history and context of Theorem A, we refer the reader to the introduction of [7]. We note here just one additional consequence. In [14, Theorem 2.5], Ol'shanskii and Sapir proved that if a multiple HNN extension of a free group has Dehn function less than $n^2 \log n$ (with a somewhat technical definition of 'less than') then it has a solvable conjugacy problem. Theorem A shows that free-by-cyclic groups fall into this class, and so we have the following result.

Corollary B. *If F is a finitely generated free group and ϕ is an automorphism of F then the conjugacy problem for $F \rtimes_\phi \mathbb{Z}$ is solvable.*

Corollary B was first proved in [5] using different methods.

In [7], we proved Theorem A in the case of *positive* automorphisms. That proof proceeded via an analysis of van Kampen diagrams in the

Date: 10 October, 2006.

2000 *Mathematics Subject Classification.* 20F65, (20F06, 20F28, 57M07).

Key words and phrases. free-by-cyclic groups, automorphisms of free groups, isoperimetric inequalities, Dehn functions.

The first author's work was supported in part by Fellowships from the EPSRC and by a Royal Society-Wolfson Research Merit Award. The second author's work was supported in part by a Junior Research Fellowship at Merton College, Oxford, and by NSF Grant DMS-0504251. We thank these organisations for their support.

universal cover of the mapping torus $R \times [0,1]/\langle (x,0) \sim (f(x),1)\rangle$, where R is a 1-vertex graph with fundamental group F and f is the obvious homotopy equivalence with $f_* = \phi$.

Such f are the prototypes for the *improved relative train track maps* of Bestvina, Feighn and Handel [2]. Our strategy for proving Theorem A in the general case is to refine and study these maps so as to tease-out features that allow us to adapt the crucial arguments from [7]. A vital ingredient in this approach is the identification of basic units that will play the role in the general case that single edges (letters) played in the positive case. We achieved this in [8] with the development of *beads*, whose claim to the role was clinched by the *Beaded Decomposition Theorem*.

With this technical innovation in hand, we now set about the task of adapting the arguments of [7] to the general case, following the proof from [7] as closely as possible and providing the (often fierce) technical details needed to translate each step into the more general context provided by [8]. We shall not repeat the proofs of technical lemmas from [7] when the adaptation is obvious. Nor shall we repeat our account of the intuition underlying our overall strategy of proof and intermediate strategies at key stages.

Unfortunately, at times we are obliged to break from the narrative that parallels [7] in order to deal with phenomena that do not arise in the case of positive automorphisms — Section 8, for example. But we as far as possible we have organised matters so that, having taken account of the new phenomena, we can return to the main narrative with the new phenomena controlled and packaged into concise terminology. Thus, with considerable technical exertions in our wake, we are able to arrange matters so that the final stages of the proof of our main theorem consist only of references to the corresponding sections of [7] with a brief explanation of what changes, if any, must be made in the general setting.

We have already noted that, from the analysis of improved relative train tracks in [8], it emerged that beads are the correct analogue for the role played by 'letters' in the positive case. An important manifestation of this is that Theorem A can be reduced to a statement concerning the existence of a linear bound (in terms of $|\partial\Delta|$) on the number of beads along the bottom of any corridor in a van Kampen diagram Δ in the universal cover of the mapping tori that we consider. In contrast to the positive case, however, the existence of such a bound does not immediately imply Theorem A, because there is no global bound on the length of a bead.

Nevertheless, proving a bound on the number of beads is by far the bulk of our work, occupying Sections 7–12, which closely follow [7, Sections 6–10] (with different numbering and modified structure). In Section 13 we explain how the bound on the number of beads, together with the ideas from the *Bonus Scheme* in Section 12, finally gives Theorem A. In Section 14 we explain how to deduce estimates on the geometry of van Kampen diagrams for *all* mapping tori of free group automorphism from the specially-crafted ones that we work with during our main proof. The key estimate – the linear bound on the length of t-corridors – admits the following algebraic formulation. This clarifies the manner in which our results concerning the geometry of van Kampen diagrams give rise to a non-deterministic quadratic time algorithm for the word problem in free-by-cyclic groups (for an alternative approach see [15]).

Fix a set of generators $\mathcal{B}$ for F and let d_F be the corresponding word metric. We consider words over the alphabet $(\mathcal{B} \cup \{t\})^{\pm 1}$, where t is a generator of the righthand factor of $F \rtimes_\phi \mathbb{Z}$. A *bracket* β in a word w is a decomposition $w \equiv w_1(w_2)w_3$; the subword w_2 is the *content* of β, and the initial and terminal letters of w_2 are its *sentinels*. A second bracket β', giving $w \equiv w'_1(w'_2)w'_3$ is *compatible* with β if $w'_2 \subset w_i$ for some $i \in \{1, 2, 3\}$ or $w_2 \subset w'_i$. A *t-complete* bracketing is a set of pairwise compatible brackets $\beta_1, \ldots, \beta_m$ such that the sentinels of each β_i are $\{t, t^{-1}\}$ and every $t^{\pm 1}$ in w is a sentinel of a unique bracket. In such a bracketing, the content of each bracket is equal in $F \rtimes_\phi \mathbb{Z}$ to an element of F.

Theorem C. *There exists a constant $K = K(\phi, \mathcal{B})$ such that any word $w \equiv e_1 \ldots e_n$ that represents the identity in $F \rtimes_\phi \mathbb{Z}$ admits a t-complete bracketing $\beta_1, \ldots, \beta_m$ such that the content c_i of each β_i satisfies $d_F(1, c_i) \leq Kn$.*

In an appendix to this paper we explain how our proof of Theorem A allows one to reprove the main result of [11].

We suggest that readers approach this paper as follows. First, they must be familiar with the structure of the argument in [7] and the vocabulary of beads in [8]. This will enable them to skim smoothly through Sections 2–5 of the current paper. Next, they can gain an accurate overview of the proof of Theorem A by reading the introduction to each of Sections 2–13 together with the titles of their subsections (and the introductions to subsections when they exist). There is then no alternative but to delve into the details of the proof.

Section 14 can be read independently. The argument in Appendix A is easy to understand in outline, but the proof appeals to detailed results from Sections 7, 11 and 12.

Contents

2. The Structure of Diagrams

Associated to any finite group-presentation $\Gamma = \langle \mathcal{A} \mid \mathcal{R} \rangle$ one has the standard combinatorial 2-complex $K(\mathcal{A} : \mathcal{R})$ with fundamental group Γ and directed edges labelled by the $a \in \mathcal{A}$. There is a 1-1 correspondence between words in the letters $\mathcal{A}^{\pm 1}$ and combinatorial loops in the 1-skeleton of $K(\mathcal{A} : \mathcal{R})$. Words such that $w = 1$ in Γ correspond to loops that are null-homotopic. Van Kampen's Lemma explains the connection[1] between free equalities demonstrating the membership $w \in \langle\langle \mathcal{R} \rangle\rangle$ and combinatorial null-homotopies for the corresponding loops.

Such a null-homotopy is given by a van Kampen diagram over $\langle \mathcal{A} \mid \mathcal{R} \rangle$, which is a 1-connected, combinatorial planar 2-complex Δ in $\mathbb{R}^2$ with a basepoint; each oriented edge is labelled by a generator $a_i^{\pm 1}$ with $a_i \in \mathcal{A}$ and the boundary label on each face is some $r_j^{\pm 1}$ with $r_j \in \mathcal{R}$ (read from a suitable basepoint). There is a unique label-preserving

[1]For a complete account of the equivalences in this subsection, see [6].

map from the 1-skeleton of Δ to the 1-skeleton of the standard 2-complex $K(\mathcal{A} : \mathcal{R})$, and this extends to a combinatorial map $\Delta \to K(\mathcal{A} : \mathcal{R})$.

Van Kampen's Lemma implies that the number of faces in a least-area van Kampen diagram with boundary label w is the least number N of factors among free equalities $w = \prod_{j=1}^{N} u_j r_j u_j^{-1}$. Thus the *Dehn function* of $\langle \mathcal{A} \mid \mathcal{R} \rangle$ can be defined to be the minimal function $\delta(n)$ such that every null-homotopic edge-loop of length at most n in $K(\mathcal{A} : \mathcal{R})$ is the restriction to $\partial\Delta$ of a combinatorial map $\Delta \to K(\mathcal{A} : \mathcal{R})$ where Δ is a 1-connected, planar combinatorial 2-complex. When described in this manner, it is natural to call the Dehn function the *combinatorial isoperimetric function* of $K(\mathcal{A} : \mathcal{R})$; the combinatorial isoperimetric function of an arbitrary compact combinatorial 2-complex is defined in the same way.

There is a standard diagrammatic argument for showing that the Dehn functions of quasi-isometric groups are $\simeq$ equivalent — see [1]. In that argument, it is unimportant that the complexes considered have only one vertex. Thus if K is any compact combinatorial 2-complex with fundamental group Γ, then the combinatorial isoperimetric function of K is $\simeq$ equivalent to the Dehn function of Γ. We shall exploit the freedom stemming from this equivalence. Specifically, we shall prove Theorem A by establishing a quadratic upper bound on the combinatorial isoperimetric function of a carefully-crafted 2-complex M with fundamental group $F \rtimes_{\phi^r} \mathbb{Z}$, where $r > 0$. In other words, we identify a constant $C > 0$ such that every null-homotopic combinatorial loop of length at most n in $M^{(1)}$ is the boundary of a combinatorial map to M from a 1-connected planar 2-complex with at most Cn^2 2-cells. In fact, we prove something more refined than this (see Section 4 below).

Remark 2.1. *Note that we are free to pass from $F \rtimes_{\phi} \mathbb{Z}$ to the finite-index subgroup $F \rtimes_{\phi^r} \mathbb{Z}$ because the $\simeq$ class of the Dehn function of a group is an invariant of commensurability.*

Henceforth we shall use the term *van Kampen diagram* to refer to the domain of a combinatorial map to M from a 1-connected planar 2-complex, with oriented edges *labelled* by letters representing the oriented edges of the target. (Note that this agrees with the standard terminology in the special case $M = K(\mathcal{A} : \mathcal{R})$.) Such a diagram is said to be *least-area* if it has the least number of 2-cells among all diagrams with the same boundary label.

2.1. **The Mapping Torus.** Let G be a compact graph and let $f : G \to G$ be a continuous map that sends each edge e_i of G to an

immersed edge-path $u_i = \varepsilon_1 \dots \varepsilon_m$ in G. We attach to each vertex $v \in G$ a new edge t_v joining v to $f(v)$. We then attach one 2-cell to this augmented graph for each edge e_i; the 2-cell is attached along the edge path $t_v^{-1} e_i t_{v'} u_i^{-1}$, where v and v' are the initial and terminal vertices of e_i and where the inverse is taken in the path groupoid (i.e. u_i^{-1} is u_i traversed backwards). The resulting 2-complex is the *mapping torus* of f, which we shall denote $M(f)$.

In this paper we are primarily concerned with van Kampen diagrams over $M(f)$, where f is a homotopy equivalence representing a given free-group automorphism ϕ. In this case $\pi_1(M(f)) \cong \pi_1(G) \rtimes_\phi \mathbb{Z}$. The 1-cells in such a diagram Δ_0 are either labelled by some t_u or by an edge $e \in G$. We will refer to all of the edges t_u as *t-edges* and, when it does not cause confusion, denote them simply by t. For the other edges in Δ_0, it is necessary to distinguish between the edge and its *label* in G.

Notation 2.2 (Labels $\check{\rho}$)**.** *If an edge ε in a van Kampen diagram over $M(f)$ is labelled by an edge in G, then we write $\check{\varepsilon}$ to denote that label. More generally, if an edge-path ρ in such a diagram contains no t-edges, we write $\check{\rho}$ to denote the path in G that labels ρ.*

2.2. **Time, folded t-corridors, singularities and bounded cancellation.** Assume we are in the setting of the previous paragraph. A *t-corridor* (more simply, *corridor*) is then defined exactly as in [7, Section 1.4], and we have the corresponding notion of *time* (which may be thought of as a map to $\mathbb{R}$ that is constant on non-t edges, integer-valued on vertices, and sends the endpoints of each t-edge to integers that differ by 1. As in [7, Subsections 1.5, 1.6], we see that each least-area diagram is the union of its corridors, and we may assume that the tops of all corridors are *folded*. (In Subsection 3.1 we shall specify how this folding is to be done, but for the results in this subsection it is not necessary to prescribe it.)

We write $\bot(S)$ and $\top(S)$ to denote the *top* and *bottom* of a (folded) corridor, respectively. *Singularities* are defined exactly as in [7].

We restrict our attention to least-area disc diagrams. The argument used to prove [7, Lemma 2.1] applies *verbatim* in the present setting to prove:

Lemma 2.3. *If S and S' are distinct corridors in a least-area diagram, then $\bot(S) \cap \bot(S')$ consists of at most one point.*

Let L be the maximum length of $f(E)$ for E an edge in G. As in [7, Proposition 2.3] we have

Proposition 2.4 (Bounded singularities)**.**

1. *If the tops of two corridors in a least-area diagram meet, then their intersection is a singularity.*
2. *There exists a constant B depending only on ϕ such that less than B 2-cells hit each singularity in any least-area diagram over $M(f)$.*
3. *If Δ is a least-area diagram over $M(f)$, then there are less than $2|\partial\Delta|$ non-degenerate singularities in Δ, and each has length at most LB.*

Proof. Except for one minor difficulty, the proof from [7] translates directly to the current setting. The minor difficulty is that in the current context the map f is a homotopy equivalence rather than a group automorphism, and f^{-1} is not defined as a topological map. Thus, given a path ρ, we need a canonical path σ in G such that $f_\#(\sigma) = \rho$, where $f_\#$ is tightening rel endpoints.

Consider $\widetilde{M(f)}$, the universal cover of $M(f)$. Its 1-skeleton consists of a collection of trees (copies of the universal cover of G) joined by t-edges. Consider a lift to $\widetilde{M(f)}$ of the unique edge-path $\tau_0\rho\tau_1^{-1}$ such that the τ_i are t-edges. Both endpoints of this lift lie in one of the trees $T \cong \tilde{G}$; define $\widetilde{\sigma}$ to be the unique injective path which joins them in T, and define σ to be the image of $\widetilde{\sigma}$ in $M(f)$. □

As in [7, Lemma 2.4], the above result yields as a special case (cf. [12] and [2, Lemma 2.3.1, pp.527–528]):

Lemma 2.5 (Bounded Cancellation Lemma)**.** *There is a constant B, depending only on f, so that if I is an interval consisting of $|I|$ edges on the bottom of a (folded) corridor S in a least-area diagram over $M(f)$, and every edge of I dies in S, then $|I| < B$.*

2.3. **Past, Future and Colour in Diagrams.** These concepts, for edges and 2-cells in van Kampen diagrams Δ, are defined exactly as in [7, Section 3]. The *immediate past* (or *ancestor*) of an edge at the top of a corridor in any diagram is the unique edge at the bottom of the corridor that lies in the same 2-cell; the *entire past* of an edge is defined by taking the transitive closure of the relation "is the immediate past of". The past of a 2-cell is defined similarly. The *future* of an edge e_0 is the set of edges that have e_0 in their past. The future of 2-cells is defined similarly. The evolution of edges is described by a graph $\mathcal{F}$ whose vertices are the 1-cells e of Δ, which has an edge connecting each e to its immediate ancestor. Note that $\mathcal{F}$ is a forest. Its connected components define *colours* in Δ; each edge not labelled t is assigned a unique colour, as is each 2-cell. Note that colours are in bijection with

a subset of the edges of the boundary of the diagram. The union of the 2-cells in a corridor S that have colour μ will be denoted $\mu(S)$.

As in [7], simple separation arguments yield the following observations.

Lemma 2.6. *Each $\mu(S)$ is connected and intersects each of $\top(S)$ and $\bot(S)$ in an interval.*

Lemma 2.7 (cf. Lemma 5.9, [7])**.** *Let $\varepsilon_1, \varepsilon_2$ and ε_3 be three (not necessarily adjacent) edges that appear in order of increasing subscript as one reads from left to right along the bottom of a corridor. If the future of ε_2 contains an edge of $\partial\Delta$ or of a singularity, then no edge in the future of ε_1 can cancel with any edge in the future of ε_3.*

Again following [7], given a diagram Δ we define $\mathcal{Z}$ to be the set of pairs (μ, μ') such that the coloured regions $\mu(S)$ and $\mu'(S)$ are adjacent in some corridor S. The proof of [7, Lemma 6.3] establishes:

Lemma 2.8.

$$|\mathcal{Z}| \leq 2\,|\partial\Delta| - 3.$$

3. Adapting Diagrams to the Beaded Decomposition

We refer the reader to [8] for the definitions and results which we require here about improved relative train track maps, nibbled futures, monochromatic paths, hard splittings and the language of *beads* — including (J, f)-atoms, GEPs and ΨEPs and what it means for a path to be (J, f)-beaded. We shall proceed under the assumption that the reader is familiar with each of these terms, and work axiomatically with the following outputs from [8].

Theorem 3.1 (Beaded Decomposition Theorem, [8])**.** *For every $\phi \in \mathrm{Out}(F_r)$, there exist positive integers k, r and J such that ϕ^k has an improved relative train-track representative $f_0 : G \to G$ with the property that every $(f_0)^r_\#$-monochromatic path in G is (J, f_0)-beaded.*

Beads are either monochromatic paths (in case they are atoms) or else GEPs or ΨEPs (which may be monochromatic, but do not have to be). Thus, by the above theorem and [8, Proposition 6.10], any nibbled future of a (J, f_0)-bead is (J, f_0)-beaded. Any hard splitting of an edge-path is inherited by its (nibbled) futures, by definition. And if one refines a hard splitting by decomposing the factors in a hard splitting, the result is again a hard splitting ([8, Lemma 2.6]). Thus we have:

Corollary 3.2. [8, Theorem 8.4] *Let $f = (f_0)^r_\#$ be as in the Beaded Decomposition Theorem above. If an edge-path σ in G is (J, f_0)-beaded, then any f-nibbled future of σ is (J, f_0)-beaded. In particular, $f_\#(\sigma)$ is also (J, f_0)-beaded.*

Remark 3.3. *An important point to recall from* [8] *is that the decomposition of an edge-path into (J, f_0)-beads is canonical.*

The value of the constant J in the Beaded Decomposition Theorem will be of no importance in what follows, so we drop it from the terminology. Similarly, we will fix the map f_0. Once we have passed to the power $f = (f_0)^r_\#$, the above results remain true when f is replaced by an iterate. Therefore, we refer simply to "beads" and "beaded paths".

3.1. **Refolding corridors according to the Beaded Decomposition.**

Henceforth[2]*, we consider only diagrams over the mapping torus of $M(f)$, where f is an iterate of $(f_0)^r_\#$ as in the Beaded Decomposition Theorem. In Section 5, we will fix the map f once and for all.*

We return to the matter of how best to fold the tops of corridors in least area diagrams over $M(f)$. Given an arbitrary least-area diagram, we refold the tops of corridors in order of increasing time. The process begins with edges at the minimal time on the boundary of the diagram, where there is no folding to be done provided the boundary label is reduced.

Focussing on a particular corridor S, our folding up to time(S) defines the histories of all edges up to this time and hence assigns colours to the edges on $\perp(S)$, decomposing it as a concatenation of monochromatic paths, one for each of the colours $\mu(S)$. Theorem 3.1 decomposes each of these labels as a hard splitting of beads σ_i. The hardness of the splitting means that after tightening the $f(\sigma_i)$, their concatenation will be a tightening of $f_\#(\check{\mu}(S))$. We insist that the first step in the tightening of the naive top of S, is that determined by the tightening of labels just described: i.e. we first tighten beads *within* colours, each according to a left-to-right convention (which labels inherit from the orientation of the corridors within the diagram). Then, as a second step, we tighten (again with a left-to-right convention) the concatenation of the tightened images of the colours. A diagram which is folded according to these conventions will be called *well-folded*.

The key point of this convention is that the hard splitting of the label on each colour is carried into the future — of course the futures of the

[2]There exceptions to this in Theorem 4.1, Section 14 and Appendix A

original beads may split into a concatenation of several beads, and some beads at the ends of each colour may be cancelled by interaction with neighbouring colours, but *each bead (more precisely[3], bead-labelled arc) in the beaded decomposition of each coloured interval on $\top(S)$ is contained into the future of a unique bead-labelled arc of the same colour on $\bot(S)$.* Thus $\top(S)$ is a concatenation of beads, each with a definite colour, where neighbouring beads are separated by a hard splitting if they are of the same colour but perhaps not if they are of a different colour. (It also becomes sensible to discuss the future of a bead in a [well-folded] diagram.)

We henceforth suppose (usually without comment) that our diagram has been refolded according to this convention.

Definition 3.4. *[cf. Definition 7.2] The* bead length *of $[S]_\beta$, of a corridor S in a well-folded diagram is the number of beads along $\bot(S)$.*

Remark 3.5. *It is important to note that the decomposition of $\bot(S)$ and $\top(S)$ into coloured intervals is* not *a hard splitting in general. Indeed it is the analysis of the cancellation between these intervals as one flows S forwards in time that forms the meat of this paper.*

3.2. **Abstract Futures of Beads.** Given an edge-path ρ in G, expressed as a concatenation of monochromatic edge-paths $\rho = \rho_1 \dots \rho_m$, consider the van Kampen diagram $\Delta(l, \rho)$ with boundary label equal to $t^{-l}\rho t^l \overline{f^l_\#(\rho)}$; this is a simple stack of corridors. The above convention dictates how we should fold the corridors of Δ and determines the future at each time up to l for each bead in the beaded decompositions of the ρ_i.

We define the (full) *abstract future of a bead in ρ* to be (the label on) its future in $\Delta(l, \rho)$.

4. Linear Bounds on the Length of Corridors

In any least-area diagram, each corridor has at least two edges on the boundary, namely its t-edges. The *length* of a corridor S is defined to be the number of 2-cells that it contains. The area of a least-area diagram is the sum of the lengths of its corridors, and therefore Theorem A is an immediate consequence of:

Theorem 4.1. *Let ϕ be an automorphism of a finitely generated free group and let f be a topological representative for a positive power of ϕ. There is a constant K, depending only on f, so that each corridor in a least-area diagram Δ over $M(f)$ has length at most $K|\partial\Delta|$.*

[3] we shall generally drop this cumbersome distinction in the sequel

Note that Theorem A actually depends only on establishing Theorem 4.1 for a single topological representative f^k of a suitable power of our given free group automorphism ϕ; in the next section we shall articulate what that suitable power is. The bulk of this paper will then be devoted to proving the existence of the constant K for this particular f^k. (In Section 14 we shall deduce Theorem 4.1 from this special case.)

Having restricted attention to a particular f^k, we may further restrict our attention to diagrams that are well-folded in the sense of Subsection 3.1, since refolding the corridors of an arbitrary a diagram does not change the configuration of corridors or their length. In a well-folded diagram, the top of each corridor S is a concatenation of beads, and the vast majority of our work (up to and including Section 12) goes into proving the following result.

Theorem 4.2. *If f and k are as above, then there is a constant K_1 such that all corridors S in well-folded, least-area diagrams Δ over $M(f^k_\#)$, have bead length $[S]_\beta \leq K_1\,|\partial\Delta|$.*

The linear bound on the length of S that we require for Theorem 4.1 does not follow directly from this estimate because there is no uniform bound on the length of certain beads, namely GEPs and ΨEPs. However, we shall see in Section 13 that the ideas developed in [7] to implement the bonus scheme adapt to the current setting to provide the following estimate:

Proposition 4.3. *There are constants J and K_2, depending only on f, such that the beads β on $\perp(S)$ of length greater than J satisfy*

$$\sum_\beta |\beta| \leq K_2\,|\partial\Delta|.$$

The constant J in the above statement is the one from Theorem 3.1.

5. Replacing f by a Suitable Iterate

In order to establish the bound on the length of corridors required to prove Theorem 4.1, we must analyse how corridors grow as they flow into the future and assess what cancellation can take place to inhibit this growth. This is much more difficult than in [7] because now we must cope with the cancellation that takes place within colours. But in common with our approach in [7], we can appeal to Remark 2.1 repeatedly in order to replace our topological representative f by some iterate of f that affords a more stable situation in which cancellation phenomena are more amenable to analysis.

In the present setting, we have to be a little careful about specifying what we mean by "an iterate", because we wish to consider only topological representatives whose restriction to each edge is an immersion, and this property is not inherited by powers of the map. To avoid this problem, we deem the phrase[4] *replacing f by an iterate*, to mean that for fixed $k \in \mathbb{N}$, we pass from consideration of $f : G \to G$ to consideration of the map $f^k_\# : G \to G$ that sends each edge E in G to the tight edge-path $f^k_\#(E)$ that is homotopic rel endpoints to $f^k(E)$.

When we replace f by $f^k_\#$, we leave behind the mapping torus $M(f)$ and consider instead $M(f^k_\#)$, which although homotopic to a k-sheeted covering of $M(f)$ is distinct from it.

A corridor in a van Kampen diagram over $M(f^k_\#)$ can be divided into a stack of k corridors in order to yield a van Kampen diagram over $M(f)$. This observation will play little role in our arguments, but it highlights one reason for hoping to simplify diagrams by passing to an iterate of f: the van Kampen diagrams over $M(f^k_\#)$ are a proper subset (after subdivision[5] of Δ) of the diagrams over $M(f)$; in the diagrams of this subset, corridors flow unhindered for at least k steps in time.

5.1. **Finding the desired iterate.** We have already passed to a large iterate in order to obtain the Beaded Decomposition Theorem. In the present subsection we pass to further iterates in order to control the behaviour of the images of beads.

Before settling on a specific f for the remainder of the paper, we must remove an irritating ambiguity concerning the ordering of strata in the filtration associated to the train track structure. This is required in order to render the choices in Section 6 coherent.

Definition 5.1. *Suppose that $f : G \to G$ is an improved relative train track map, and that H_i, H_j are strata for f. We say that H_i and H_j are* interchangeable *if one can reorder the strata, so that one still has an improved relative train track structure, but the order of H_i and H_j is reversed.*

If H_i and H_j are interchangeable, and $i > j$, then no iterate of any edge in H_i crosses an edge in H_j (and neither do the iterates of any edges occurring in the iterated images of edges in H_i).

[4]and obvious variations on it

[5]the obvious subdivision of a diagram Δ is called the *k-refinement*

Convention 5.2. *We suppose that for any improved relative train track map that we consider, if H_i and H_j are interchangeable strata so that H_i is an exponential stratum and H_j is a parabolic stratum then $i > j$.*

We further assume that if $H_i = \{E_i\}$ and $H_j = \{E_j\}$ are interchangeable parabolic strata and $n \mapsto |f^n(E_i)|$ grows exponentially while $n \mapsto |f^n(E_j)|$ grows polynomially, then $i > j$. And if both these functions grow polynomially, then the degree of polynomial growth of the former is at least as great as the latter.

In the following lemma, ω is the number of strata in the train track structure for f. Also recall that an edge ε in a path σ is said to be *displayed* if there is a hard splitting $\sigma = \sigma_1 \odot \varepsilon \odot \sigma_2$. The definition of a *displayed sub edge-path* is entirely analogous, and will be used later.

Lemma 5.3. *One can replace f by an iterate to ensure that if ρ is any atom then either the beads of $f^\omega_\#(\rho)$ are Nielsen paths and* GEP*s only, or else there is a displayed edge ϵ in $f^\omega_\#(\rho)$ so that*

(1) ε is of highest weight amongst all displayed edges in all $f^k_\#(\rho)$, for $k \geq 1$, and
(2) the growth of $n \mapsto |f^n_\#(\epsilon)|$ is at least as large as that of any displayed edge in any $f^k_\#(\rho)$.

Proof. Lemma 5.3 from [8] contains all but statement (2), whose validity is assured by Convention 5.2. □

Our next two results capture the *end stability* that [7, Proposition 4.5] provided in the case of positive automorphisms. This is the first stage in our analysis at which we encounter an awkward point that does not arise in [7], namely there may exist beads (more specifically atoms) ρ such that $f_\#(\rho)$ is a single vertex.

Definition 5.4. *A* vanishing bead (atom) *ρ is one with $f_\#(\rho)$ a single vertex.*

Lemma 5.5. *There exists a constant k_0, depending only on f so that the map $f_0 = f^{k_0}_\#$ satisfies the following properties. Let ρ be a non-vanishing bead, let $i \in \{1, \ldots, \omega\}$, and let σ_i be the leftmost bead in $(f_0)_\#(\rho)$ of weight at least i.*

(1) If σ_i is not a GEP *or a* ΨEP *then the leftmost bead of weight at least i in $(f_0)^j_\#(\rho)$ is the same for all $j \geq 1$. Furthermore, in this case σ_i is a single (displayed) edge or a Nielsen bead.*
(2) If σ_i is a GEP *or a* ΨEP *then the leftmost bead of weight at least i in $(f_0)^j_\#(\rho)$ is contained in the (abstract) future of σ_i for all $j \geq 1$.*

Proof. If σ is a bead then all iterated images of σ are beaded paths, and a simple finiteness argument shows that there is a bound on the number of beads which are not GEPs or ΨEPs. □

An entirely similar argument applies to rightmost beads, of course. In order to deal with the different types of beads, we also need the following variant.

Lemma 5.6. *There exists a constant k_1, depending only on f, so that the map $f_1 = f_\#^{k_1}$ satisfies the following properties. Let ρ be a non-vanishing bead and let σ be the leftmost bead in $(f_1)_\#^j(\rho)$ which is not a Nielsen bead.*

1. *If σ is not a* GEP *or a* ΨEP *then for all $j \geq 1$ the leftmost bead in $(f_1)_\#^j(\rho)$ which is not a Nielsen bead is σ. Furthermore, in this case σ is a (displayed) edge.*
2. *If σ is a* GEP *or a* ΨEP *then for all $j \geq 1$ the leftmost bead in $(f_1)_\#(\rho)$ which is not a Nielsen bead is in the future of σ.*

We are finally in a position to articulate all of the properties that we want to arrange for f by replacing it with an iterate.

Proposition 5.7. *There is a constant D_2 that depends only on f, so that if we replace f by $f_\#^{D_2}$ then,*

1. *the conclusion of* [8, Lemma 5.1] *holds with $k_1 = 1$: in particular, if ε is an exponential edge of weight i, then $f(\varepsilon)$ is longer than the unique indivisible Nielsen path of weight i (if it exists);*
2. *the conclusion of* [8, Theorem 8.1] *holds with $D_1 = 1$;*
3. *the conclusion of Lemma 5.3 holds;*
4. *the conclusions of Lemmas 5.5 and 5.6 hold; and*
5. *if ρ is a bead then $f_\#(\rho)$ contains at least three displayed copies of any exponential edge that is displayed in any $f_\#^j(\rho)$, $j \geq 1$. Moreover, the leftmost (and rightmost) such displayed edge ε is contained in a displayed path of the form $f(\varepsilon)$.*

Power Decree: *For the remainder of the paper, we will assume that $f : G \to G$ is an improved relative train track map that satisfies the properties in Proposition 5.7. We shall also operate under Convention 5.2.*

Let L be the maximal length of $f(E)$, for edges $E \in G$.[6]

[6]In [7], the symbol 'M' was used for the analogous quantity. We use L here (and in [8]) in order to avoid confusion with the mapping torus $M(f)$.

6. Preferred Futures of Beads

The reader who is comparing our progress to [7] will find that we are now in the position that we were at the start of Section 5 of that paper. Thus we now want to define the preferred future of a bead ρ (in three senses[7]) and then begin a study of fast beads.

Unfortunately, the definition of the preferred future of a bead in a diagram is much more cumbersome than the analogue in [7].

6.1. **Abstract Preferred Futures and Growth.** First we note that if beads (or more generally edge paths in G) are ever going to vanish in the sense of Definition 5.4, then they do so immediately.

Lemma 6.1. *If σ is an edge path in G and $f_\#^k(\sigma)$ is a vertex for some $k \geq 1$, then $f_\#(\sigma)$ is already a vertex.*

Proof. For all vertices $v \in G$, $f(v)$ is a fixed point of f. Therefore, the endpoints of $f_\#^j(\sigma)$ are the same for all $j \geq 1$. If $f_\#^k(\sigma)$ is a point, then the endpoints of $f_\#^k(\sigma)$ are equal, hence the tight path $f_\#(\sigma)$ is a loop. Since f is a homotopy equivalence, this loop must be trivial. □

Definition 6.2 (Abstract preferred futures)**.** *The (immediate)* preferred future *of a non-vanishing bead σ is a particular bead in the beaded decomposition of $f_\#(\sigma)$, as defined below. The k-*step preferred future *is then defined by an obvious recursion.*

1. *If σ is a* GEP *then $f_\#(\sigma)$ is also a* GEP*, and we define the preferred future of σ to be $f_\#(\sigma)$.*
2. *If σ is a* ΨEP *then either σ or $\overline{\sigma}$ has the form $\sigma = E\overline{\tau}^k\nu\gamma$. If it is σ, then by* [8, Corollary 6.11]*, $f_\#(\sigma)$ is either of the form $\sigma' \odot \xi$, where σ' is a* ΨEP *(which has the same weight as σ), or else of the form $E \odot \xi$, where E has the same weight as σ and is the unique highest weight edge in $f_\#(\sigma)$. In the first case, the preferred future of σ is σ'. In the second case, the preferred future of σ is E. The preferred future of a* ΨEP *σ where $\overline{\sigma}$ has the above form is defined in an entirely analogous way.*
3. *If σ is a Nielsen path then the preferred future of σ is $f_\#(\sigma) = \sigma$.*
4. *Finally, we consider a non-vanishing atom σ.*
   (a) If the beaded decomposition of $f_\#(\sigma)$ consists entirely of Nielsen paths and GEP*s, then we fix a highest weight* GEP *to be the preferred future of σ; otherwise, we fix a highest weight Nielsen path.*
   (b) If not, then let ε be the edge described in Lemma 5.3, fix a displayed occurrence of ε in $f_\#(\sigma)$ (in case ε is exponential,

[7] in $f_\#(\rho)$, in a diagram, and in a concatenation of beaded paths

choose a displayed occurrence that is neither leftmost nor rightmost[8]*) and define this to be the preferred future of* ε*.*

Remark 6.3. *Suppose that* ε *is an edge in* G*, considered as a bead, and suppose that* ε *is not contained in a zero-stratum. Then* ε *has a preferred future, which is an edge contained in the same stratum as* ε*. We always assume that the preferred future of* ε *is a (fixed) occurrence of* ε *in* $f_\#(\varepsilon)$ *which satisfies the requirements of the above definition. This situation is very close in spirit to the definition of preferred future in* [7]*.*

We now divide the beads into classes according to the growth of the paths $f_\#^k(\sigma)$, $k = 1, 2, \ldots$. Specifically, we define left-fast and left-slow beads in accordance with [7, Subsection 5.1].

Definition 6.4 (Left-fast beads)**.** GEP*s and Nielsen paths are* left-slow*.*

Suppose that α *is an atom or a* ΨEP*. Then* α *is* left-fast *if the distance between the left end of* $f_\#^k(\alpha)$ *and the left end of the preferred future of* α *in* $f_\#^k(\alpha)$ *grows at least quadratically with* k*, and* left-slow *otherwise.*

Note that if a ΨEP σ is left-fast then it is $\overline{\sigma}$ which it is of the form $E\overline{\tau}^k\nu\gamma$.

Remark 6.5. *We only care that fast growth be super-linear, but it happens that this is the same as being at least quadratic (cf.* [9]*).*

The concepts of *right-fast* and *right-slow* beads are entirely analogous.

6.2. **Preferred future in diagrams.** In this subsection we define the notion of 'preferred futures' within van Kampen diagrams. We also define 'biting' and 'consumption', which are the analogues in this paper of 'consumption' from [7, Section 5].

The folding convention of Subsection 2.2 expresses $\bot(S)$ as the concatenation of coloured paths $\mu(S)$, each labelled by a monochromatic path in G. The Beaded Decomposition Theorem gives us a hard splitting into beads

$$\check{\mu(S)} = \check{\beta}_1 \odot \check{\beta}_2 \odot \cdots \odot \check{\beta}_{m_\mu},$$

and it is convenient to refer to the sub-paths $\beta_i \subseteq \bot(S)$ carrying the labels $\check{\beta}_i$ as beads, as we did in Subsection 2.2.

[8]this exists by Proposition 5.7

If $\mu_1, \ldots, \mu_k$ are the colours appearing in S, in order, then the label on $\top(S)$ is obtained by tightening

$$f_\#(\mu_1\check{}(S)) \cdots f_\#(\mu_k\check{}(S)).$$

The path $f_\#(\mu_1\check{}(S)) \cdots f_\#(\mu_k\check{}(S))$ is called the *semi-naive future* of S.

We have adopted a left-to-right convention to remove any ambiguity in how one tightens the semi-naive future to obtain the label of $\top(S)$.

We previously defined the (immediate) future of a bead $\beta \subset \perp(S)$ to consist of those edges of $\top(S)$ whose immediate past lies in β. Since it is integral to what we shall do now, we re-emphasize:

Lemma 6.6. *The immediate future of a bead $\beta \subset \perp(S)$ is a (possibly empty) interval equipped with a hard-splitting into beads.*

If ρ is the immediate future of β, then ρ is also an interval in the semi-naive future of S, and hence its label $\check{\rho}$ is a specific sub-path of $f_\#(\check{\beta})$. [Note that one has more than the path $\check{\rho}$ here, one also has its position within $f_\#(\check{\beta})$; thus, for example, we would distinguish between the two visible copies of $\check{\rho}$ in $f_\#(\check{\beta}) = \check{\rho}\sigma\check{\rho}$.]

Definition 6.7 (Preferred and tenuous futures in Δ)**.** *Consider a bead $\beta \subset \mu(S) \subset \perp(S)$ in Δ whose immediate future $\rho \subset \top(S)$ determines the subpath $\check{\rho}_0$ of $\check{\beta}$ in G.*

If the (abstract) preferred future $\check{\beta}_+$ of $\check{\beta}$, as defined in Definition 6.2, is entirely contained in $\check{\rho}_0$, then the corresponding sub-path β_+ of ρ is the preferred future *of β.*

If $\check{\rho}_0$ does not contain $\check{\beta}_+$, then β does not have a preferred future. In this situation we say that the future of β is tenuous*.*

Remark 6.8. *Note that, if it exists, the preferred future of a bead $\beta \subset \mu(S)$ is a bead in the beaded decomposition of both ρ and the μ-coloured interval of $\top(S)$.*

Also, if a bead happens to be a single edge ε whose label is not contained in a zero stratum, the preferred future is a single (displayed) edge, with the same label as ε.

Definition 6.9 (Biting and consumption)**.** *If the future of a bead $\beta \subset \perp(S)$ is tenuous, we say that β is* bitten in S*. If, in the notation of (6.7),* no edge *of the preferred future of $\check{\beta}$ appears in $\check{\rho}$, then we say that β is* consumed *in S.*

Remark 6.10. *The above definition says in particular that any bead whose label is a vanishing atom is consumed.*

Let $\beta' \subset \perp(S)$ be a bead whose label is non-vanishing. If β' is bitten in S, there is a specific edge ε in the semi-naive future of S that,

during the tightening process, is the first to cancel with an edge ε' in the interval labelled by the preferred future of $\check{\beta}'$. The edge ε is in the immediate future of a bead β, necessarily of a different colour than β'.

Definition 6.11. *In the above situation, we say that* β bites β' from the left *if* β *lies to the left of* β' *in* S*, and that* β bites β' from the right *if* β *lies to the right of* β' *in* S*. We say that the edges* ε *and* ε' *discussed above* exhibit *the biting.*

The above concepts of biting and consumption replace the single, simpler, notion of consumption from [7, Section 5]: there, since the preferred future was a single edge, if it was bitten it was consumed. In [7], a frequently used concept was for an edge to be 'eventually consumed'. In this paper, we need the following replacement:

Definition 6.12. *Suppose that* $\rho_1 \subset \mu_1(S)$ *and* $\rho_2 \subset \mu_2(S)$ *are beads in* $\perp(S)$*. We say that* ρ_1 *is* eventually bitten *by* ρ_2 *if there is a corridor* S' *which contains a preferred future* β_1 *of* ρ_1 *and a bead* β_2 *in the future of* ρ_2 *so that* β_2 *bites* β_1 *in* S'*.*

With these definitions in hand, we have the following, which is an appropriate replacement for [7, Lemma 5.3].

Lemma 6.13 (cf. Lemma 5.3, [7])**.** *There exists a constant* C_0 *with the following property: if* ρ *is a bead such that* $f_\#(\rho)$ *contains a left-fast displayed edge* E *and if* $UV\rho$ *is a (tight) path with* $V\rho = V \odot \rho$ *and* $|V| \geq C_0$ *then for all* $j \geq 1$ *the preferred future of* E *is not bitten when* $f^j(UV\rho)$ *is tightened. Moreover,* $|f^j_\#(UV\rho)| \to \infty$ *as* $j \to \infty$*.*

Proof. We first prove the result in the special case that $V\rho$ is a nibbled future of a left-fast edge E_1, where ρ is the preferred future of E_1. In other words, we will prove the existence of a constant C_0' so that if $|V| \geq C_0'$ then the statement of the lemma holds for the particular path $UV\rho$. (We will later reduce to this special case.)

Note that V and $V\rho$ are monochromatic paths, and thus admit a beaded decomposition. Suppose first that V does not contain any beads of length greater than J. In this case, the proof is entirely parallel to that of [7, Lemma 5.3], where we count using the number of non-vanishing beads rather than the number of edges.

In case V contains long GEPs or long ΨEPs, we note that the cancellation by U on the left, and possibly by one of the edges in the GEP or ΨEP on the right can only decrease the length of a GEP or ΨEP by at most $2B$ at each iteration. Thus it is straightforward to include long GEPs and ΨEPs into the above calculation. We now turn to the general case.

Suppose that V is an arbitrary path so that $V\rho = V \odot \rho$. Then V can shrink of its own accord (it needn't be beaded), and can be cancelled by the future of U. However, there is certainly a constant C_0 so that if $|V| \geq C_0$ then by the time this shrinking of V combined with cancelling by the future of U can have reduced V to the empty path, the future of the edge E has at least C_0' edges to the left of its preferred future. We are then in the special case that we dealt with first. □

The following two lemmas are proved in an entirely similar manner to [7, Lemma 5.5]. Recall that displayed edges are particular types of beads, and the (abstract) preferred futures of beads were defined in Definition 6.2. Recall from Remark 6.8 that the preferred future of a displayed edge whose label is not contained in a zero stratum is a single displayed edge.

Lemma 6.14. *Let $\chi_1\sigma\chi_2$ be a tight path in G. Suppose that χ_1 and χ_2 are monochromatic and that, for $i = 1, 2$, the edge E_i is displayed in χ_i and that E_i is not in a zero stratum. Suppose that σ is a concatenation of beaded paths. Then the preferred futures of E_1 and E_2 cannot cancel each other in any tightening of $f_\#(\chi_1)f_\#(\sigma)f_\#(\chi_2)$.*

Suppose that S is a corridor in a well-folded diagram, and that $\mu_1(S)$ and $\mu_2(S)$ are non-empty paths in $\bot(S)$, where μ_1 and μ_2 are colours. Suppose further that for $i = 1, 2$ there is a displayed edge ε_i such that $\check{\varepsilon}_i$ is not contained in a zero stratum. Then the edges in the semi-naive future of S corresponding to the preferred futures of ε_1 and ε_2 do not cancel each other when folding the semi-naive future of $\bot(S)$ to form $\top(S)$.

Lemma 6.15. *Let S be a corridor and suppose that ε_1 and ε_2 are edges in $\bot(S)$ whose labels lie in parabolic strata. In the naive future of each ε_i (that is, before even the beads have been tightened), there is a unique edge ε_i' with the same label as ε_i. At no stage during the tightening of $\top(S)$ can ε_1' cancel with ε_2'.*

Corollary 6.16. *A displayed edge in any coloured interval $\mu(S)$ which is labelled by a parabolic edge $\check{E}_i \in H_i$ can only be consumed by an edge whose label is in $\overline{G \smallsetminus G_i}$.*

6.3. **Abstract paths, futures and biting.** In many of the arguments in later sections, we wish to work with concatenations of beaded paths in G rather than sides of corridors in diagrams. This is done as in Subsection 3.2 by associating to such a path $\rho = \rho_1 \dots \rho_m$, with the ρ_i beaded, the van Kampen diagram $\Delta(l, \rho)$ with boundary label $t^{-l}\rho t^l \overline{f_\#^l(\rho)}$. But we modify the usual definition of colour by defining

the colours on the bottom of the first (earliest) corridor not to be single edges but rather to be intervals labelled ρ_i. We then use the definitions of the previous subsection (biting, preferred future *etc.*) to define the associated concepts *for beads in* ρ.

We emphasize, ρ itself need not beaded; only the ρ_i are. We also emphasize that edges do not have preferred futures, only beads do.

However, some beads are single, displayed edges, and when considered as beads they do have a preferred future.

7. Counting Fast Beads

This section is the analogue of [7, Section 6]; it is here that the proof of Theorem A begins in earnest.

Let Δ be a minimal area van Kampen diagram, folded according to the convention of Section 2.2, and fix a corridor S_0 in Δ. As explained in Section 4, the core of our task is to bound the number of beads in the decomposition of $\perp(S_0)$. In order to do so, we must undertake a detailed study of the preferred futures of these beads.

First we dispense with the case that $\check{\beta}$ is a vanishing atom.

Lemma 7.1. *Suppose that $\mathcal{S}$ is the collection of beads in S_0 which are not vanishing atoms. If $\sum_{\beta\in\mathcal{S}} |\beta| = D$ then $|S_0| \leq B(D+1)$.*

Proof. This follows immediately from the Bounded Cancellation Lemma. □

Narrowing our focus in the light of this lemma, we define:

Definition 7.2 (Bead norm)**.** *Given a concatenation $\rho = \rho_1 \dots \rho_m$ of beaded paths, we define the* bead norm *of ρ, denoted $\|\rho\|_\beta$, to be the number of non-vanishing beads in the concatenation. (This is poor notation, since the norm depends on the decomposition into the ρ_i and not just the edge-path ρ. But in the contexts we shall use it, specifically $\perp(S_0)$, it will always be clear which decomposition we are considering.)*

Remark 7.3. *All beads have length at least* 1*. Thus bead norm is dominated by length. In particular, estimates concerning Bounded Singularities and Bounded Cancellation remain true when distance is replaced by bead norm; cf. Lemma 7.6.*

Remark 7.4. *An important advantage of bead norm over edge-length is that when one takes the repeated images $f^k_\#(\chi)$ of a monochromatic path, its length can decrease, due to cancellation within beads, whereas bead norm cannot.*

In Definition 3.4 we defined the bead length $[S]_\beta$ of a corridor S in a well-folded diagram. It is convenient for our future arguments to concentrate on non-vanishing atoms, and hence on bead norm rather than bead length. However, an immediate consequence of the Bounded Cancellation Lemma is the following bi-Lipschitz estimate:

Lemma 7.5. *Suppose S is a corridor in a well-folded corridor. Then*

$$\|S\|_\beta \leq [S]_\beta \leq B\|S\|_\beta.$$

7.1. **The first decomposition of S_0.** [cf. [7], Subsection 6.1]

Let β be a bead in S_0 that is not a vanishing atom. As we follow the preferred future of β forwards in time, one of the following events must occur:

1. The last preferred future of β intersects the boundary of Δ nontrivially.
2. The last preferred future of β intersects a singularity nontrivially.
3. The last preferred future of β is bitten in a corridor S.

We remark that, unlike in [7], these events are not mutually exclusive; this is because a bead can consist of more than one edge.

We shall bound the bead norm of S_0 by finding a bound on the number of non-vanishing beads in each of the three cases.

We divide Case (3) into two sub-cases:

3a. The preferred future of β is bitten by a bead that is not in the future of S_0.

3b. The preferred future of β is bitten by a bead that is in the future of S_0.

7.2. **Bounding the easy bits.** [cf. [7], Subsection 6.2]

Label the non-vanishing beads which fall into the above classes $S_0(1)$, $S_0(2)$, $S_0(3a)$ and $S_0(3b)$, respectively. We shall see, just as in [7], that $S_0(3b)$ is by far the most troublesome of these sets.

The following lemma is proved in an entirely similar way to [7, Lemmas 6.1 and 6.2], using the Bounded Cancellation Lemma and simple counting arguments.

Lemma 7.6.

(1) $\|S_0(1)\|_\beta \leq |\partial\Delta|$.
(2) $\|S_0(2)\|_\beta \leq 2B|\partial\Delta|$.
(3) $\|S_0(3a)\|_\beta \leq B|\partial\Delta|$.

We have thus reduced our task of bounding $\|S_0\|_\beta$ to bounding the numbers of beads in $S_0(3b)$, i.e. to understanding cancellation *within*

the future of S_0. The bound on the number of beads in $S_0(3b)$ is proved in an analogous way to [7], and takes up a large part of the remainder of this paper (through Section 12).

7.3. **The chromatic decomposition.** [cf. [7], Subsection 6.3]

Fix a colour μ and consider the interval $\mu(S_0)$ in $\perp(S_0)$ consisting of beads coloured μ.

We shall subdivide $\mu(S_0)$ into five (disjoint but possibly empty) subintervals according to the fates of the preferred futures of the beads.

Let $l_\mu(S_0)$ be the rightmost bead β in $\mu(S_0)$ such that $f_\#(\check{\beta})$ contains a left-fast displayed edge ϵ so that the preferred future of ϵ is eventually bitten from the left from within the future of S_0. Let $A_1(S_0, \mu)$ be the set of beads in $\mu(S_0)$ from the left end up to and including $l_\mu(S_0)$.

Let $A_2(S_0, \mu)$ consist of those beads which are not in $A_1(S_0, \mu)$ but whose preferred futures are bitten from the left from within the future of S_0.

Let $A_3(S_0, \mu)$ denote those beads which do not lie in $A_1(S_0, \mu)$ or $A_2(S_0, \mu)$ and which fall into the set $S_0(1) \cup S_0(2) \cup S_0(3a)$.

All of the beads which are not in $A_1(S_0, \mu)$, $A_2(S_0, \mu)$ or $A_3(S_0, \mu)$ must have their preferred future bitten from the right from within the future of S_0.

Analogous to the definition of $l_\mu(S_0)$, we define a bead $r_\mu(S_0)$: the bead $r_\mu(S_0)$ is the leftmost bead β' so that $f_\#(\check{\beta}')$ contains a right-fast displayed edge whose preferred future is eventually bitten from the right from within the future of S_0.

Let $A_4(S_0, \mu)$ denote those beads which are not in $A_1(S_0, \mu)$, $A_2(S_0, \mu)$ or $A_3(S_0, \mu)$ and which lie strictly to the left of $r_\mu(S_0)$.

Finally, let $A_5(S_0, \mu)$ denote those edges not in $A_1(S_0, \mu)$, $A_2(S_0, \mu)$, $A_3(S_0, \mu)$ or $A_4(S_0, \mu)$ which lie to the right of $r_\mu(S_0)$ (include $r_\mu(S_0)$ in $A_5(S_0, \mu)$ if it has not already been included in one of the earlier sets).

Now Lemma 7.6 immediately implies

Lemma 7.7.

$$\sum_\mu \|A_3(S_0, \mu)\|_\beta \leq (3B + 1)|\partial\Delta|.$$

We also have

Lemma 7.8. *Let C_0 be the constant from Lemma 6.13 above. Then*

(1) $\|A_1(S_0, \mu)\|_\beta, \|A_5(S_0, \mu)\| \leq C_0$*; and*
(2) $|A_1(S_0, \mu) \setminus l_\mu(S_0)|, |A_5(S_0, \mu) \setminus r_\mu(S_0)| \leq C_0$.

Proof. We prove the bounds only for $A_1(S_0, \mu)$, the proofs for $A_5(S_0, \mu)$ being entirely similar.

The entire future of beads in $A_1(S_0, \mu)$ other than $l_\mu(S_0)$ must be eventually consumed from the left from within the future of S_0; cf. [7, Lemma 5.9].

If $\|A_1(S_0, \mu)\|_\beta$ or $|A_1(S_0, \mu) \smallsetminus l_\mu(S_0)|$ were greater than C_0 then we would conclude from Lemma 6.13 that no left-fast bead in the immediate future of $l_\mu(S_0)$ could be bitten at any stage from the left from within the future of S_0, contrary to the definition of $l_\mu(S_0)$. □

As we continue to follow the proof from [7], our next goal is to reduce the task of bounding the bead norm of S_0 to that of bounding the number of Nielsen beads contained in $A_2(S_0, \mu)$ and $A_4(S_0, \mu)$. We focus exclusively on $A_4(S_0, \mu)$, the arguments for $A_2(S_0, \mu)$ being entirely similar.

In outline, our argument proceeds in analogy with the subsections beginning with [7, Subsection 6.4], commencing with the decomposition of $A_4(S_0, \mu)$ into subintervals $C_{(\mu,\mu')}$. But we quickly encounter a new phenomenon that requires an additional section of argument — HNP cancellation, which does not arise in the case of positive automorphisms.

7.4. **The decomposition of $A_4(S_0, \mu)$ into the $C_{(\mu,\mu')}$.** All beads in $A_4(S_0, \mu)$ are eventually bitten from the right from within the future of S_0. For a colour $\mu' \neq \mu$, define a subset $C_{(\mu,\mu')}$ of $A_4(S_0, \mu)$ as follows: given a bead $\sigma \in A_4(S_0, \mu)$, there is a bead σ' in S_0 so that σ is eventually bitten by σ'. If σ' is coloured μ' then $\sigma \in C_{(\mu,\mu')}$.

The sets $C_{(\mu,\mu')}$ form intervals in S_0.

8. HNP-Cancellation and Reapers

The results of the previous section reduce the task of bounding $\|S_0\|_\beta$ to that of establishing a bound on the sum of the bead norms of the monochromatic intervals $C_{(\mu,\mu')}$. In [7], the corresponding intervals (also labelled $C_{(\mu,\mu')}$) contained no exponential edges. In the current context, however, there may be exponential edges *trapped* in Nielsen paths, which may themselves be contained in beads of any type. This raises the concern that our attempts to control the length of the $C_{(\mu,\mu')}$ in the manner of [7] will be undermined by the *release* of these trapped edges when the Nielsen path is bitten, leading to rapid growth in subsequent nibbled futures of the Nielsen path. Our purpose in this section is to develop tools to control this situation, specifically Lemmas 8.22 and 8.23.

We must also deal with a second threat that arises from the phenomenon described in Example 8.6; we call this *Half Nielsen Path (HNP-) cancellation*.

Recall that a ΨEP is an edge path ρ in G; it is associated to a GEPand either ρ or $\bar\rho$ is of the form $E\bar\tau^k\bar\nu\gamma$ where E is an edge with $f_\#(E) = E\odot\tau^m$, where τ and ν are Nielsen paths, and $\bar\gamma\nu$ is a terminal segment of τ (and $m, k > 0$). These are the prototypes of the following types of paths.

Definition 8.1. *Suppose that E is a linear edge with $f_\#(E) = E\odot\tau^m$, where τ is a Nielsen path and $m > 0$. Suppose further that ν is a Nielsen path and γ an edge-path so that $\bar\gamma\nu$ is a terminal segment of τ.*

A PEP *is a path ρ so that either ρ or $\bar\rho$ has the form $E\bar\tau^k\bar\nu\gamma$ where $k > 0$.*

Remark 8.2. *Every* ΨEP *is a* PEP*, but an arbitrary* PEP*has no* GEP*associated to it.*

It is important to note that in the following definition the PEP being discussed is *not* assumed to be a bead in the decomposition of $\bot(S)$. (Beads along $\bot(S)$ are monochromatic whereas we want to discuss HNP cancellation, as in Definition 8.7, in the context of adjacent colours interacting.)

Definition 8.3 (HNP cancellation)**.** *Let S be a corridor in a well-folded diagram, let ε and ε' be edges in the naive (unfolded) future of $\bot(S)$ that cancel in the passage to $\top(S)$ and assume that ε is to the left of ε'.*

Suppose further that the past of ε is e with label $\check e = E$ a linear edge and that ε' is in the future of an edge e_γ whose label is an edge γ.

We call the cancellation of ε and ε' left HNP-cancellation *and write $\varepsilon\rceil\ \varepsilon'$ if the interval from e to e_γ in $\bot(S)$ (inclusive) is labelled by a* PEP *of the form $E\bar\tau^k\bar\nu\phi\gamma$, where τ is a Nielsen path so that $\tau = \xi\nu$, where ξ and ν are Nielsen paths, and $\overline{\phi\gamma}$ is a terminal sub edge-path of ξ.*

Right HNP-cancellation *is defined by reversing the roles of ε and ε' and insisting upon a* PEP *in $\bot(S)$ of the form $\bar\gamma\bar\phi\nu\tau^k\bar E$. It is denoted $\varepsilon\ \lfloor\varepsilon'$.*

When we are unconcerned about the distinction between left and right, we refer simply to HNP-cancellation.

We extend this definition to concatenations of beaded paths in G by using the obvious stack-of-corridors diagram as in Subsection 3.2.

Remark 8.4. *HNP-cancellation occurs at the 'moment of death' of the* PEP*; see* [8, Section 6] *for an explanation of the significance of this moment and an analysis of it (in the language of* ΨEP*s).*

Lemma 8.5. *Suppose that $E\overline{\tau}^k\overline{\nu}\varnothing\gamma$ is a* PEP *which exhibits an HNP-cancellation, as in Definition 8.3. Then $\varnothing$ is empty, so γ is the first edge of $\overline{\xi}$.*

Proof. The assumption that HNP-cancellation occurs means that we can restrict our attention to cancellation when tightening

$$f(E\overline{\tau}^k\overline{\nu}\varnothing\gamma).$$

This can be written as

$$E\tau^m f(\overline{\tau}^k\overline{\nu})f(\varnothing\gamma).$$

The path $\overline{\tau}^k\overline{\nu}\varnothing\gamma$ admits a hard splitting $\overline{\tau}\odot\cdots\odot\overline{\tau}\odot\overline{\nu}\odot\varnothing\gamma$. Therefore, under any choice of tightening, the m copies of τ cancel with the k copies of $f(\overline{\tau})$ (partially tightened), then with $f(\overline{\nu})$; they then begin to interact with $f(\varnothing\gamma)$. Just as in the proof of [8, Proposition 6.9], under the assumptions of [8, Lemma 5.1], there is only a single edge in $\varnothing\gamma$ whose future can interact with $f(E)$ when tightening. □

We now present the deferred example that explains the need to consider HNP-cancellation. This will also lead us to a further definition — *HNP biting* — that encodes a genuinely troublesome situation where HNP cancellation must be accounted[9] for. Fortunately, many other instances of HNP-cancellation are swept-up by our general cancellation and finiteness arguments, allowing us to avoid a detailed analysis of the possible outcomes.

The problem at the heart of the following example did not arise in [7] because the natural realisation of a positive automorphism does not map any linear edge across other linear edges.

Example 8.6. *Suppose that u is a Nielsen path, and that E_1 and E_2 are edges so that $f(E_i) = E_iu^k$ for $i = 1, 2$ and some integer $k > 0$. For any integer j, the path $\tau_j = E_1u^j\overline{E_2}$ is an indivisible Nielsen path.*

Suppose that E_3 is an edge so that $f(E_3) = E_3\tau_j^l$, for some integers j and l (with $l > 0$). For ease of notation, we will assume that $l = 1$.

Consider the path $\rho = E_3\overline{\tau}_j^rE_2$, for some $r > 0$. Then ρ is a PEP*.*

In the iterated images $f_\#(\rho)$, the visible copy of E_2 has a unique future labelled E_2, which we will call the 'preferred future' of E_2 for the

[9]We usually account for it by excluding it from our definitions. When it cannot be excluded, we often sidestep it, using the notions of 'robust future' and 'robust past' given in Definitions 8.12 and 8.13 below.

purposes of this example. After $r+1$ iterations of ρ under $f_\#$ (and any choice of tightening at each stage), the future of E_3 cancels the preferred future of the visible copy of E_2. If we encode the evolution of ρ in a stack diagram as in Subsection 3.2 then the cancellation of E_2 is HNP-cancellation.

In the following discussion, we assume that the reader is familiar with [7], in particular the vocabulary of teams and reapers.

The phenomenon described in the above example causes problems when the sub-path $\rho_1 = \overline{\tau}_j^r E_2$ of ρ is monochromatic and E_2 is displayed in ρ_1. In this situation, it shows that the most obvious adaptation of [7, Lemma 6.7] would be false. It is for this reason that we must exclude HNP-biting in Definition 9.7.

Similarly, because Example 8.6 renders a naive version of the results of [7, Section 8] false, HNP-biting must be excluded from the Two Colour Lemma and the associated results in Section 10.

A situation in which we cannot exclude HNP-biting by decree arises in the analysis of teams and in particular the definition of a *reaper* (subsection 8.3). Suppose that ρ labels some interval in the bottom of a corridor, with many copies of $\overline{u}$ to its immediate right. In this case, the edge ε_2 labelled E_2 will consume copies of $\overline{u}$ in the first r units of time, but its future will then be cancelled (assuming no other cancellation occurs from either side, and that there are no singularities, etc.). Since ε_2 was acting as the reaper of a team, we must find a continuing manifestation of it at subsequent times, for otherwise we will lose control over the length of teams (r being arbitrary) and the structure of our main argument will fail. This problem is solved by introducing the *robust future* of ε_2 (Definition 8.12), which in this case is an edge labelled E_1 that 'replaces' the preferred future of ε_2 when it is cancelled.

Definition 8.7. *Suppose that χ_1 and χ_2 are beaded paths in G and $\chi_1\chi_2$ is tight. Suppose that there is a bead $\rho_1 \subset \chi_1$ and a bead $\rho_2 \subset \chi_2$ so that*

1. *either ρ_1 is a displayed edge γ in χ_1 which is linear or else ρ_1 is a displayed ΨEP in χ_1 of the form $E\bar{\tau}^k\bar{\nu}\gamma$, where γ is a linear edge;*
2. *when tightening $f_\#(\chi_1)f_\#(\chi_2)$ to form $f_\#(\chi_1\chi_2)$, ρ_1 bites ρ_2 and the edge ε' in the exhibiting pair $(\varepsilon', \varepsilon)$ (see Definition 6.11) is in the future of γ;*
3. *moreover[10], $\varepsilon' \rhd \varepsilon$.*

[10]The PEP implicit in the symbol $\rhd$ is not the ΨEP in (1).

Under these circumstances we say that ρ_2 is left-HNP-bitten *by ρ_1 and we write $\rho_1 \blacklozenge \rho_2$. There is an entirely analogous definition of* right-HNP-biting *$\rho_1 \blacklozenge \rho_2$, and when we are unconcerned about the direction we will refer simply*[11] *to* HNP-biting.

We make the analogous definition for HNP-biting within diagrams.

Definition 8.8. *Suppose that χ_1 and χ_2 are beaded paths and that ρ_1 is a bead in χ_1. We say that ρ_1 is* eventually HNP-bitten *by χ_2 if ρ_1 is eventually bitten by χ_2 (Definition 6.12) and this biting is HNP-biting.*

We make the analogous definition within diagrams.

Definition 8.9. *Suppose that E and E' are edges in G. We say that E and E' are* indistinguishable *if there is a Nielsen path τ and an integer $s > 0$ so that $f(E) = E\tau^s$ and $f(E') = E'\tau^s$.*

The edges E_1 and E_2 in Example 8.6 are indistinguishable.

8.1. **Parabolic HNP-cancellation and robust futures.** The following is a simple (but key) observation, and has an obvious application to HNP-cancellation of edges of parabolic weight.

Lemma 8.10. *Suppose that τ, ν, ν' and σ are Nielsen paths, with σ irreducible and $\tau = \nu'\overline{\sigma}\nu$. Suppose further that γ is the initial edge of σ, and that $f(\gamma) = \gamma \odot \xi^l$ for some Nielsen path ξ. Then σ has the form $\gamma\xi^r\overline{\gamma'}$ where r is some integer and γ' is an edge so that γ and γ' are indistinguishable.*

Moreover, suppose that E is an edge so that $f(E) = E \odot \tau^m$, and let $\rho = E\overline{\tau}^i\overline{\nu}\gamma$ be a PEP *with $0 \leq i < m$. Then $f_\#(\rho)$ has the form $E \odot \tau^{m-i-1}\nu'\gamma'\overline{\xi}^j$ where γ and γ' are indistinguishable.*

Proof. The first assertion is an immediate consequence of the structure of indivisible Nielsen paths of parabolic weight, and the second is then obvious (a detailed analysis of the Nielsen paths of parabolic weight is undertaken in [8, Section 1]). □

Definition 8.11. *In general, non-displayed edges ε in diagrams do not have preferred futures. But if $\check{\varepsilon}$ has parabolic weight, there is a unique edge of the same weight in $f_\#(\check{\varepsilon})$, and it is natural to define the (immediate) preferred future of ε to be the corresponding edge in the immediate future of ε. (If ε happens to be displayed, this agrees with our earlier definition.)*

[11] We swap orientation in Definition 8.8 so as to emphasize this point immediately.

In Section 10, when proving the Pincer Lemma, we will have to exclude HNP-biting. This will also be the case in the applications of the Pincer Lemma in Sections 11 and 12. Thus, in following the future of a linear edge γ when HNP-cancellation occurs, we would like to ignore the preferred future (which disappears), and rather follow the future of the interchangeable edge γ' from Lemma 8.10 above. Thus we make the following

Definition 8.12 (Robust Futures for Parabolic Edges)**.** *Suppose that ε is a (not necessarily displayed) edge in a colour $\mu(S)$, and that $\check{\varepsilon}$ is contained in a parabolic stratum. If the preferred future of ε is cancelled from the left [resp. right] by HNP-cancellation in $\top(S)$, then Lemma 8.10 provides an edge γ' that is indistinguishable from $\check{\varepsilon}$ and survives in the tightened path $f_{\#}(E\bar{\tau}^k\bar{\nu}\emptyset\gamma)$ [resp. its reverse] considered in Definition 8.3.*

We define the robust future *of an edge $\varepsilon \subseteq \bot(S)$ as follows. If the preferred future of ε survives in $\top(S)$, then the robust future of ε is just the preferred future of ε. If the preferred future is cancelled by HNP-cancellation, then the robust future of ε is the above edge labelled γ', provided this survives in $\top(S)$. Otherwise there is no robust future.*

Definition 8.13 (Robust Pasts for Linear Edges)**.** *Let ε' be an edge of $\top(S)$ and suppose that both it and its immediate past are labelled by linear edges. If ε' is not the robust future of any edge then the robust past of ε' is the past of ε'. But if ε'* is *the (immediate) robust future of ε then the robust past of ε' is ε.*

Just as for preferred futures, the notions of robust future and robust past can be extended arbitrarily many steps forwards or backwards in time by iterating the definition.

8.2. **A setting where we require cancellation lemmas.** Consider the following situation. Let $\chi_1\sigma\chi_2$ be a tight path in G with χ_1 and χ_2 monochromatic and σ a path with a preferred decomposition into monochromatic paths (each of which comes equipped with a beaded decomposition). We will analyse the possible interaction between χ_1 and χ_2 in iterates of $\chi_1\sigma\chi_2$ under f (where the tightening follows the convention of Subsection 6.3).

As ever, the following lemma remains valid with left/right orientation reversed.

Lemma 8.14. *Suppose that χ_1, χ_2 and σ are as above, and suppose that each non-vanishing bead in χ_2 is eventually bitten by a bead from χ_1 in some iterated image $f^k_{\#}(\chi_1\sigma\chi_2)$ of $\chi_1\sigma\chi_2$.*

Suppose further that ρ is a bead in χ_2 so that $f_\#(\rho)$ has parabolic weight, and that ρ is eventually left-HNP-bitten by a bead from χ_1 in the evolution of $\chi_1\sigma\chi_2$. Then ρ is the rightmost non-vanishing bead in χ_2.

Proof. Pass to the iterate $f_\#^{k-1}(\chi_1\sigma\chi_2)$ so that the preferred future of ρ lies in a PEP π, which exhibits the (eventual) HNP-biting of ρ in the tightening to form $f_\#^k(\chi_1\sigma\chi_2)$. Let ρ_1 be the preferred future of ρ in $f_\#^{k-1}(\chi_1\sigma\chi_2)$. Since $f_\#(\rho)$ has parabolic weight, ρ_1 has parabolic weight, and is either a displayed edge or a displayed ΨEP or GEP. We must prove that no bead to the right of ρ_1 is eventually bitten by the future of χ_1.

By Definition 8.7 and Lemma 8.5 the PEP π has the form $\gamma\bar\tau^k\bar\nu\varepsilon$, where

(1) γ is an edge so that $f(\gamma) = \gamma \odot \tau^m$;
(2) γ is either a displayed edge in the future of χ_1 in $f_\#^{k-1}(\chi_1\sigma\chi_2)$ or else if the rightmost edge in a displayed ΨEP; and
(3) ε is contained in ρ_1.

Let α be the displayed edge or ΨEP containing γ.

Let ρ_1' be the terminal part of ρ_1 from ε to its right end, and let χ_2' be the terminal part of the future of χ_2 in $f_\#^{k-1}(\chi_1\sigma\chi_2)$, from ε to its right end.

Since ρ_1 is displayed, we have $\chi_2' = \rho_1' \odot \beta$ for some path β.

By Lemma 8.10, when tightening to form $f_\#^k(\chi_1\sigma\chi_2)$, the edge ε is replaced by an indistinguishable edge ε' which comes from the future of α. Suppose that δ is that part of $f_\#(\alpha\rho_1')$ from ε' to the right end. Since α is a (linear) edge or a ΨEP, the edge ε' survives in all iterates of α (under any choices of cancellation. Similarly, since ε and ε' are indistinguishable, ε' survives in all iterates of δ (under any choices of tightening). This implies that we have a hard splitting $f_\#(\alpha\chi_2') = f_\#(\alpha\rho_1') \odot f_\#(\beta)$, and the fact that α is displayed implies that no bead in β can be eventually bitten by the future of χ_1, as required. □

In applications of Lemma 8.14 (and of Lemmas 8.22 and 8.23 below), we usually take $\chi_1 = \mu_1\check{}(S)$ and $\chi_2 = \mu_2\check{}(S)$, where μ_1 and μ_2 are colours and S is some corridor, and we will choose σ to be the label of that part of $\perp(S)$ which lies strictly between $\mu_1(S)$ and $\mu_2(S)$.[12] Since the folding conventions of Subsections 2.2 and 6.3 are compatible, and because of the hardness of our splittings, the interaction between μ_1

[12]However, it will also be convenient sometimes to take χ_1 to be a subinterval of $\mu_1\check{}(S)$ consisting of an interval of beads.

and μ_2 in the future of S can be analysed by studying the interaction between the futures of χ_1 and χ_2 in iterated images of $\chi_1\sigma\chi_2$ under f.

8.3. **Reapers.** In [7] proving the existence of reapers was straightforward (see [7, Section 9]). In the current context, however, we have to work harder to prove that a suitable incarnation of a reaper exists, because of the phenomena discussed in the preceding subsection. At the heart of our difficulties is the fact that Nielsen atoms need not be single edges.

Definition 8.15. *A* beaded Nielsen path *in a corridor S is a subinterval $\sigma \subset \bot(S)$ so that $\check{\sigma}$ is a beaded path all of whose beads are Nielsen paths.*

Note that in the above definition we do not assume that σ is a single colour, or even that each bead in $\check{\sigma}$ is contained in a single colour. Examples of beaded Nielsen paths include that part of a GEP between the extremal edges, and the sub-paths $\overline{\tau}^i$ of a PEP $E\overline{\tau^k}\bar{\nu}\phi\gamma$.

Although the beads in a beaded Nielsen path might not be displayed in a path $\mu(\check{S})$, it is still possible to define the future of a bead in a beaded Nielsen path, and the notions of preferred future and biting still make sense. We will use this observation in the sequel.

The following notion is parallel to that of [7, Definition 10.1], which was pivotal in the bonus scheme (cf. Section 12 below). Here, it plays a more central role.

Definition 8.16 (Swollen present and swollen future)**.** *Suppose S is a corridor and that $I \subseteq \bot(S)$ is a beaded Nielsen path in S. The* swollen present *of I is the*[13] *maximal subinterval $I' \subseteq \bot(S)$ such that (i) $I \subseteq I'$; (ii) I' is a beaded Nielsen path in S; and (iii) the beads of I are beads of I'.*

The left swollen present *of I is that part of the swollen present from the left end up to the right end of I, whilst the* right-swollen present *goes from the left end of I to the right end of the swollen present.*

If the actual future of I is a beaded Nielsen path the (immediate) swollen future $sw_1(I)$ *of I is the swollen present of the (actual) future of I. With a similar qualification, the* swollen future $sw_k(I)$ *at* time$(S)+k$ *is defined to be $sw_1(sw_{k-1}(I))$.*

With the same qualifications, the left and right swollen futures are defined in the obvious ways.

[13] uniqueness is immediate from the observation that if a terminal sub-path σ of a Nielsen path τ is itself Nielsen then σ is a concatenation of beads in τ.

The first qualification in the above definition is required because it is possible that the immediate future of a beaded Nielsen path is not a beaded Nielsen path. Thus we must be careful only to apply this concept in cases where we know the swollen future to exist.

Definition 8.17 (Reapers)**.** *Suppose that S is a corridor and $I \subset \perp(S)$ is a beaded Nielsen path in S with nonempty swollen future $sw_1(I)$. Suppose that α is an edge in $\perp(S)$ immediately adjacent to I on the left. We say that α is a* left-reaper for I *if (i) $\check{\alpha}$ is a linear edge; (ii) $\check{\alpha}$ bites some of the future of $\check{I}$ in $f_\#(\check{\alpha}I)$; and (iii) the robust future of α is immediately adjacent to $sw_1(I)$ in $\top(S)$.*

There is an entirely analogous definition of right-reapers*. As usual, when we are unconcerned about the direction we will refer to* reapers*.*

Definition 8.18 (Left-edible)**.** *Let S be a corridor in a well-folded diagram, and $I \subset \perp(S)$ a beaded Nielsen path. We say that I is* left-edible *if each bead in I is eventually bitten by a bead coloured μ in the future of S, where $\mu(S)$ lies to the left of I.*

Right-edible *paths are defined with a reversal of the left-right orientation.*

In the remainder of this section we work towards proving Propositions 8.19 and 8.21.

Proposition 8.19. *Let S be a corridor in a well-folded diagram and $I \subset \perp(S)$ a left-edible path so that $|I| \geq B + J$. Then the immediate future of I in $\top(S)$ is left-edible.*

The following lemma is straightforward, and allows us to focus our attention on the time when cancellation between colours begins.

Lemma 8.20. *Let S be a corridor in a well-folded diagram and let $I \subset \perp(S)$ be a left-edible colour, all of whose beads are eventually bitten by beads coloured μ. Let S^I be the corridor in the future of S so that the first biting of a bead in the left swollen future of I by something coloured μ occurs in S^I. Then the left swollen future of I in $\perp(S^I)$ is left-edible.*

In the following statement B is the Bounded Cancellation Constant from Proposition 2.5 and J is the constant from the Beaded Decomposition Theorem 3.1. The corridor S^I is as in Lemma 8.20 above, and I^λ is the left swollen future of I in S^I.

Proposition 8.21. *Suppose that S is a corridor in a well-folded diagram and $I \subset \perp(S)$ is a left-edible path, all of whose beads are eventually bitten by beads coloured μ. Suppose also that $|I| \geq B + J$. Then*

(1) *the immediate future of I^λ in $\top(S^I)$ has an associated left reaper α, which is coloured μ; and*
(2) *for each bead in the immediate future of I^λ, when it is eventually bitten the biting is by the robust future of α.*

8.4. **Two Cancellation Lemmas.** The following lemma is useful in the proof of Lemma 9.8 below. We record it now because a variation on it (Lemma 8.23) is needed in the proof of Proposition 8.21.

We revert to the setting described in Subsection 8.2.

Lemma 8.22. *Assume that in the iterates of $\chi_1\sigma\chi_2$ (i.e. forward-images under $f_\#$) each bead in χ_2 is eventually bitten by a bead in χ_1. Suppose that χ_2 has weight i, where H_i is an exponential stratum, and that all beads of weight i in χ_2 are Nielsen beads. Let ρ be a bead of weight i in χ_2.*

(1) *If ρ is not bitten in $f_\#(\chi_1\sigma\chi_2)$ but is eventually bitten in the image $f_\#^k(\chi_1\sigma\chi_2)$ then ρ is entirely consumed in $f_\#^k(\chi_1\sigma\chi_2)$.*
(2) *If ρ is bitten but not entirely consumed in $f_\#(\chi_1\sigma\chi_2)$ then ρ is the rightmost bead in χ_2.*

Proof. There is at most one indivisible Nielsen path of weight i and the lemma is vacuous unless there is exactly one.

Let β be a bead in χ_2 of weight i, and suppose that an edge η in the future of χ_1 is the edge which cancels the rightmost edge in the preferred future of β to exhibit the biting of β by χ_1. Since β is an indivisible Nielsen path, it has edges of weight i on both ends, as does its preferred future, and so η has weight i. Suppose that the past of η in $\chi_1\sigma\chi_2$ has weight i. Then by [8, Theorem 8.1] and Assumption 5.7, η is either a displayed edge in the future of χ_1, or else is contained in a Nielsen bead. Suppose first that η is contained in a Nielsen bead τ. Since η is to cancel with an edge in β, the path τ must have weight i. Hence $\tau = \bar{\beta}$, and β is entirely consumed when it is bitten.

Suppose then that η is displayed in the future of χ_1. By Assumption 5.7.(5) we may assume that the edge η is contained in a displayed path of the form $f(\eta)$. Since $f(\eta)$ is i-legal, and β is not, it is not possible for the illegal turn in β (of weight i) to be cancelled by any iterates of η. However, $|f(\eta)| > |\beta|$, by Assumption 5.7(1), so it is not possible for the displayed copy of $f(\eta)$ to be cancelled by the future of β. Therefore, in this case β must be the rightmost bead in χ_2.

Furthermore, suppose that β and η are as above, and the past of η in $\chi_1\sigma\chi_2$ has weight i, and suppose moreover that β is not bitten in $f_\#(\chi_1\sigma\chi_2)$. Then β is bitten by η in some $f_\#^k(\chi_1\sigma\chi_2)$, and $k \geq 2$. Thus we may assume that the immediate past of η is also displayed and is η.

By applying Lemma 5.5 and noting that the rightmost edge of β must be $\bar{\eta}$, we see that the sub-path between the immediate past of β and the immediate past of η has the form $\cdots\bar{\eta}\omega\eta\cdots$ for some path ω. The path ω must start and finish at he same vertex, and in order for the written copy of $\bar{\eta}$ to cancel with the written copy of η it must be that $f_\#(\omega)$ is a point. However, ω is not a point, because otherwise the past of β and the past of η would already cancel. This contradicts the fact that f is a homotopy equivalence. The same argument shows that if η is contained in a Nielsen bead and β is not bitten in $f_\#(\chi_1\sigma\chi_2)$ then β cannot be bitten by η.

Therefore, if β is bitten by an edge η whose past in $\chi_1\sigma\chi_2$ has weight i then β is close to the left end of χ_2, and is either entirely consumed when bitten or is the rightmost bead in χ_2.

We may now assume that the bead ρ is cancelled by an edge η whose past in χ_2 has weight greater than i. The above arguments show that we may assume that the immediate past of η also has weight greater than i, and by Lemma 5.5 we may assume that this past is contained in a displayed edge, a GEP, or a ΨEP. It is easy to see that the immediate past of η cannot have exponential weight and cannot be a GEP. Thus we may assume that the immediate past of η is either the edge on the left end of a ΨEP of the form $\gamma\nu\tau^k\overline{E}$, (and that the edge γ is parabolic) or else is displayed and parabolic.

Lemma 5.5 and the above arguments imply that this immediate past of η must be a linear edge, and the above arguments now imply that if ρ is bitten in a corridor it must be entirely consumed. □

The following variant of Lemma 8.22 is the one we need in the proof of Proposition 8.21. We continue to study $\chi_1\sigma\chi_2$ as in Subsection 8.2.

Lemma 8.23. *Suppose that χ_2 is a beaded Nielsen path and each of its beads is eventually bitten by a bead in χ_1 in some iterated image of $\chi_1\sigma\chi_2$ under f.*

Let ρ be a bead in χ_2 which is not bitten in $f_\#(\chi_1\sigma\chi_2)$. If ρ is bitten but not consumed in some iterated image of $\chi_1\sigma\chi_2$ then ρ is the rightmost bead in χ_2.

Proof. We follow the proof of Lemma 8.22 above, with the added wrinkle that there may be parabolic weight Nielsen paths to consider in χ_2. In this case there needn't be a unique Nielsen path of weight i.

Suppose that ρ is as in the statement of the Lemma. If ρ has exponential weight, then the arguments of the proof of Lemma 8.22 give the required properties. If ρ has parabolic weight, Lemma 6.15 implies that when ρ is bitten by an edge η in the future of χ_1, the immediate past of

η has weight greater than that of ρ. Also, this immediate past must be parabolic. Arguing as in the proof of Lemma 8.22, one sees that either ρ is entirely consumed when bitten, or else ρ is the rightmost bead in χ_2. □

Corollary 8.24. *Suppose that I is a beaded Nielsen path in $\bot(S)$ for some corridor S of a well-folded diagram, and suppose that all beads of I are eventually bitten from the left by beads in a single colour μ. Then, with the possible exception of B beads on the left end and one bead on the right (the final one bitten), whenever μ bites a Nielsen bead in the future of I, it consumes it entirely.*

Proof of the Proposition 8.19

Proof. If the immediate future of I in $\top(S)$ were not left-edible, then Corollary 8.24 would ensure that no bead in I which is not bitten in S is ever bitten by μ. However, the assumption on the length of I (and the Bounded Cancellation Lemma) ensure that there *are* beads in I not bitten in S. The fact that I *is* left-edible therefore ensures that the future of I in $\top(S)$ is also left-edible. □

Proof of the Proposition 8.21

Proof. Let S' be the corridor containing the immediate past of I^λ. Lemma 8.23 implies that in $\top(S')$ there is an edge ρ in μ which cancels a whole Nielsen path in the future of I.

Since $|I| \geq B + J$, there is a bead in I not bitten in $\top(S)$. The proof of Lemma 8.23 now implies that there is a reaper as in the statement of the proposition. □

9. Non-fast and Unbounded Beads

With the technical exertions of the previous section behind us, we are now able to return to the main argument, picking up the flow of [7] at Subsection 6.6. Thus our next purpose is to reduce the task of bounding the bead norm of the intervals $C_{(\mu,\mu')}$ to that of bounding the lengths of certain long blocks of Nielsen atoms. These blocks are the analogue of the intervals $C_{(\mu,\mu')}(2)$ from [7], and will be the building blocks of the *teams* introduced in Section 11 (in analogy with [7, Section 9]).

Definition 9.1. *Suppose that $\rho = \gamma\nu\tau^k\overline{E_i}$ is a* PEP *(with $k \geq 0$). We say that ρ is* left-slow *if γ is empty or a concatenation of left-slow beads.*

There is an entirely analogous definition of right-slow PEP*s of the form* $\rho = E_i\overline{\tau}^k\overline{\nu\gamma}$.

Often, we will just speak of *slow* PEPs, since a single PEP can only be left-slow or right-slow, but not both.

Definition 9.2. *Suppose that the bead* ρ *is such that* $f_\#(\rho)$ *is not a Nielsen bead. Then the function* $n \mapsto |f_\#^n(\rho)|$ *grows at least linearly. In this case, we call* ρ *an* unbounded bead.

Definition 9.3. *A beaded path is called* right-tame *if all of its beads are* GEP*s, slow* ΨEP*s, Nielsen paths and atoms which do not have a right-fast displayed edge in their immediate future.*

The next lemma follows immediately from the definition.

Lemma 9.4. $A_4(S_0, \mu)$ *is a right-tame path.*

Lemma 9.5. *Suppose that* α *is a non-vanishing atom which is not right-fast. Then either all of the beads in* $f_\#(\alpha)$ *are Nielsen paths and* GEP*s, or else the preferred future of* α *is parabolic.*

Proof. The only modification to Lemma 5.3 is the exclusion of exponential edges in the second case, which is valid because such an edge would obviously contradict the fact that α is not right-fast. □

Definition 9.6. *Suppose that* σ *is a right-tame path. The* untrapped weight *of* σ *is the largest* j *so that* $f_\#(\sigma)$ *contains a bead of weight* j *which is not Nielsen.*

Definition 9.7. *Suppose that, for some pair* $(\mu, \mu') \in \mathcal{Z}$*, the untrapped weight of* $C_{(\mu,\mu')}$ *is* j*. For each* $1 \leq i \leq j$*, define* ρ_i *to be the leftmost bead in* $C_{(\mu,\mu')}$ *so that* $f_\#(\rho_i)$ *has an unbounded bead of weight at least* i *that is not HNP-bitten in the future of* S_0*.*[14]

Let $\mathcal{E}_i$ *denote those beads in* $C_{(\mu,\mu')}$ *from the right end up to and including* ρ_i*, and let* $\mathcal{D}_i = \mathcal{E}_i \smallsetminus \mathcal{E}_{i+1}$.

The following is the analogue of [7, Lemma 6.7]

Lemma 9.8. *For all* $1 \leq i \leq \omega$ *there is a constant* $C_1(i)$ *so that for each of the paths* $C_{(\mu,\mu')}$ *and decomposition into intervals* $\mathcal{D}_i$ *as above, we have*

$$\|\mathcal{D}_i\|_\beta \leq C_1(i).$$

Proof. As far as possible, we try to follow the proof of [7, Lemma 6.7]. However, due to the phenomena described in Section 8, the proof here is somewhat more complicated.

[14]Note that it is possible that $\rho_i = \rho_{i+1}$ for some i.

We go forward to the time, t say, which is one step before the moment when μ' first starts to bite the preferred futures. By virtue of Remark 7.4, and the definition of $\mathcal{D}_i$, there are at least as many beads in the future of $\mathcal{D}_i$ at time t as there are in S_0. Therefore, it is sufficient to bound the number of beads in the future of $\mathcal{D}_i$ at time t; to ease the notation, we write $\mathcal{D}_i$ for this future, i.e. pretend that $t = \mathrm{time}(S_0)$.

It is possible that there exist beads $\rho \in \mathcal{D}_i$ so that $f_\#(\rho)$ has weight greater than i. In such a case, all of the beads in $f_\#(\rho)$ of weight greater than i are Nielsen beads.

Consider the highest weight k for which there is a bead ρ in $\mathcal{D}_i$ with $f_\#(\rho)$ of weight k, and suppose that $k > i$. Suppose first that ρ has exponential weight. Then by Lemma 8.22 either $\mathcal{D}_i$ has bead norm at most B (and length at most $\ell = JB(B+1)$), or else ρ is entirely consumed when it is bitten. In the first case ρ is the leftmost bead in $\mathcal{D}_i$, and also in $C_{(\mu,\mu')}$. A similar argument applies when ρ has parabolic weight.

Thus, excluding cases where $|\mathcal{D}_i| < \ell$, we may treat the Nielsen beads of weight higher than i as indivisible units, which are entirely consumed when bitten. We are therefore in the situation of the proof of [7, Lemma 6.7], where the unbounded beads in $\mathcal{C}_i$ grow apart at a linear rate, and so must be cancelled quickly. Otherwise, the proof is entirely parallel to the one from [7]. □

We are trying to reduce the task of bounding the bead norm to that of bounding the size of intervals consisting entirely of Nielsen beads, which are each consumed by a reaper. In order to make this reduction, we still have some HNP-biting to deal with. In order to deal with this, we need an analogue of [7, Lemma 9.4].

Recall that L is the maximal length of $f(E)$ where E is an edge in G.

Proposition 9.9 (cf. Lemma 9.4, [7])**.** *There is a constant C_4 depending only on f which satisfies the following properties. If I is an interval on $\top(S)$ labelled by a beaded path all of whose beads are Nielsen atoms, then the path labelling the past of I in $\bot(S)$ is of the form $u\alpha v$ where α is a beaded path all of whose beads are Nielsen atoms and $|u|$ and $|v|$ are less than C_4.*

If the past of I begins (respectively ends) with a point fixed by f, then u (respectively v) is empty.

In particular, $|I| \leq |\alpha| + 2LC_4$.

Proof. The interval $I \subset \top(S)$ is a beaded path, all of whose beads are Nielsen paths of length at most J. Therefore, along I there are points

where I admits a hard splitting and these points occur with a frequency of at least one every J edges. Since these points are vertices, the set of labels of points at which the splitting occurs is finite. Consider the path from $\top(S)$ to $\bot(S)$ starting from one of these vertices. The label of this path is $w\bar{t_i}$ where w is a (possibly empty) path in G of length at most L, and t_i is one of the edges from the mapping torus $M(f)$. (We are about to use a finiteness argument and it will be important that the repetition we infer includes the labels of the points on $\bot(S)$. Thus it is important which of the t-edges this path includes.)

Since the data we record — the label of the vertex on $\top(S)$, the path $w\bar{t_i}$ and the label of the end of this path on $\bot(S)$) — run over a finite set, there is a constant C' such that in the interval within C' vertices of the left end of I there will be repetition of these data. Since the vertices occur at least every J edges, this repetition occurs within $C'J$ of the left end of I.

Once we have found this repetition, we have an interval $\lambda \subset \bot(S)$, an interval $\eta \subset \top(S)$ and a path w_0 of length at most L such that $f_\#(\lambda) = w_0\eta\bar{w_0}$. Therefore, the free homotopy class of $f_\#(\lambda)$ is the same as that of $\eta = f_\#(\eta)$, since η is a beaded path all of whose beads are Nielsen paths. Since f is a homotopy equivalence, the free homotopy class of λ must be the same as that of η.

Suppose that $\eta = p_1 \dots p_m$ where each p_i is an indivisible Nielsen path. Now, λ is tight, so $\lambda = \sigma p_i p_{i+1} \dots p_m p_1 \dots p_{i-1}\bar{\sigma}$, for some path σ. Thus, if '$\sim$' denotes free homotopy,

$$f(\lambda) \sim f_\#(\sigma) p_i \dots p_{i-1} f_\#(\bar{\sigma}),$$

which tightens to

$$w_0 p_1 \dots p_m \bar{w_0}.$$

By the Bounded Cancellation Lemma, tightening the path $f(\lambda)$ as written above reduces the length of $f_\#(\sigma)$ by less than B, and the result has length at most $2L + |\eta|$. This implies that $|f_\#(\sigma)| < L + B$. Therefore, $\|\sigma\|$ is bounded, and by a small increase we may also assume that $i = 1$. By considering only one vertex out of every $B(L + B)$, we can find such a path η where there is some p_j in the middle of λ such that the path from the copy of $p_j \subset \top(S)$ to the copy of $p_j \subset \bot(S)$ is a single edge labelled t, for some j.

We have argued that, for some path η of bounded length which lies on the left end of I, the past of η is of the form $u\eta u'$ where $|u|$ and $|u'|$ are bounded, and the paths from the splitting points in $\eta \subset I$ to $\bot(S)$ consist of single edges labelled t.

Consider the analogous situation on the right end of I. We can find a path $\eta' \subset I$ lies at the right end of I such that the past of η' is of

the form $v'\eta'v$ where $|v|$ and $|v'|$ are bounded and the paths from the vertices of $\eta' \subset I$ to $\bot(S)$ consist of single edges labelled t.

Consider the paths along $\bot(S)$ and $\top(S)$ from the left end of η to the right end of η'. We have a path $\rho \subset \bot(S)$ with fixed points of f on either end which maps to a Nielsen path $f_\#(\rho) \subset I \subset \top(S)$. The same argument as in the proof of [8, Lemma 1.14] then shows that $\rho = f_\#(\rho)$. Hence ρ is a beaded path, all of whose beads are Nielsen paths, and the paths u and v on either side of ρ are of bounded length as required. This proves the first assertion in the statement of the lemma.

The second assertion follows similarly, and the final assertion follows immediately from the first. □

Consider a pair $(\mu, \mu') \in \mathcal{Z}$, and recall the definition of the subintervals $\mathcal{E}_i$ from Definition 9.7.

Proposition 9.10. *There is a constant C_5, depending only on f so that the following holds. For each $(\mu, \mu') \in \mathcal{Z}$, the interval $C_{(\mu,\mu')} \smallsetminus \mathcal{E}_1$ in $A_4(S_0, \mu)$ has the form uNv where u and v are such that $\|u\|_\beta, \|v\|_\beta \leq C_5$ and N is a beaded path all of whose beads are Nielsen beads.*

Proof. By Lemma 8.14, for each adjacency of colours (μ, μ') there can only be one bead in $\mu(S)$ which is eventually HNP-bitten by μ'.

The result now follows from Proposition 9.9 and the definition of $\mathcal{E}_1$. □

Definition 9.11. *For $(\mu, \mu') \in \mathcal{Z}$, define $C_{(\mu,\mu')}(2) := N$, the beaded Nielsen path from Proposition 9.10.*

The sum of our arguments to this point has reduced the task of bounding the sum of the bead norms of the intervals $\mu(S_0)$ in S_0 to that of bounding the sum of the lengths of the intervals $C_{(\mu,\mu')}(2)$ for pairs $(\mu, \mu') \in \mathcal{Z}$.

We summarise the results from this section as follows.

Proposition 9.12. *There is a constant C_1, depending only on f, so that*

$$\|C_{(\mu,\mu')}\|_\beta \leq \|C_{(\mu,\mu')}(2)\|_\beta + C_1.$$

Remark 9.13. *Since the intervals $C_{(\mu,\mu')}(2)$ consist entirely of Nielsen beads, we have the following obvious relationship between length and bead norm:*

$$|C_{(\mu,\mu')}(2)| \leq \|C_{(\mu,\mu')}(2)\|_\beta \leq J|C_{(\mu,\mu')}(2)|.$$

Therefore, in order to finish the bound on bead norm, it is sufficient to bound the total lengths of the intervals $C_{(\mu,\mu')}(2)$.

It is important for the remainder of the paper that the path $C_{(\mu,\mu')}(2)$ is a beaded path that consists entirely of Nielsen atoms. This is a stronger statement than just asserting it is a Nielsen path, since we require a decomposition into beads of uniformly bounded size, each of which is a Nielsen path. This makes the path $C_{(\mu,\mu')}(2)$ very similar to the long blocks of constant letters which played such a prominent role in [7]

At this point the reader may benefit from consulting [7, Section 7], which outlines the strategy for the remainder of the proof of Theorem A (the strategy from the positive case still holds here). For the remainder of this paper, we will mostly continue without reminding the reader of this strategy.

10. The Pleasingly Rapid Disappearance of Colours

We are now at the point in our arguments where we need to formulate and prove the Pincer Lemma, as in [7, Section 8]. In [7] the Pincer Lemma was proved by counting colours which *essentially vanished*, which is to say they came to consist entirely of constant letters. For positive automorphisms, this is a well-defined event and can only occur once for each colour. For general automorphisms, the analogues of constant letters are indivisible Nielsen paths. However, since Nielsen paths can contain non-constant edges, indivisible Nielsen paths are not indivisible in an absolute sense (the terminology refers to the fact that an indivisible Nielsen path cannot be split into two Nielsen paths). Thus, it is possible that a colour can be labelled by a Nielsen path at some time t but not at some later time $t + k$. There are two ways to circumvent this problem. The first is to concentrate on the times when a colour decreases in weight, whilst the second is to focus on the times when a colour becomes Nielsen and seek compensation when a colour subsequently ceases to be Nielsen. We mostly pursue the second idea but there are aspects of the first also.

The version of the Pincer Lemma which we need in this paper is Theorem 10.27.

The ideas in the proof of the Pincer Lemma here are very similar to those in [7] but the execution is somewhat different.

Definition 10.1. *Suppose that I is a non-empty beaded Nielsen path and that U and V are beaded paths. We say that I is* stably Nielsen *in the path UIV if the future*[15] *of I in $f_\#(UIV)$ is also a non-empty Nielsen beaded path.*

[15] as defined in (3.2)

Suppose that μ_1, μ_2 and μ_3 are colours in a well-folded diagram and that the intervals $\mu_1(S), \mu_2(S)$ and $\mu_3(S)$ are non-empty and adjacent in $\bot(S)$. If $\mu_2\check{(}S)$ is a non-empty Nielsen path, then we say that $\mu_2(S)$ is stably Nielsen *if, in the above sense, $\mu_2\check{(}S)$ is stably Nielsen in $\mu_1\check{(}S)\mu_2\check{(}S)\mu_3\check{(}S))$.*

Lemma 10.2 (Relative Buffer Lemma)**.** *Let $i \in \{1, \dots, \omega - 1\}$ and let $I \subset \bot(S)$ be an edge-path labelled by edges in G_i. Suppose that the colours $\mu_1(S)$ and $\mu_2(S)$ lie either side of I, adjacent to it. Provided that the whole of I does not die in S, no edge in the future of $\mu_1(S)$ with label in $G \setminus G_i$ will ever cancel with an edge in the future of $\mu_2(S)$ with label in $G \setminus G_i$.*

Proof. Given Lemmas 5.5 and 5.6, the proof of [7, Lemma 8.1] applies modulo changes of terminology. □

We now need the following 'two-sided' version of Proposition 8.19.

Lemma 10.3. *Let μ_1, μ_2, μ_3 and S be as in Definition 10.1, and suppose that $\mu_2(S)$ is stably Nielsen. Then for all corridors S' in the future of S, if $\mu_1(S')$ and $\mu_3(S')$ are nonempty then $\mu_2(S')$ is a (possibly empty) Nielsen path.*

Proof. Whilst $\mu_1(S')$ and $\mu_3(S')$ are non-empty, any bead in μ_2 which is bitten must be bitten by a bead coloured either μ_1 or μ_3. Let I_1 be the set of (Nielsen) beads in $\mu_2(S)$ which are eventually bitten by a bead coloured μ_1 (and are bitten whilst $\mu_1(S')$ and $\mu_3(S')$ are non-empty). Define I_2 to be those beads in $\mu_2(S)$ which are bitten by a bead coloured μ_3 (with the same proviso).

Suppose that I_1 and I_2 are non-empty. They form intervals, and I_1 is to the left of I_2.

Proposition 8.21, and the fact that $\mu_2(S)$ is stably Nielsen, implies that unless I_1 is immediately consumed there is a left reaper coloured μ_1 associated to I_1, and similarly there is a right reaper coloured μ_3 associated to I_2. The properties of reapers in Definition 8.17 imply the result.

In case one or both of I_1 and I_2 are empty (or immediately consumed), there is at most one reaper to consider, but the result follows in the same way. □

Lemma 10.4 (Buffer Lemma)**.** *Suppose, for some corridor S in a well-folded diagram, that $I \subset \bot(S)$ is a beaded Nielsen path and that $\mu_1(S)$ and $\mu_2(S)$ lie either side of I, immediately adjacent to it. Suppose further that $\check{I}$ is stably Nielsen in $\mu_1\check{(}S)\check{I}\mu_2\check{(}S)$. Provided that the whole*

of I does not die in S, no bead in $\mu_1(S)$ can be eventually bitten by a bead coloured μ_2 (and vice versa), unless it is (eventually) HNP-bitten.

Proof. Given Lemmas 5.5, 5.6 and 10.3, and the exclusion of HNP-biting, the proof of [7, Lemma 8.1] applies. □

The proof of the following lemma follows that of [7, Lemma 8.1].

Lemma 10.5 (Weighted Buffer Lemma)**.** *Suppose, for some corridor S in a well-folded diagram, that $I \subset \perp(S)$ is a beaded path consisting of Nielsen beads and beads of weight at most i, and that $\mu_1(S)$ and $\mu_2(S)$ lie either side of I, immediately adjacent to it. Suppose further that the only beads of $f_\#(\mu_1\check{(}S)\check{I}\mu_2\check{(}S))$ that are in the future of I and have weight greater than i are Nielsen beads.*

Then, provided that the whole of I does not die in S, no bead in $\mu_1(S)$ can be eventually bitten by a bead coloured μ_2 (and vice versa), unless it is (eventually) HNP-bitten.

10.1. **The Two Colour Lemma.** Example 8.6 can be used to construct examples where the above two results are false if HNP-biting is not excluded. The same is true of the results in this section. This accounts for the caution that the reader will note in Sections 11, 12 and 13, where we are careful to ensure that the Pincer Lemma is applied only to pincers that involve no HNP-biting.

Definition 10.6 (Stable f-neutering)**.** *Suppose that U and V are beaded paths, that for some k the futures of V in $f_\#^k(UV)$ and $f_\#^{k+1}(UV)$ are Nielsen, but that the future of V in $f_\#^{k-1}(UV)$ contains a non-Nielsen bead.*

Denote the futures of U and V in $f_\#^{k-1}(UV)$ by U^{k-1} and V^{k-1}, respectively. Let β be the rightmost non-Nielsen bead in $f_\#(V^{k-1})$. If the biting of β in the tightening of $f_\#(U^{k-1})f_\#(V^{k-1})$ to form $f_\#^k(UV)$ is not HNP-biting then we say that U stably left f-neuters *V in k steps.*

The definition of stable right f-neutering *is identical with the roles of U and V reversed, and when we are unconcerned about the direction we will refer simply to* stable f-neutering.

In the light of Proposition 8.19, once stably f-neutered, the subsequent futures of V remain beaded Nielsen paths.

Proposition 10.7 (Two Colour Lemma, cf. Proposition 8.4 [7])**.** *There exists a constant T_0, depending only on f, so that if U and V are beaded paths and U stably f-neuters V then it does so in at most T_0 steps.*

Proof. Denote the future of U in $f^i_\#(UV)$ by U^i and the future of V by V^i.

As in the proof of [7, Proposition 8.4], we will decompose each of the paths V^i into an *unbounded part* and a *bounded part*. The bounded part will be an interval on the right end of V^i whose immediate (abstract) future is a beaded Nielsen path. The unbounded interval lies on the left end of V^i, and we will bound its length.

This would be a straightforward adaptation of the proof from [7] if Proposition 9.12 provided a bound of the length of that part of $C_{(\mu,\mu')}$ not contained in $C_{(\mu,\mu')}(2)$. However, the bound in Proposition 9.12 is just a bound on bead norm. Thus, we need to deal with the possibility of long GEPs and ΨEPs.

The following enumerated claims will together yield an upper bound on the length of the unbounded part of V^i, which in the course of the proof will be decomposed into V^i_{fast} and V^i_{nc}

Three of the claims concern the existence of a constant k_j that depends only on f; we use the abbreviation $\exists k_j = k_j(f)$.

Claim 1: $\exists k_1 = k_1(f)$ such that any GEP in V^i has length less than k_1.

This follows in a straightforward way from the Buffer Lemma 10.4 and the fact that the obvious preferred future of the rightmost edge in any GEP in V^i must eventually cancel with an edge from the future of U^i.

Next we consider long ΨEPs in V^i. Suppose that ρ is a ΨEP in V^i. Then the label on ρ or $\bar\rho$ has the form $E\bar\tau^k\bar\nu\gamma$, where τ is Nielsen path, $f(E) = E \odot \tau^m$ and $\bar\gamma\nu$ is a terminal segment of τ. We consider a number of different cases. First we dismiss a case that follows immediately from Lemma 6.13 and from the fact that exponential edges are left-fast:

Claim 2: If $\check\rho = E\bar\tau^k\bar\nu\gamma$ and γ is an exponential edge then the right end of ρ lies within C_0 of the left end of V^i.

Next we consider V^i_{fast}, which is defined to consist of those beads from the left end of V^i up to and including the rightmost bead in V^i whose immediate (abstract) future contains a left-fast bead.

Claim 3: $\exists k_2 = k_2(f)$ such that $|V^i_{\text{fast}}| \le k_2$.

This follows immediately from Lemma 6.13 unless the rightmost bead in V^i_{fast} is a ΨEP. (Note that this rightmost bead is not a GEP, since a GEP does not have a left-fast bead in its immediate abstract future.)

Suppose, then, that the rightmost bead in V^i_{fast} is a ΨEP, say ρ. If $\check\rho = E\bar\tau^k\bar\nu\gamma$, then we are done by Claim 2. So suppose that $\check\rho = \bar\gamma\nu\tau^k\bar E$.

Let ε be the edge in ρ whose label is $\bar{E}$. The preferred future of ε is to be cancelled by an edge in the future of U^i. By an obvious finiteness argument (as in the proof of [7, Proposition 8.4]), there is a constant p so that the path V^p contains no left-fast beads. This gives a bound on the amount of time before the future of ρ is bitten, and hence a bound on the amount that the future of ρ can shrink before then. Suppose that V^j is the first future of V^i in which the future of ρ has been bitten. Because the preferred future of ε is to be cancelled, Proposition 8.21 and the Buffer Lemma 10.4 imply that the length of the future in V^j of ρ is bounded above by a constant depending only on f.

The required bound on $|V^i_{\text{fast}}|$ is now at hand: Lemma 6.13 bounds the length of $V^i_{\text{fast}} \smallsetminus \rho$, and the combination of the bound on j and the bound on the length of the future of ρ in V^j gives a bound on the length of ρ. This completes the proof of Claim 3. We remark that the above argument also gives a bound on the amount of time it takes for V^1_{fast} to be entirely consumed.

We now define a set V^i_{nc} as follows: Let ρ_{nc} be the rightmost bead in V^i whose immediate abstract future is not Nielsen. We define V^i_{nc} as follows:

(1) if $\rho_{\text{nc}} \in V^i_{\text{fast}}$ then $V^i_{\text{nc}} = \emptyset$;
(2) if ρ_{nc} is not a ΨEP, then V^i_{nc} consists of those beads from (but not including) the rightmost bead in V^i_{fast} up to and including ρ_{nc};
(3) if ρ_{nc} is a ΨEP with label of the form $\bar{\gamma}\nu\tau^k\bar{E}$ or ρ_{nc} is a ΨEP with label of the form $E\bar{\tau}^k\bar{\nu}\gamma$ and γ is not a Nielsen path, then V^i_{nc} consists of those beads in V^i from (but not including) the rightmost bead in V^i_{fast} up to and including ρ_{nc};
(4) finally, if ρ_{nc} is a ΨEP with label of the form $E\bar{\tau}^k\bar{n}\bar{u}\gamma$ and γ is either empty or a Nielsen path, then V^i_{nc} consists of that interval from (but not including) the rightmost bead in V^i_{fast} up to and including the leftmost edge in ρ_{nc} (the label of this leftmost edge is E).

Note that in Case 4 the bead ρ_{nc} is certainly not contained in V^i_{fast}.

Claim 4: $\exists k_3 = k_3(f)$ such that $|V^i_{\text{nc}}| \leq k_3$.

The proof of Claim 3 above established an upper bound on the time before all of V^i_{fast} is entirely consumed, and hence also on the time before the future of V^i_{nc} begins to be consumed. We now follow the proof of Lemma 9.8, which establishes an upper bound on the time that can elapse before the final non-constant bead in V^i is bitten. We will be done if we can bound this time from below by a positive constant times $|V^i_{\text{nc}}|$.

In the current setting, we have non-constant beads in V^i_{nc} that may not be growing apart like those in the proof of Lemma 9.8.[16] But there *is* a lower bound on the rate at which the surviving futures of these beads can come together. Hence the length of V^i_{nc} provides a lower bound on the amount of time that must elapse before V^j becomes stably Nielsen, since the future of V^i_{nc} must be entirely consumed before this time. (Note that in Case 4, the preferred future of the edge $\bar{E}$ in ρ_{nc} must be eventually consumed by the future of U^i.) This proves Claim 4.

The *unbounded part* of V^i is the union of V^i_{fast} and V^i_{nc}, whilst the *bounded part* is the remainder of V^i. The sum of the previous four claims bound the length of the unbounded part of V^i by a constant that depends only on f.

There is a similar bound on the number of edges in U^i that have an edge in their future that cancels with an edge in the future of V^i. (Here we need the hypothesis that the path V^k becoming stably Nielsen does not arise from HNP-biting.)

At this stage, we can follow the proof of [7, Proposition 8.4] directly. After an amount of time bounded by a constant that depends only on f, either the future of V becomes stably Nielsen or empty, or else there is a repetition of the following data: (i) the unbounded part of V^i plus the leftmost $B + J$ edges of the bounded part; (ii) a terminal segment of U^i containing all of the edges that can ever interact with the future of V. Once we have such a repetition, if the future of V has not become stably Nielsen or vanished then it never will, contrary to hypothesis. □

We need a weighted version of neutering and the two-colour lemma.

Definition 10.8 ((f,i)-neutering)**.** *Fix $i \in \{1,\dots,\omega\}$ and let U and V be beaded paths. Suppose that for some k the future of V in $f^k_\#(UV)$ has weight less than i, but that the future of V in $f^{k-1}_\#(UV)$ has weight at least i.*

*Denote the futures of U and V in $f^{k-1}_\#(UV)$ by U_{k-1} and V_{k-1}, respectively. Let β be the rightmost bead in $f_\#(V_{k-1})$ of weight at least i. If the biting of β in the tightening of $f_\#(U_{k-1})f_\#(V_{k-1})$ to form $f^k_\#(UV)$ is not HNP-biting then we say that U (f,i)-*neuters *V in at most k steps.*

Proposition 10.9 (Weighted Two Colour Lemma)**.** *There exists a constant T_0', depending only on f, so that for any $i \in \{1,\dots,\omega\}$, if U*

[16] This is because we are now measuring length rather than bead-norm.

and V are beaded paths and U (f,i)-neuters V then it does so in at most T_0' steps.

Proof. We decompose the futures of U and V in $f_\#^k(UV)$ as in Lemma 10.7.

The proof is similar to that of Lemma 10.7, except that when we appeal to the proof of Proposition 9.8 we assume that we have a path $\mathcal{E}_j$ with $j \geq i$. Otherwise, the proof of Lemma 10.7 above and that of [7, Proposition 8.4] can now be followed *mutatis mutandis*. □

By replacing T_0 by T_0' if necessary, we may assume that $T_0 \geq T_0'$. We henceforth make this assumption.

10.2. **The disappearance of colours: Pincers and implosions.**

Definition 10.10. *Consider a pair of non-constant edges ε_1 and ε_2 which cancel in a corridor S_t of Δ, and suppose that, for $i = 1, 2$, the immediate past of ε_i lies in a bead of some $\mu_i(S_t)$ that is either a unbounded atom, a* GEP *or a* ΨEP*. Suppose further that the cancellation of ε_1 and ε_2 is not HNP-cancellation, and that $\mu_1 \neq \mu_2$. Consider the paths p_1, p_2 in $\mathcal{F} \subset \Delta$ tracing the histories of ε_1 and ε_2. Suppose that at time τ_0 the paths p_1 and p_2 lie in a common corridor S_b. Under these circumstances, we define the* pincer $\Pi = \Pi(p_1, p_2, \tau_0)$ *to be the sub-diagram of Δ enclosed by the chains of 2-cells along p_1 and p_2, and the chain of 2-cells connecting them in S_b.*

We define S_Π to be the earliest corridor of the pincer in which $\mu_1(S_\Pi)$ and $\mu_2(S_\Pi)$ are adjacent. Define $\widetilde{\chi}(\Pi)$ to be the set of colours $\mu \notin \{\mu_1, \mu_2\}$ such that there is a 2-cell in Π coloured μ. Finally, define

$$\mathrm{Life}(\Pi) = \mathrm{time}(S_\Pi) - \mathrm{time}(S_b).$$

See [7, Section 8] for illustrative pictures.

Proposition 10.11 (Unnested Pincer Lemma, cf. Proposition 8.7 [7])**.**

There exists a constant T_1, depending only on f, such that for any pincer Π

$$\mathrm{Life}(\Pi) \leq T_1(1 + |\widetilde{\chi}(\Pi)|).$$

In the proof of Proposition 8.7 (Regular Implosions) in [7], the strategy was to identify a constant T_1 such that over each period of time of length T_1 within a pincer, at least one colour became constant. There are a number of impediments to implementing this strategy in the current situation. The first is that Nielsen paths can consist of edges which are not constant edges, so if a colour *becomes Nielsen* then it may cease to be Nielsen at some stage in the future. In order to overcome this impediment, we make the following

Definition 10.12. *Suppose that for some colour μ and some corridor S, the path $\mu(\check{S})$ is stably Nielsen, and let ν_1 and ν_2 be the colours immediately on either side of μ in S. If there is some corridor S' in the future of S in which $\mu(\check{S}')$ is not Nielsen and S' is the earliest such corridor, then we say that μ is* resuscitated *in S'. By Lemma 10.3, at least one of ν_1 and ν_2 is not adjacent to μ in S', so either $\nu_1(S')$ or $\nu_2(S')$ is empty. If $\nu_i(S')$ is empty, we say that ν_i* sacrifices itself *for μ.*

Remark 10.13. *A colour can sacrifice itself for at most one colour.*

A colour may become stably Nielsen and be resuscitated a number of times, but a different colour must sacrifice itself for each resuscitation.

The concept of 'becoming stably Nielsen' is analogous to that of a colour 'essentially vanishing' in [7, Section 8]*. However, the concept of 'resuscitation' does not have an analogue in* [7]*.*

Fix a pincer Π and assume that $\mathrm{Life}(\Pi) > 1$. The strategy to prove Proposition 10.11 is to identify a constant T_1 so that during the life of Π, in each $T_1/2$ steps of time there is a colour that becomes stably Nielsen (perhaps vanishing) In order to obtain the bound in the statement of Proposition 10.11, we then count the colours which become stably Nielsen or vanish, and the colours which sacrifice themselves for those that are resuscitated. A colour can therefore be counted twice – once for disappearing (or for the last time it becomes stably Nielsen), and once as a sacrifice – but no colour is counted more than twice. Thus Proposition 10.11 is an immediate consequence of the following result whose proof will occupy the remainder of this subsection.

Proposition 10.14. *There is a constant T_1, depending only on f, so that for any pincer Π in a minimal area van Kampen diagram over $M(f)$, in any interval of time of length $T_1/2$, at least one colour in $\tilde{\chi}(\Pi)$ becomes stably Nielsen or vanishes.*

Definition 10.15 (p-implosive arrays)**.** *Let p be a positive integer and S a corridor. A* p-implosive array *of colours in S is an ordered tuple $A(S) = [\nu_0(S), \dots, \nu_r(S)]$, with $r > 1$, such that*

1. *each pair of colours $\{\nu_j, \nu_{j+1}\}$ is separated in S only by a stably Nielsen (or empty) path;*
2. *in each of the corridors $S = S^1, S^2, \dots, S^p$ in the future of S, no $\nu_j(S^i)$ is empty or a stably Nielsen path, $j = 1, \dots, r-1$;*
3. *in S^p,* either *an edge coloured ν_0 from a unbounded atom, a* GEP *or a* ΨEP *cancels with an edge coloured ν_r from a unbounded atom, a* GEP *or a* ΨEP *(and hence the colours ν_j with $j = 1, \dots, r-1$ are consumed entirely), or* else *each of the colours ν_j ($j = 1, \dots, r-1$) become stably Nielsen or vanish, while ν_0*

and ν_r are not Nielsen in $f_\#(\nu_0(\check{S}^p)\cdots\nu_r(\check{S}^p))$ (although they may nevertheless become stably Nielsen or even disappear in S^p because of colours external to the array).

Arrays satisfying the first of the conditions in (3) are said to be of *Type I*, and those satisfying the second condition are said to be of *Type II*. (These types are not mutually exclusive).

The *residual block* of an array of Type II is the stably Nielsen path which lies between $\nu_0(S^p)$ and $\nu_r(S^p)$ (if either $\nu_0(S^p)$ begins or $\nu_r(S^p)$ ends with an interval of Nielsen atoms include these in the residual block). Note that the residual block may be empty. The *enduring block* of the array is the set of stably Nielsen paths in $\perp(S)$ that have a future in the residual block.

Note that there may exist some *unnamed colours* between $\nu_j(S)$ and $\nu_{j+1}(S)$; if they exist, these form a stably Nielsen path.

Remark 10.16. *Let $[\nu_0(S),\dots,\nu_r(S)]$ be a p-implosive array.*

1. *Any q-implosive sub-array of $[\nu_0(S),\dots,\nu_r(S)]$ has $q=p$.*
2. *If an edge of ν_i cancels with an edge of ν_j and $j-i>1$, then this cancellation can only take place in S^p. If the edges cancelling come from displayed unbounded atoms,* GEP*s or* ΨEP*s, then the sub-array $[\nu_i(S),\dots,\nu_j(S)]$ is p-implosive of Type I.*
3. *If u,v and w are beaded edge-paths such that u, v and $f_\#(uwv)$ are Nielsen paths then w is a Nielsen path. It follows that the residual block of any array of Type II contains edges from at most two of the colours ν_j, and if there are two colours then they are consecutive, ν_j, ν_{j+1}.*
4. *Likewise, the enduring block of an implosive array of Type II is an interval involving at most two of the ν_j and if there are two such colours they must be consecutive.*

Lemma 10.17. *Let Π be a pincer. The ordered list of colours along each corridor before* $\mathrm{time}(S_\Pi)$ *in a pincer Π must contain a p-implosive array for some p.*

Proof. The definition of p-implosive array is designed so that when a colour becomes stably Nielsen (or disappears) in a pincer there is a p-implosive array. See the proof of [7, Lemma 8.10] for more details. □

Definition 10.18. *Suppose that $A(S)=[\nu_0(S),\dots,\nu_r(S)]$ is a p-implosive array. We say that $A(S)$ is an* HNP-implosive array *if either*

1. *$A(S)$ is of Type I and in S^p the cancellation between ν_0 and ν_r is HNP-biting, or*

(2) *$A(S)$ is of Type II and in S^p, for some $0 < i < r$, ν_0 and ν_i are involved in HNP-biting or for some $0 < j < r$, ν_j and ν_r are involved in HNP-biting.*

In order to follow the arguments from [7], we need to sharpen Lemma 10.17: HNP-cancellation can beget p-implosive arrays with p arbitrarily large, and therefore we must argue for the frequent occurrence of p-implosive arrays that are not HNP-implosive. A first step in this direction is given by the following

Lemma 10.19. *Let Π be a pincer, and let μ_1 and μ_2 be the colours associated to the bounding-paths p_1 and p_2 of Π. Then there is no HNP-biting between beads in μ_1 and μ_2 within Π.*

Proof. Follows from Lemmas 8.14 and 8.22. □

When we are unconcerned about p in a p-implosive array, we refer merely to an *implosive array*. The first restriction to note concerning implosive arrays is this:

Lemma 10.20. *If $[\nu_0(S), \dots, \nu_r(S)]$ is implosive of Type I, then $r \leq B$. If it is implosive of Type II, then $r \leq 2B$.*

Proof. In Type I arrays, the interval $\nu_1(S^p) \cdots \nu_{r-1}(S^p) \subset \perp(S^p)$ is to die in S^p, so the bound is an immediate consequence of the Bounded Cancellation Lemma. For Type II arrays, one applies the same argument to the intervals joining $\nu_0(S^p)$ and $\nu_r(S^p)$ to the residual block. □

Proof of Proposition 10.14. We give a suitable formulation of 'short' so that in any corridor S within Π, S contains a short p-implosive array. Proposition 10.14 then follows from an obvious finiteness argument.

Let $A(S) = [\nu_0(S), \dots, \nu_r(S)]$ be the implosive array guaranteed to exist by Lemma 10.17, and suppose that $p \geq 2T_0$ (if not then a colour becomes stably Nielsen or vanishes within $2T_0$ of time(S)).

We can decompose each of the colours $\nu_j(S)$ in analogy with [7], using the decomposition in Section 7.3 above.

We fix a constant Λ_1 so that if $\|A(S)\| > \Lambda_1$ then one of the following must occur in S^{T_0}:

(1) there is a block of displayed Nielsen atoms in some $\nu_j(S^{T_0})$ of length at least $J + 4B$,
(2) there is a displayed GEP in some $\nu_j(S^{T_0})$ of length at least $J + 4B + 2$,
(3) there is a displayed ΨEP in some $\nu_j(S^{T_0})$ of length at least $J + 4B + L + 1$, or

(4) there is an interval of unnamed colours in $A(S)$ (which form a stably Nielsen block) of length at least $J + 4B$ between $\nu_0(S^{T_0})$ and $\nu_r(S^{T_0})$.

In the remainder of the proof, we shall use the term *block* to refer generically to the identified interval in whichever of the above cases we find ourselves. Increasing Λ_1 if necessary, we may assume that the past of the block in S satisfies the relevant condition from (1) – (4) with the bound increased by $2BT_0$.

For such a block I in S^{T_0}, consider the first edge on either side of this block which is not contained in a Nielsen path. These edges may be on one end of a GEP or a ΨEP (including the GEP or ΨEP from condition (2) or (3)), or may be contained in unbounded atoms. Call these edges ε_1 and ε_2.

The Buffer Lemma 10.4 implies that either (i) one of ε_1 and ε_2 must be 'stabbed in the back' – we do not exclude the possibility that this stabbing happens by HNP-biting, or (ii) there is HNP-cancellation across the above block.

We first dispose of case (ii). Suppose, for ease of notation, that the edge ε_1 HNP-bites the edge ε_2 across the above block I. Let ε_1 have weight k. Then all edges in I and ε_2 must have weight less than k. Let ε_2' be the first edge to the right of I that has weight at least k. Then the Relative Buffer Lemma 10.2 implies that either ε_1 or ε_2' must be stabbed[17] in the back (again, this could be by HNP-biting).

We have argued that some edge must be stabbed in the back. Suppose that this stabbing is of an edge ε in S^{T_0} and that ε has weight k_1. Consider first the possibility that ε is stabbed in the back via HNP-biting. Then this occurs by an edge ε' of weight at least $k_1 + 1$. Now, either this stabbing in the back occurs within T_0 of S^{T_0}, or by the Weighted Two Colour Lemma (10.9) there is another block as in (1) – (4) above. This block has higher weight than the previous block, and as above leads to another stabbing in the back. If this stabbing is HNP-biting, pass to a yet higher weight stabbing, and so on.

Eventually (after less than ω iterations of this argument), we get an edge ε stabbed in the back with the stabbing not HNP-biting. Suppose that ε has weight k_2. Suppose for ease of notation that ε is to the left of the long block, and suppose that ε is coloured ν_i. Because of the block of Nielsen atoms to the non-stabbing side of ε, the Two Colour Lemma (Proposition 10.7) implies that if the edge ε' which stabs ε in the back is coloured by ν_j then $i - j > 1$; we then write $\nu_j \searrow \nu_i$.

[17]Note that if there is no such edge ε_2' in Π then ε_1 must be stabbed in the back, by Lemmas 10.2 and 10.19.

Passing to an innermost pair $\nu_{l_1} \searrow \nu_{l_2}$ between ν_i and ν_j we can see that there are no blocks in S^{T_0} satisfying any of (1) – (4) above, for otherwise there would be a further stabbing, leading to a related pair of colours between our innermost pair, contradicting the innermost nature of this pair.

Once there are no such blocks, we have a bound on the length of the p-implosive array implicit in the relation $\nu_{l_1} \searrow \nu_{l_2}$. An obvious finiteness argument now finishes the proof. □

We have already seen how Proposition 10.14 implies Proposition 10.11. Just as in [7, Section 8], we must now deal with the possibility of 'nested pincers'.

10.3. **Super-buffers.**

Definition 10.21. *We consider sequences of* 5*-tuples of tight edge-paths in* G*.*

$$U_k := \Big(u_{k,1}, u_{k,2}, u_{k,3}, u_{k,4}, u_{k,5}\Big), \quad k = 1, 2, ...$$

with $|u_{k,1}|$ *and* $|u_{k,2}|$ *at most* $C_0 + C_1 + 2B(B+1) + 1$*, while* $|u_{k,2}|$ *and* $|u_{k,4}|$ *are at most* $C_0 + C_1 + J$ *and* $|u_{k,3}| \leq 4B(B+1) + 1$.[18] *We fix an integer* T_1' *sufficiently large to ensure that for any sequence of length* T_1' *there will be a repetition, i.e. some* $t_1 < t_2 \leq T_1'$ *with*

$$\Big(u_{t_1,1}, u_{t_1,2}, u_{t_1,3}, u_{t_1,4}, u_{t_1,5}\Big) = \Big(u_{t_2,1}, u_{t_2,2}, u_{t_2,3}, u_{t_2,4}, u_{t_2,5}\Big).$$

We also choose $T_1' \geq T_1$*.*

With appropriate changes of terminology and the results of the previous subsection in hand, the proof of [7, Proposition 8.21] yields:

Lemma 10.22. *Let* $V = V_1V_2V_3$ *be a tight concatenation of three beaded paths in* G*. If the future of* V_2 *is not stably Nielsen in* $f_\#^{T_1'}(V)$ *then the future of* V_2 *is not stably Nielsen in* $f_\#^k(V)$ *for any* $k \geq 0$*.*

10.4. **Nesting and the Pincer Lemma.** Let $\lambda_0 = J + 2B(T_0+1)+1$, which is the obvious analogue of the constant of the same name in [7, Section 8]. As in [7, Remark 9.5], it is convenient to assume that $LC_4 < \lambda_0$, and we increase λ_0 to make this so. (This makes certain statements in Section 11 easier, but has no serious affect.)

[18] The purpose of these constants is just as in [7, Definition 8.19], with appropriate changes due to Lemmas 7.8 and 7.1 and Proposition 9.12.

Definition 10.23. *Consider one pincer Π_1 contained in another Π_0. Suppose that in the corridor $S \subseteq \Pi_0$ at the top of Π_1 (where its boundary paths $p_1(\Pi_1)$ and $p_2(\Pi_1)$ come together) the future in $\top(S)$ of at least one of the edges containing $p_1(\Pi_1) \cap \top(S)$ or $p_2(\Pi_1) \cap \top(S)$ is not contained in any stably Nielsen path and this future* [19] *lies in a beaded path consisting of Nielsen beads and beads of weight strictly less than the weight of the edges containing $p_1(\Pi_1) \cap \top(S)$ and $p_2(\Pi_1) \cap \top(S)$, and that this beaded path has at least λ_0 non-vanishing beads. Then we say that Π_1 is* nested *in Π_0.*

Remark 10.24. *Besides the obvious translations, the above differs from* [7, Definition 8.22] *in that the path at the top of the pincer may now consist of Nielsen beads and lower weight beads, whereas in* [7] *it consisted entirely of constant letters. This more general setting does not make any of the proofs in this section harder (because of the Weighted Two Colour Lemma), but is needed because of the more complicated definition of the 'cascade of pincers' below (Definition 11.17).*

Definition 10.25. *For a pincer Π_0, let $\{\Pi_i\}_{i \in I}$ be the set of all pincers nested in Π_0. Then define*

$$\chi(\Pi_0) = \tilde{\chi}(\Pi_0) \smallsetminus \bigcup_{i \in I} \tilde{\chi}(\Pi_i).$$

The corridor S_t was defined in Definition 10.10.

Lemma 10.26. [7, Lemma 8.25] *If the pincer Π_1 is nested in Π_0 then* $\mathrm{time}(S_t(\Pi_1)) < \mathrm{time}(S_{\Pi_0})$.

Proof. The existence of the beaded path at the top of the pincer Π_1 makes this an immediate consequence of the Weighted Buffer Lemma 10.5. □

Define $T_1 = T_1' + 2T_0$. The following theorem is the main result of this section, and is the strict analogue of [7, Theorem 8.26]. The proof in the current context follows the proof from [7] *mutatis mutandis*.

Theorem 10.27 (Pincer Lemma)**.** *For any pincer Π*

$$\mathrm{Life}(\Pi) \leq T_1(1 + |\chi(\Pi)|).$$

[19] We allow this future to be empty, in which case "contained in" means that the immediate past of the long stably Nielsen path is not separated from Π_1 by any edge that has a future in $\top(S)$.

11. Teams

By virtue of Lemma 9.12, Remark 9.13 and the results of Section 7, we have reduced the task of bounding the bead norm of S_0 to that of bounding the lengths of certain blocks $C_{(\mu,\mu')}(2)$ which consist of Nielsen beads coloured μ all of which are to be eventually bitten by beads coloured μ' in the future of S_0. By Proposition 8.21, if such a block has length at least $B+J$, then there is an associated reaper, which consumes Nielsen beads in $C_{(\mu,\mu')}(2)$ at a constant rate (and entirely consumes any bead it bites, up to the final bead). Note that to each pair (μ,μ') there is at most one associated reaper.

This puts us in the situation where we can develop the technology of *teams* as in [7, Section 9]. However, there are a number of key differences to [7]: we already had to work hard in Section 8 to establish the existence of a *reaper* for $C_{(\mu,\mu')}(2)$, and now we have to work harder to identify the times $\hat{t}_1(\mu,\mu')$ and $t_1(\mathcal{T})$ attached to a pair $(\mu,\mu')\in\mathcal{Z}$ and a team $\mathcal{T}$, using the *robust* past of the reaper instead of the actual past; this is required in order that the Pincer Lemma apply to teams of genesis (G3). It is worth remarking that once we have identified the pincer $\Pi_{\mathcal{T}}$ associated to a team $\mathcal{T}$ of genesis (G3), we revert to an analysis of actual pasts (as in the definition of pincer).

Note that the colour of the edges in the robust future of an edge may not always be the same, contrary to the actual future. In fact, whenever the robust past is not the actual past, the colour changes. This explains a slight difference between Definition 11.3 below and [7, Definition 9.1].

Consider an interval $C_{(\mu,\mu')}(2)$ so that $|C_{(\mu,\mu')}(2)|>B+J$, and let ϵ^μ be the reaper associated to $C_{(\mu,\mu')}(2)$ in Proposition 8.21 above. Let t_0 be the time at which ϵ^μ first bites a Nielsen bead in $C_{(\mu,\mu')}(2)$, and let β_μ be the rightmost bead in the future of $C_{(\mu,\mu')}(2)$ at this time. Note that β_μ is a Nielsen bead. Let ϵ_μ be the rightmost edge in β_μ.

Remark 11.1. *Since $|C_{(\mu,\mu')}(2)|>B+J$, and each bead of $C_{(\mu,\mu')}(2)$ is to be bitten by μ', the colour of ϵ^μ is μ'.*

Lemma 11.2. *Suppose that the immediate past of ϵ_μ exists (i.e. that ϵ_μ does not lie on $\partial\Delta$). Then the immediate past of ϵ_μ lies in some bead σ, and σ contains the immediate past of each edge in β_μ.*

The above lemma, applied at each stage in the past, implies that we can follow the past of the edge ϵ_μ and deduce consequences about the past of all edges in β_μ.

We now define a time $\hat{t}_1(\mu,\mu')$ as follows: We go back to the last point in time when (i) the past of ϵ_μ and the robust past of ϵ^μ lay in

a common corridor; and (ii) ϵ_μ is contained in a beaded Nielsen path whose swollen present is immediately adjacent to the robust past of ϵ^μ.

We denote this corridor $S_\uparrow$.

Definition 11.3. *The robust past of ϵ^μ at time $\hat{t}_1(\mu,\mu')$ is called the* reaper*, and is denoted $\hat{\rho}(\mu,\mu')$. The interval $\hat{\mathfrak{T}}(\mu,\mu')$ is the maximal beaded Nielsen path in $\perp(S_\uparrow)$ all of whose beads are eventually bitten by $\hat{\rho}(\mu,\mu')$. The* pre-team *$\hat{\mathcal{T}}(\mu,\mu')$ is defined to be the set of pairs $(\mu_1,\mu_2)\in\mathcal{Z}$ so that (i) the robust past of ϵ^μ is coloured μ_2 at some time between $\hat{t}_1(\mu,\mu')$ and t_0; and (ii) $\hat{\mathfrak{T}}(\mu,\mu')$ contains some edges coloured μ_1. The number of beads in $\hat{\mathfrak{T}}(\mu,\mu')$ is denoted $\|\hat{\mathcal{T}}\|$.*

As in [7, Section 9], we will define *teams* to be pre-teams satisfying a certain maximality condition (see Definition 11.6 below).

Remark 11.4. *Just as in* [7, Remark 9.2]*, if $\hat{t}_1(\mu,\mu')<\mathrm{time}(S_0)$ then near the right-hand end of $\hat{\mathfrak{T}}(\mu,\mu')$ one may have an interval of colours ν for which $\nu(S_0)$ is empty.*

Lemma 11.5 (cf. Lemma 9.3, [7])**.** *If $\hat{t}_1(\mu,\mu')\geq\mathrm{time}(S_0)$ then*

$$\sum_{(\mu_1,\mu_2)\in\hat{\mathcal{T}}(\mu,\mu')}|C_{(\mu,\mu')}(2)|\leq\|\hat{\mathcal{T}}(\mu,\mu')\|+B(B+1).$$

Proof. The extra $B(B+1)$ is to account for the beads consumed before the reaper comes into play. Otherwise the proof is just as in [7]. □

11.1. **The Genesis of pre-teams.** [cf. Subsection 9.2, [7]]

We consider the various events that may occur at $\hat{t}_1(\mu,\mu')$ which prevent us pushing the pre-team back one step in time. Recall that $S_\uparrow$ is the corridor at time $\hat{t}_1(\mu,\mu')$ which contains $\hat{\mathcal{T}}(\mu,\mu')$. Suppose that μ_2 is the colour of $\hat{\rho}(\mu,\mu')$.

There are four types of events:

- (G1) The immediate past of $C_{(\mu,\mu_2)}(S_\uparrow)$ is separated from the robust past of $\hat{\rho}(\mu,\mu')$ by an intrusion of $\partial\Delta$.
- (G2) We are not in Case (G1), but the immediate past of $C_{(\mu,\mu_2)}(S_\uparrow)$ is separated from the robust past of $\hat{\rho}(\mu,\mu')$ because of a singularity.
- (G3) The immediate past of $C_{(\mu,\mu_2)}(S_\uparrow)$ is still in the same corridor as the robust past of $\hat{\rho}(\mu,\mu')$, but the swollen present of the immediate past of $C_{(\mu,\mu_2)}(S_\uparrow)$ is not immediately adjacent to the robust past of $\hat{\rho}(\mu,\mu')$.
- (G4) We are not in any of the above cases, but the immediate past of the rightmost edge in $C_{(\mu,\mu_2)}(S_\uparrow)$ is not contained in a beaded Nielsen path.

We now make the definition of a team.

Definition 11.6 (cf. Definition 9.6, [7])**.** *All pre-teams $\hat{\mathcal{T}}(\mu,\mu')$ with $\hat{t}_1(\mu,\mu') \geq \mathrm{time}(S_0)$ are defined to be teams, but the qualification criteria for pre-teams with $\hat{t}_1(\mu,\mu') < \mathrm{time}(S_0)$ are more selective.*

If the genesis of $\hat{\mathcal{T}}(\mu,\mu')$ is of type (G1) or (G2), then the rightmost component of the pre-team may form a pre-team at times before $\hat{t}_1(\mu,\mu')$. In particular, it may happen that $(\mu_1,\mu_2) \in \hat{\mathcal{T}}(\mu,\mu')$ but $\hat{t}_1(\mu,\mu') > \hat{t}_1(\mu_1,\mu_2)$ and hence $(\mu,\mu') \notin \hat{\mathcal{T}}(\mu_1,\mu_2)$. To avoid double-counting in our estimates on $\|\mathcal{T}\|$ we disqualify the (intuitively smaller) pre-team $\hat{\mathcal{T}}(\mu_1,\mu_2)$ in these settings.

If the genesis of $\hat{\mathcal{T}}(\mu,\mu')$ is of type (G4), then again it may happen that what remains to the right of $\hat{\mathcal{T}}(\mu,\mu')$ at some time before $\hat{t}_1(\mu,\mu')$ is a pre-team. In this case, we disqualify the (intuitively larger) pre-team $\hat{\mathcal{T}}(\mu,\mu')$.

The pre-teams that remain after these disqualifications are now defined to be teams.

A typical team will be denoted $\mathcal{T}$ and all hats will be dropped from the notation for their associated objects (just as in [7, Section 9]).

A team is said to be *short* if $\|\mathcal{T}\| \leq \lambda_0$ or $\sum_{(\mu_1,\mu_2)\in\mathcal{T}} |C_{(\mu_1,\mu_2)}(2)| \leq \lambda_0$. Let Σ denote the set of short teams.

Lemma 11.7 (cf. Lemma 9.7, [7])**.** *Teams of genesis (G4) are short.*

We wish our ultimate definition of a team to be such that every pair (μ,μ') with $C_{(\mu,\mu')}(2)$ non-empty is assigned to a team. The above definition fails to achieve this because of two phenomena: first, a pre-team $\mathcal{T}(\mu,\mu')$ with genesis of type (G4) may have been disqualified, leaving (μ,μ') teamless; second, in our initial discussion of pre-teams we excluded pairs (μ,μ') with $|C_{(\mu,\mu')}(2)| \leq B+J$. The following definitions remove these difficulties.

Definition 11.8 (Virtual team members)**.** *If a pre-team $\hat{\mathcal{T}}(\mu,\mu')$ of type (G4) is disqualified under the terms of Definition 11.6 and the smaller team necessitating disqualification is $\hat{\mathcal{T}}(\mu_1,\mu_2)$, then we define $(\mu,\mu') \in_v \hat{\mathcal{T}}(\mu_1,\mu_2)$ and $\hat{\mathcal{T}}(\mu,\mu') \subset_v \hat{\mathcal{T}}(\mu_1,\mu_2)$. We extend the relation $\subset_v$ to be transitive and extend $\in_v$ correspondingly. If $(\mu,\mu') \in_v \mathcal{T}$ then (μ_2,μ') is said to be a* virtual member *of the team $\mathcal{T}$.*

Definition 11.9. *If (μ,μ') is such that $1 \leq |C_{(\mu,\mu')}(2)| \leq B+J$ and (μ,μ') is neither a member nor a virtual member of any previously defined team, then we define $\mathcal{T}_{(\mu,\mu')} := \{(\mu,\mu')\}$ to be a (short) team with $\|\mathcal{T}_{(\mu,\mu')}\| = |C_{(\mu,\mu')}(2)|$.*

Lemma 11.10 (cf. Lemma 9.10, [7]). *Every $(\mu, \mu') \in \mathcal{Z}$ with $C_{(\mu,\mu')}(2)$ non-empty is a member or a virtual member of exactly one team, and there are less than $2|\partial\Delta|$ teams.*

Proof. The first assertion is an immediate consequence of the preceding three definitions, and the second follows from the fact that $|\mathcal{Z}| < 2|\partial\Delta|$. □

11.2. **Pincers associated to teams of genesis (G3).** [cf. Subsection 9.3,[7]]

In this subsection we describe a pincer $\Pi_{\mathcal{T}}$ canonically associated to each team of genesis (G3), as in [7, Subsection 9.3]. The only real difference between the definitions here and those in [7] is the use of robust past and beaded Nielsen paths. Sadly, this variation leads to complications in the cascade of pincers; see Definition 11.17 and Remark 10.24.

Definition 11.11 (cf. Definition 9.11, [7]). *The* narrow past *of a team $\mathcal{T}$ at time t consists of those beaded Nielsen paths whose beads are displayed in their colour and whose future is contained in $\mathfrak{T}$. The narrow past may have several components at each time, the set of which are ordered left to right according to the ordering in $\mathfrak{T}$ of their futures. We call these components* sections.

For the remainder of this subsection we consider only long teams of genesis (G3).

The following lemma follows from the definition of teams of genesis (G3) in a straightforward manner.

Lemma 11.12. *Let $\mathcal{T}$ be a team of genesis (G3). There exist beads $y(\mathcal{T})$ and $y_1(\mathcal{T})$ of different colours, both lying strictly between the immediate past of the swollen present of $\mathcal{T}$ and the robust past of $\hat{\rho}(\mu, \mu')$, so that $y(\mathcal{T})$ is bitten by $y_1(\mathcal{T})$ and this is not HNP-biting.*

Definition 11.13 (The Pincer $\tilde{\Pi}_{\mathcal{T}}$). *Choose a leftmost pair of beads $y(\mathcal{T}), y_1(\mathcal{T})$ satisfying Lemma 11.12, and let $x(\mathcal{T})$ be the leftmost edge in $y(\mathcal{T})$. Let $x_1(\mathcal{T})$ be the edge in $y_1(\mathcal{T})$ which is the past of the edge which cancels with the leftmost edge in the immediate future of $x(\mathcal{T})$.*

Define $\tilde{p}_l(\mathcal{T})$ to be the path in the family forest $\mathcal{F}$ that traces the history of $x(\mathcal{T})$ to $\partial\Delta$, and let $\tilde{p}_r(\mathcal{T})$ be the path that traces the history of $x_1(\mathcal{T})$.

Define $\tilde{t}_2(\mathcal{T})$ to be the earliest time at which the paths $\tilde{p}_l(\mathcal{T})$ and $\tilde{p}_r(\mathcal{T})$ lie in the same corridor.

Remark 11.14. *Since the pair $y(\mathcal{T}), y_1(\mathcal{T})$ in Definition 11.13 are the leftmost pair satisfying Lemma 11.12, any non-vanishing beads which*

lie between $\mathfrak{T}$ and this pair are involved in HNP-biting and are of lower weight than $y_1(\mathcal{T})$, by the Weighted Buffer Lemma 10.5.

Lemma 11.15. *The segments of the paths $\tilde{p}_l(\mathcal{T})$ and $\tilde{p}_r(\mathcal{T})$, together with the path joining them along the bottom of the corridor at time $\tilde{t}_2(\mathcal{T})$ form a pincer.*

Proof. Note that when choosing the beads $y(\mathcal{T})$ and $y_1(\mathcal{T})$ we excluded HNP-cancellation. That the paths in the statement of the lemma form a pincer then follows immediately from the definition of pincers. □

We denote the pincer described in Lemma 11.15 above by $\tilde{\Pi}_{\mathcal{T}}$.

11.3. **The cascade of pincers.** The Pincer Lemma argues for the regular disappearance of colours within a pincer during those times when more than two colours continue to survive along its corridors. However, when there are only two colours, the situation is more complicated.

Recall that the constant T_0 is as in Proposition 10.7, subject to the requirement that $T_0 \geq T_0'$ as in the assumption immediately after Proposition 10.9. The pincer S_Π associated to a pincer Π is defined in Definition 10.10.

Lemma 11.16. *One of the following must occur:*

1. $\mathrm{time}(S_{\tilde{\Pi}_{\mathcal{T}}}) > t_1(\mathcal{T}) - T_0$;
2. *the path $\tilde{p}_l(\mathcal{T})$ and the entire narrow past of $\mathcal{T}$ are not in the same corridor at time $t_1(\mathcal{T}) - T_0$; or*
3. *at time $t_1(\mathcal{T}) - T_0$ the path $\tilde{p}_l(\mathcal{T})$ and the narrow past of $\mathcal{T}$ are separated by a path which does not split as a beaded path whose beads are either Nielsen paths or of weight less than $\tilde{p}_l(\mathcal{T})$.*

Proof. If not, the Weighted Two Colour Lemma (Lemma 10.9) would give a contradiction, since there *is* to be interaction between the beads $y(\mathcal{T})$ and $y_1(\mathcal{T})$ at time $t_1(\mathcal{T})$, and this interaction is not HNP-biting. □

We now consider each of the three cases in turn, seeking a definition of times $t_2(\mathcal{T})$ and $t_3(\mathcal{T})$ and (possibly) a pincer $\Pi_{\mathcal{T}}$. The following definition is entirely analogous to [7, Definition 9.13], with the appropriate translations.

Definition 11.17 (cf. Definition 9.13, [7])**.**

1. *Suppose some section of the narrow past of $\mathcal{T}$ is not in the same corridor as $\tilde{p}_l(\mathcal{T})$ at time $t_1(\mathcal{T}) - T_0$: In this case*[20] *we define $t_2(\mathcal{T}) = t_3(\mathcal{T})$ to be the earliest time at which the entire narrow*

[20]this includes the possibility that $\tilde{p}_l(\mathcal{T})$ does not exist at time $t_1(\mathcal{T}) - T_0$

past of $\mathcal{T}$ lies in the same corridor as $\tilde{p}_l(\mathcal{T})$ and has length at least λ_0.

(2) *Suppose that Case (1) does not occur and* $\mathrm{time}(S_{\tilde{\Pi}_\mathcal{T}}) > t_1(\mathcal{T}) - T_0$. *We define $\Pi_\mathcal{T} = \tilde{\Pi}_\mathcal{T}$ and $t_3(\mathcal{T}) = \mathrm{time}(S_{\Pi_\mathcal{T}})$. If the narrow past of $\mathcal{T}$ at time $t_1(\mathcal{T}) - T_0$ has length less than λ_0, we define $t_2(\mathcal{T}) = t_3(\mathcal{T})$, and otherwise $t_2(\mathcal{T}) = \tilde{t}_2(\mathcal{T})$.*

(3) *Suppose that neither Case (1) or Case (2) occurs: In this case, Lemma 11.16(3) pertains. We pass to the latest time at which there is a path between $\tilde{p}_l(\mathcal{T})$ and the narrow past of $\mathcal{T}$ which has an edge of at least the same weight as $\tilde{p}_l(\mathcal{T})$ at this time and is not contained in a Nielsen path. Choose a pair of beads $y'(\mathcal{T})$, $y'_1(\mathcal{T})$ as in Lemma 11.12, as well as edges $x'(\mathcal{T})$, $x'_1(\mathcal{T})$. Let $\tilde{p}'_l(\mathcal{T})$ be the path tracing the history of $x'(\mathcal{T})$. Let $\tilde{p}'_r(\mathcal{T})$ trace the history of the edge $x'_1(\mathcal{T})$ that cancels $x'(\mathcal{T})$. Let $\tilde{t}'_2(\mathcal{T})$ be the earliest time at which the paths $\tilde{p}'_l(\mathcal{T})$ and $\tilde{p}'_r(\mathcal{T})$ lie in the same corridor and consider the pincer formed by these paths after time $\tilde{t}'_2(\mathcal{T})$ and the path joining them along the bottom of the corridor at time $\tilde{t}'_2(\mathcal{T})$.*

We now repeat our previous analysis with the primed objects $\tilde{p}'_l(\mathcal{T}), \tilde{t}'_2(\mathcal{T})$, etc. *in place of $\tilde{p}_l(\mathcal{T}), \tilde{t}_2(\mathcal{T})$,* etc., *checking whether we now fall into Case (1) or (2); if we do not then we pass to $\tilde{p}''_l(\mathcal{T})$,* etc.. *We iterate this analysis until we fall into Case (1) or (2), at which point we acquire the desired definitions of $\Pi_\mathcal{T}, t_2(\mathcal{T})$ and $t_3(\mathcal{T})$.*

Define $p_l(\mathcal{T})$ (resp. $p_r(\mathcal{T})$) to be the left (resp. right) boundary path of the pincer $\Pi_\mathcal{T}$ extended backwards in time through $\mathcal{F}$ to $\partial\Delta$. Define $p_l^+(\mathcal{T})$ to be the sequence of edges (one at each time) lying on the leftmost of the primed $\tilde{p}_l(\mathcal{T})$ from the top of $\pi_\mathcal{T}$ to time $t_1(\mathcal{T})$.

Definition 11.18 (cf. Definition 9.14, [7])**.** *Let $\mathcal{T}$ be a long team of genesis (G3). We define $\chi_P(\mathcal{T})$ to be the set of colours containing the paths $\tilde{p}_l(\mathcal{T}), \tilde{p}'_l(\mathcal{T}), \tilde{p}''_l(\mathcal{T}), \ldots$ that arise in Case (3) of Definition 11.17 but do not become $p_l(\mathcal{T})$.*

Lemma 11.19 (cf. Lemma 9.15, [7])**.**

(1) *If $\mathcal{T}$ is a long team of genesis (G3),*

$$t_1(\mathcal{T}) - t_3(\mathcal{T}) \leq T_0(|\chi_P(\mathcal{T})| + 1).$$

(2) *If $\mathcal{T}_1$ and $\mathcal{T}_2$ are distinct teams then $\chi_P(\mathcal{T}_1) \cap \chi_P(\mathcal{T}_2) = \emptyset$.*

11.4. **The length of teams.** This subsection follows [7, Subsection 9.4]. We consider the lengths of arbitrary teams.

Definition 11.20 (cf. Definition 9.16, [7]). *Let $\mathcal{T}$ be a team. Define $\mathrm{down}_1(\mathcal{T}) \subset \partial\Delta$ to consist of those edges e that are labelled by some t_i and satisfy one of the following conditions:*

1. *e is at the left end of a corridor containing a section of the narrow past of $\mathcal{T}$ that is not leftmost at that time;*
2. *e is at the right end of a corridor containing a section of the narrow past of $\mathcal{T}$ that is not rightmost at that time;*
3. *e is at the right end of a corridor which contains the rightmost section of the narrow past of $\mathcal{T}$ at that time but which does not intersect $p_l(\mathcal{T})$.*

Definition 11.21 (cf. Definition 9.17, [7]). *Define $\partial^{\mathcal{T}} \subset \partial\Delta$ to be the intersection of the narrow past of $\mathcal{T}$ with $\partial\Delta$.*

Lemma 11.22 (cf. Lemma 9.18, [7]).

(1) For distinct teams $\mathcal{T}_1$ and $\mathcal{T}_2$, the sets $\partial^{\mathcal{T}_1}$ and $\partial^{\mathcal{T}_2}$ are disjoint.
(2) For distinct teams $\mathcal{T}_1$ and $\mathcal{T}_2$, the sets $\mathrm{down}_1(\mathcal{T}_1)$ and $\mathrm{down}_1(\mathcal{T}_2)$ are disjoint.

Definition 11.23 (cf. Definition 9.19, [7]). *Suppose that $\mathcal{T}$ is a team of genesis (G3). We define $Q(\mathcal{T})$ be the set of edges ε with the following properties: $p_l(\mathcal{T})$ passes through ε before time $t_3(\mathcal{T})$, the corridor S with $\varepsilon \in \perp(S)$ contains the entire narrow past of $\mathcal{T}$, and this narrow past has length at least λ_0.*

The following lemma reduces the task of bounding the total length of teams to that of bounding the size of the sets $Q(\mathcal{T})$. Its proof follows that of [7, Lemma 9.20].

Lemma 11.24 (cf. Lemma 9.20, [7]).

(1) If the genesis of $\mathcal{T}$ is of type (G1) or (G2), then

$$\|\mathcal{T}\| \leq 2LC_4|\mathrm{down}_1(\mathcal{T})| + |\partial^{\mathcal{T}}|.$$

(2) If the genesis of $\mathcal{T}$ is of type (G3), then

$$\|\mathcal{T}\| \leq 2C_4|\mathrm{down}_1(\mathcal{T})| + |\partial^{\mathcal{T}}| + 2LC_4|Q(\mathcal{T})| + 2LC_4T_0(|\chi_P(\mathcal{T})|+1) + \lambda_0.$$

11.5. **Bounding the size of $Q(\mathcal{T})$.** Let $\mathcal{G}_3$ be the set of long teams of genesis (G3) for which $Q(\mathcal{T})$ is nonempty. Our goal for the remainder of this section is to find a bound for $\sum_{\mathcal{T}\in\mathcal{G}_3} |Q(\mathcal{T})|$.

Lemma 11.25 (cf. Lemma 9.22, [7]). *For all $\mathcal{T} \in \mathcal{G}_3$*

$$t_3(\mathcal{T}) - t_2(\mathcal{T}) = \mathrm{Life}(\Pi_{\mathcal{T}}) \leq T_1(|\chi(\Pi_{\mathcal{T}})| + 1).$$

Lemma 11.26 (cf. Lemma 9.23, [7]). *If $\mathcal{T}_1, \mathcal{T}_2 \in \mathcal{G}_3$ are distinct teams then $\chi(\Pi_{\mathcal{T}_1}) \cap \chi(\Pi_{\mathcal{T}_2}) = \emptyset$.*

Proof. The pincers $\Pi_{\mathcal{T}_i}$ are disjoint or else one is contained in the other. In the latter case, say $\Pi_{\mathcal{T}_1} \subset \Pi_{\mathcal{T}_2}$, the definition of nesting (Definition 10.23), and of the pincer associated to a team (Definition 11.17) ensure that $\Pi_{\mathcal{T}_1}$ is actually nested in $\Pi_{\mathcal{T}_2}$ (cf. Remark 10.24). □

Corollary 11.27 (cf. Corollary 9.24, [7]).

$$\sum_{\mathcal{T} \in \mathcal{G}_3} t_3(\mathcal{T}) - t_2(\mathcal{T}) \leq 3T_1|\partial\Delta|.$$

We have now reduced our task for this section to bounding the number of edges in the $Q(\mathcal{T})$ which occur before $t_2(\mathcal{T})$; this is the cardinality of the following set.

Definition 11.28 (cf. Definition 9.25, [7]). *For a team $\mathcal{T} \in \mathcal{G}_3$ we define* $\mathrm{down}_2(\mathcal{T})$ *to be the set of edges in $\partial\Delta$ that lie at the right-hand end of a corridor containing an edge in $Q(\mathcal{T})$ before time $t_2(\mathcal{T})$.*

Just as in [7], it is not necessarily the case that the sets $\mathrm{down}_2(\mathcal{T})$ are disjoint for distinct teams, and we must deal with the possibility of 'double-counting'.

The left-to-right ordering defined on paths in $\mathcal{F}$ in [7, §9] is defined in the current context exactly as in [7].

Notation: Let $\mathcal{G}_3'$ be the set of teams $\mathcal{T} \in \mathcal{G}_3$ with $\mathrm{down}_2(\mathcal{T}) \neq \emptyset$.

Lemma 11.29 (cf. Lemma 9.26, [7]). *Consider $\mathcal{T} \in \mathcal{G}_3'$. If a path p in $\mathcal{F}$ is to the left of $p_l(\mathcal{T})$ and a path q is to the right of $p_r(\mathcal{T})$, then there is no corridor connecting p to q at any time $t < t_2(\mathcal{T})$.*

Definition 11.30 (cf. Definition 9.27, [7]). *$\mathcal{T}_1 \in \mathcal{G}_3'$ is said to be* below *$\mathcal{T}_2 \in \mathcal{G}_3'$ if $p_l(\mathcal{T}_1)$ and $p_r(\mathcal{T}_1)$ both lie between $p_l(\mathcal{T}_2)$ and $p_r(\mathcal{T}_2)$ in the left-to-right ordering.*

$\mathcal{T}_1$ is to the left *of $\mathcal{T}_2$ if both $p_l(\mathcal{T}_1)$ and $p_r(\mathcal{T}_2)$ lie to the right of $p_r(\mathcal{T}_1)$.*

We say that $\mathcal{T}$ is at depth 0 *if there are no teams above it. Then, inductively, we say that a team $\mathcal{T}$ is at depth $d+1$ if d is the maximum depth of those teams above $\mathcal{T}$.*

A final depth *team is one with no teams below it.*

Note that there is a complete left-to-right ordering of those teams in $\mathcal{G}_3'$ at any given depth.

Lemma 11.31 (cf. Lemma 9.28, [7]). *If there is a team from $\mathcal{G}_3'$ below a team $\mathcal{T} \in \mathcal{G}_3'$, then $t_1(\mathcal{T}) \geq \mathrm{time}(S_0) \geq t_2(\mathcal{T})$.*

Proof. The proof from [7] works almost verbatim. In particular, the same proof shows that $\text{time}(S_0) \geq t_2(\mathcal{T})$.

To see that $t_1(\mathcal{T}) \geq \text{time}(S_0)$, suppose that $\mathcal{T}'$ is a team below $\mathcal{T}$. Associated to the team $\mathcal{T}'$ we have the beaded Nielsen path $\mathfrak{T}'$, which is to be consumed by some reaper. The definitions of nesting and of the pincer $\Pi_{\mathcal{T}'}$ ensure that this consumption of $\mathfrak{T}'$ must occur before time $t_1(\mathcal{T})$. On the other hand, $\mathfrak{T}$ has a non-empty future or past in S_0. □

With the preceding results in hand, a direct translation of the proof of Lemma 9.29, [7] finishes the work of this section:

Lemma 11.32 (cf. Lemma 9.29, [7])**.** *There exist sets of colours $\chi_c(\mathcal{T})$ and $\chi_\delta(\mathcal{T})$ associated to each team $\mathcal{T} \in \mathcal{G}_3'$ such that the sets associated to distinct teams are disjoint and the following inequalities hold.*

For each fixed team $\mathcal{T}_0 \in \mathcal{G}_3'$ (of depth d say), the teams of depth $d+1$ that lie below $\mathcal{T}_0$ may be described as follows:

- *There is at most one distinguished team $\mathcal{T}_1$, and*

$$\|\mathcal{T}_1\| \leq 2B\Big(T_1(1 + |\chi(\Pi_{\mathcal{T}_0})|) + T_0(|\chi_P(\mathcal{T}_0)| + 1)\Big).$$

- *There are some number of final-depth teams.*
- *For each of the remaining teams $\mathcal{T}$ we have*

$$|\text{down}_2(\mathcal{T}_0) \cap \text{down}_2(\mathcal{T})| \leq T_1\Big(1 + |\chi_c(\mathcal{T})|\Big) + T_0\Big(|\chi_\delta(\mathcal{T})| + 2\Big).$$

Corollary 11.33 (cf. Corollary 9.30, [7])**.** *Summing over the set of teams $\mathcal{T} \in \mathcal{G}_3'$ that are not distinguished, we get*

$$\sum_{\mathcal{T}} \Big|\text{down}_2(\mathcal{T})\Big| \leq 2\Big|\bigcup_{\mathcal{T}} \text{down}_2(\mathcal{T})\Big| + \sum_{\mathcal{T}} T_1\Big(1+|\chi_c(\mathcal{T})|\Big) + \sum_{\mathcal{T}} T_0\Big(|\chi_\delta(\mathcal{T})|+2\Big).$$

Summing over the same set of teams again, we finally obtain:

Corollary 11.34.

$$\sum_{\mathcal{T}} |\text{down}_2(\mathcal{T})| \leq |\partial\Delta|(2 + 3T_1 + 5T_0).$$

12. The Bonus Scheme

This section closely follows [7, Section 10]. We have at last reached a stage where the proofs from [7] can be translated without significant modification.

In the previous section we defined teams and obtained a global bound on $\sum \|\mathcal{T}\|$. If $C_{(\mu,\mu')}(2)$ is non-empty then (μ, μ') is a member or virtual

member of a unique team. If the team is such that $t_1(\mathcal{T}) \geq \text{time}(S_0)$, then no member of the team is virtual and we have the inequality

$$\|\mathcal{T}\| \geq \sum_{(\mu_1,\mu_2)\in\mathcal{T}} |C_{(\mu_1,\mu_2)}| - B(B+1),$$

established in Lemma 11.5. This inequality might fail in case $t_1(\mathcal{T}) < \text{time}(S_0)$. The *bonus scheme* assigns additional edges to teams in order to compensate for this failure.

By definition, at time $t_1(\mathcal{T})$ the reaper $\rho = \rho_{\mathcal{T}}$ lies immediately to the right of $\mathfrak{T}$. The beads of $\mathfrak{T}$ not consumed from the right by ρ by $\text{time}(S_0)$ have a preferred future in S_0. This preferred future, if contained in a single colour, lies in $C_{(\mu_1,\mu_2)}(2)$ for some member $(\mu_1, \mu_2) \in \mathcal{T}$. It could also intersect more than one colour [21]. However, not all beads in the $C_{(\mu_1,\mu_2)}(2)$ need arise in this way: some may not have a Nielsen bead as an ancestor at time $t_1(\mathcal{T})$. And if (μ_1, μ_2) is only a virtual member of $\mathcal{T}$, then no bead of $C_{(\mu_1,\mu_2)}(2)$ lies in the future of $\mathfrak{T}$. The *bonus* beads in $C_{(\mu_1,\mu_2)}(2)$ are a certain subset of those that do not have a Nielsen bead as an ancestor at time $t_1(\mathcal{T})$. They are defined as follows.

Definition 12.1. *Let $\mathcal{T}$ be a team with $t_1(\mathcal{T}) < \text{time}(S_0)$ and consider a time t with $t_1(\mathcal{T}) < t < \text{time}(S_0)$.*

The swollen future *of $\mathcal{T}$ at time t is defined as in Definition 8.16 with respect to the interval $\mathfrak{T}$, which lies at time $t_1(\mathcal{T})$.*

Let ϵ be a non-Nielsen bead that lies immediately to the left of the swollen future of $\mathcal{T}$, but whose immediate ancestor is not a right linear edge in this position. If the path from ϵ to the reaper $\rho_{\mathcal{T}}$ of $\mathcal{T}$ is a GEP*, then we say that ϵ is a* rascal*. Otherwise, if ϵ provides more Nielsen beads than the reaper consumes, then ϵ is a* terror*.*

In both cases, the bonus provided by *ϵ is the set of beads in the swollen future of $\mathcal{T}$ in S_0 that have ϵ as their most recent ancestor which is not a Nielsen bead, and which are eventually consumed by $\rho_{\mathcal{T}}$.*

The set $\text{bonus}(\mathcal{T})$ *is the union of the bonuses provided to $\mathcal{T}$ by all rascals and terrors.*

Lemma 12.2 (cf. Lemma 10.2, [7])**.** *For any team $\mathcal{T}$,*

$$\sum_{(\mu_1,\mu_2)\in\mathcal{T} \text{ or } (\mu_1,\mu_2)\in_v\mathcal{T}} |C_{(\mu_1,\mu_2)}(2)| \leq \|\mathcal{T}\| + |\text{bonus}(\mathcal{T})| + B + J.$$

Note that the GEP which contains a rascal in the above definition is not displayed. We now proceed to bound the total bonus provided

[21] Since Nielsen beads have bounded length, and there is a bound on the number of adjacencies of colours, there are relatively few such beads.

to teams by all rascals and terrors. Terrors are straightforward to deal with.

Lemma 12.3 (cf. Lemma 10.3, [7])**.** *The sum of the lengths of the bonuses provided to all teams by terrors is less than* $2L|\partial\Delta|$.

Proof. Let ϵ by a terror, associated to a team $\mathcal{T}$. Since the region from ϵ to the reaper of $\mathcal{T}$ is not a GEP, ϵ must be right-fast. Therefore, it will be separated from the team to which it is associated after one unit of time. Hence the bonus that ϵ provides is at most L.

That there can be at most one terror per adjacency of colours follows in a straightforward manner from Lemma 5.6 and the definition of terror.

Thus the total contribution of all terrors is less than $2L|\partial\Delta|$. □

In parallel with [7, Definition 10.4], we make the following

Definition 12.4. *Fix a team* $\mathcal{T}$ *with* $t_1(\mathcal{T}) < \mathrm{time}(S_0)$ *and consider the interval of time* $[\tau_0(\epsilon), \tau_1(\epsilon)]$*, where* $\tau_0(\epsilon)$ *is the time at which a rascal* ϵ *appears at the left end of the swollen future of* $\mathcal{T}$*, and* $\tau_1(\epsilon)$ *is the time at which the robust future of* ϵ *is no longer to the immediate left of the future of the swollen future of* $\mathcal{T}$*.*

In the case where the robust future $\hat{\epsilon}$ *of* ϵ *at time* $\tau_1(\epsilon)$ *is cancelled from the left by an edge* e'*, we define* $\tau_2(\epsilon)$ *to be the earliest time when the pasts of* $\hat{\epsilon}$ *and* e' *are in the same corridor. The path in* $\mathcal{F}$ *that traces the past of* $\hat{\epsilon}$ *is denoted* p_ϵ *and the past following the ancestors of* e' *from* $\tau_2(\epsilon)$ *to* $\tau_1(\epsilon)$ *is denoted* p'_ϵ*. The pincer*[22] *formed by* p_ϵ*,* p'_ϵ *and the corridor joining them at time* $\tau_2(\epsilon)$ *is denoted* Π_ϵ*.*

The only essential difference between the above definition and [7, Definition 10.4] is the use of the robust future of ϵ rather than the pp-future.

With this definition in hand, the remaining results from [7, Section 10] may be translated directly, yielding in particular:

Proposition 12.5 (cf. Lemma 10.13, [7])**.** *Summing over all teams that are not short, we have*

$$\sum_{\mathcal{T}} |\mathrm{bonus}(\mathcal{T})| \leq \Big((B{+}3)(3T_1{+}2T_0)L{+}6BT_1{+}4BT_0{+}2\lambda_0{+}2B{+}5L{+}1\Big)|\partial\Delta|.$$

[22] we include the degenerate case here where the "pincer" has no colours other than those of ϵ and e'.

13. From Bead Norm to Length

The output of the results up to now is a bound for the bead norm of our corridor S_0. In order to complete the proof of Theorem 4.1 in the case of the specified IRTT f (which implies Theorem A) we need to turn this into a bound on the length of S_0. For this we need to bound the total length of the GEPs and ΨEPs in S_0 which have length more than J (or indeed any other fixed length). In this section we explain how the techniques of the bonus scheme can be used to establish such a bound.

If a bead ρ in $\mu(S_0)$ has length greater than J, it is either a GEP or a ΨEP. If it is a ΨEPthen we may trace its past: at each time, this past is either of length at most J or else is a ΨEP or a GEP. Whilst this past remains a ΨEP, the number of Nielsen paths will decrease with each backwards step in time, so at some point in the past of ρ, it must become a GEP.

Suppose now that ρ is a GEP. The past of a GEP is either a GEP or else has length at most J. Thus, the length of the GEP decreases as we go into the past until eventually it is of length at most J.

There is a strong analogy between teams of genesis (G4) and long GEPs and ΨEPs. On one end of a long bead is a linear edge which consumes the Nielsen beads in the middle. This linear edge can be considered as a reaper. On the other end of a GEP is a linear edge which can be considered as a rascal. The moment when the past of a ΨEP becomes a GEP is analogous to $\tau_1(\epsilon)$ from the bonus scheme, and so a ΨEP in S_0 can be thought of as a team with a rascal ϵ with $\tau_1(\epsilon) \leq \text{time}(S_0)$. Similarly, a long GEP in S_0 can be thought of as a team with a rascal ϵ so that $\tau_1(\epsilon) > \text{time}(S_0)$.

We can define the bonus associated to such a rascal exactly as we did in the previous section. Since we are in the setting of genesis type (G4), all of the Nielsen beads in a long GEP or ΨEP are in the bonus. Thus it is enough to bound the total of the bonuses associated to long GEPs and ΨEPs.

The only thing we need to be able to follow the bonus scheme directly is a bound on the number of long GEPs and ΨEPs in S_0.

Lemma 13.1. *The number of beads of length greater than J in S_0 is less then $4|\partial\Delta|$.*

Proof. Let ρ be a bead in S_0 of length greater than J, and assign a time $\tau_1(\rho)$ to ρ as described above. If ρ is a GEP then $\tau_1(\rho) > \text{time}(S_0)$, whilst if ρ is a ΨEP then $\tau_1(\rho) \leq \text{time}(S_0)$.

Let ρ' be the past or future of ρ at time $\tau_1(\rho) - 1$. Consider the 'event' at time $\tau_1(\rho)$ which stops the robust future of ρ' being a GEP.

This 'event' is either an intrusion of the boundary, a singularity, or else there is an associated pincer caused by a cancellation from another colour. There are less than $|\partial\Delta|$ events of each of the first two types.

The Buffer Lemma ensures that there is at most one event of the third type for each adjacency of colours. An application of Lemma 2.8 completes the proof. □

A bound on the total length of long beads in S_0 now follows exactly as in the bonus scheme from Section 12 (the detailed arguments being in [7, Section 10]).

13.1. **The end of the main road.** In Section 4 we discussed how Theorem A follows from Theorem 4.2 and Proposition 4.3. The bound that we just established on the total length of long beads in S_0 proves Proposition 4.3. The output of our estimates in the previous sections bounded the bead norm of S_0 by a linear function of $|\partial\Delta|$, and Theorem 4.2 follows from this because

$$[S]_\beta \leq B\|S\|_\beta,$$

(see Lemma 7.5).

Thus the proof of Theorem A is finally at an end, and the reader can join us in wondering why a statement as simple and engaging as Theorem 4.1 should require such a complicated proof.

14. Corridor Length Functions and Bracketing

In this section we prove Theorem 4.1 in full generality and deduce the Bracketing Theorem from it. Our proof of Theorem 4.1 proceeds via a discussion of *corridor length functions* for more general semidirect products and mapping tori. Such functions should be regarded as measuring the complexity of van Kampen diagrams in the spirit of isoperimetric and isodiametric functions. We prove the following results (see Subsection 14.2 for precise definitions of the terms involved).

Proposition 14.1. *Let G_1 and G_2 be compact combinatorial complexes with fundamental group Π, and for $i = 1, 2$ let $f_i : G_i^{(1)} \to G_i^{(1)}$ be an edge-path map of 1-skeleta inducing $\phi \in \mathrm{Out}(\Pi)$. Then the t-corridor length function for the mapping torus $M(f_1)$ is $\simeq$ equivalent to that of $M(f_2)$.*

Proposition 14.2. *If Π is finitely generated and $\Gamma = \Pi \rtimes_\phi \mathbb{Z}$ is finitely presented, then for every positive integer p, the corridor length function of Π is $\simeq$ equivalent to that of $\Gamma_p = \Pi \rtimes_{\phi^p} \mathbb{Z}$*

In the previous section we completed the proof of Theorem 4.1 in the case of one particular IRTT representative f of a certain power of an arbitrary free-group automorphism ϕ. The above results complete the proof in the general case. Before turning to the proof of these results, we explain how the Bracketing Theorem stated in the introduction is obtained by applying Theorem 4.1 to the most naive topological representation of a free group automorphism ϕ.

14.1. **The Bracketing Theorem.** The terms in the following theorem were defined in the introduction.

Theorem C. *There exists a constant $K = K(\phi, \mathcal{B})$ such that any word $w \equiv e_1 \dots e_n$ that represents the identity in $F \rtimes_\phi \mathbb{Z}$ admits a t-complete bracketing $\beta_1, \dots, \beta_m$ such that the content c_i of each β_i satisfies $d_F(1, c_i) \leq Kn$.*

Proof. We work with the mapping torus M of the obvious realisation of ϕ on the graph with one vertex whose edges are indexed by $\mathcal{B}$. Given a word w, we consider a minimal-area van Kampen diagram over M with boundary label w. We insert a bracket $w_1(w_2)w_3$ if and only if there is a t-corridor whose ends are labelled by the initial and terminal letters of w_2. (One must allow t-corridors of zero length in this description; one would exclude them by making the easy reduction to words that have no proper sub-words that are null-homotopic.)

These brackets are pairwise compatible because distinct t-corridors cannot cross. And because every t-edge in the boundary of a van Kampen diagram is the end of a (perhaps zero-length) corridor, the bracketing is complete. The content of the bracket is the freely reduced form of the label along the top or bottom of the corridor (according to the orientation of the sentinels). In the former case, the length of the corridor bounds the length of this label, and in the latter case one has to multiply the length by at most $L = \max\{|\phi(b)| : b \in \mathcal{B}\}$. □

14.2. **Corridor length functions.** If Π is a group with finite generating set $\mathcal{A}$ and $\phi \in \mathrm{Aut}(\Pi)$ is such that $\Gamma = \Pi \rtimes_\phi \mathbb{Z}$ is finitely presented, then Γ has a finite presentation of the form

$$\big\langle \mathcal{A},\, t \mid \mathcal{R},\, t^{-1}at = \hat{\phi}(a)\ (a \in \mathcal{A}) \big\rangle,$$

where t is the generator of the visible $\mathbb{Z}$, the relations $\mathcal{R}$ involve only the letters $\mathcal{A}$, and $\hat{\phi}(a) \in F(\mathcal{A})$ is equal to $\phi(a)$ in Π.

We are concerned with the geometry of t-corridors in van Kampen diagrams over such presentations. Thus we associate to the presentation the *t-corridor length function* $\Lambda : \mathbb{N} \to \mathbb{N}$, which is defined as follows. For each $w \in F(\mathcal{A} \cup \{t\})$ with $w = 1$ in Γ, we choose a van

Kampen diagram for w in which the length of the longest t-corridor is as small as possible, and we define $\lambda_t(w)$ to be this length. We then define

$$\Lambda(n) := \max\{\lambda_t(w) \mid w =_\Gamma 1,\ |w| \leq n\}.$$

More generally, since we have a well-defined notion of van Kampen diagram and t-corridor in the setting of mapping tori of *edge-path maps*[23] of combinatorial complexes, we can define the *t-corridor length function* for such a complex.

14.3. **Invariance under change of topological representative.** The scheme of the following proof follows the standard method of showing that features of the geometry of van Kampen diagrams are preserved under quasi-isometry. However, one has to be careful to deal only with fibre-preserving maps in order to retain control over the t-corridor structure.

Proof of Proposition 14.1. We have a cocompact action of $\Gamma = \Pi \rtimes_\phi \mathbb{Z}$ on the universal cover $X_i = \tilde{M}(f_i)$ for $i = 1, 2$, where the action of Π leaves invariant the connected components $C_{i,m}$ of the preimage of $G_i \subset M(f_i)$ and the generator t of $\mathbb{Z}$ acts so that $t^r.C_{i,m} = C_{i,m+r}$.

The cocompactness of the actions means that there exist constants δ_1, δ_2 so that every vertex in $C_{i,m}$ is within a distance δ_i of any Π-orbit of vertices in $C_{i,m}$, where distance is measured in the combinatorial metric on the 1-skeleton (unit edge lengths).

We define Γ-equivariant quasi-isometries between the 1-skeleta of the X_i as follows. First we pick base vertices $x_i \in C_{i,0}$ and define $g_1 : \gamma.x_1 \mapsto \gamma.x_2$ and $g_2 : \gamma.x_2 \mapsto \gamma.x_1$. Then, for each vertex $v \in C_{i,m} \smallsetminus \Gamma.x_i$ we choose a closest element $v' \in \Gamma.x_i \cap C_{i,m}$ and define $g_i(v) := g(v')$. Next, we extend to the edges in $C_{i,m}$ by sending each to a shortest edge path connecting the images of its vertices. Finally, we extend g_i to t-edges in X_i so that it sends each such homeomorphically onto the t-edge joining the images of its endpoints.

With the maps g_1, g_2 in hand, we can now push van Kampen diagrams back and forth between X_1 and X_2 as in the standard proof of the qi-invariance of Dehn functions (cf. [10], page 143). Thus, given a loop ℓ in the 1-skeleton of X_1, labelled $u_1 t^{\varepsilon_1} u_2 \dots u_l t^{\varepsilon_l}$ we consider the loop $g_1 \circ \ell$ in $X_2^{(1)}$ and fill it with a van Kampen diagram Δ so as minimize the length of the longest t-corridor. We will be done if we can bound $\lambda_t(\ell)$ by a linear function of this length.

Viewing Δ as a map from a cellulated 2-disc to X_2, we compose it with g_2 to obtain a map to X_1. This new map is obtained from Δ by

[23] an *edge-path map* is a cellular map that sends edges to edge-paths

simply changing the labels on the edges: the t-edges are unchanged while the edges labelled by 1-cells in G_2 are now labelled by edge-paths in the 1-skeleton of G_1 whose length is bounded by the constants of the quasi-isometry g_2; the boundary label of the diagram will be $\ell' = v_1 t^{\varepsilon_1} v_2 \dots v_l t^{\varepsilon_l}$, where the v_j are edge-paths of uniformly bounded length and each v_j is contained in the same component C_{1,m_j} as u_j. (This is the point at which we use the fact that we chose our quasi-isometries to respect fibres.) The faces of this diagram can be filled with van Kampen diagrams in X_1; in the case of 2-cells with no t-labels, we use only lifts of 2-cells from G_1; in the case of 2-cells labelled $t^{-1}\rho t\sigma$ we divide them into (short) t-corridors in the obvious manner. The result[24] is a van Kampen diagram for ℓ' in X_1 whose t-corridors are in bijection with those of Δ and whose length is bounded by k times the length of those in Δ, where k is a constant that depends only on our quasi-isometries.

To complete the desired diagram filling our original loop ℓ, we need an annular diagram between ℓ and ℓ' that does not disrupt the structure of t-corridors in Δ'. To this end, we join the vertices of u_j to those of v_j by paths in C_{i,m_j} of minimal length and fill the resulting loop with a diagram mapping to C_{i,m_j}; this gives a diagram Δ'' with holes corresponding to the occurrences of $t^{\pm 1}$ in ℓ. Next, if the arc joining the termini of u_j and v_j is labelled ρ_i, then we insert a t-corridor into the hole associated to $\dots u_j t u_{j+1} \dots$, where the bottom of the t-corridor is labelled ρ_j. (If t is replaced by t^{-1}, the bottom of the corridor is the arc σ_{j+1} joining the initial vertex of u_{j+1} to that of v_{j+1}.) To complete the construction of Δ, one uses 2-cells in $C_{i,m_{j+1}}$ to fill the loop formed by the top of the t-corridor and σ_{j+1}. □

Corollary 14.3. *If Π is finitely generated and $\Gamma = \Pi \rtimes_\phi \mathbb{Z}$ is finitely presented then, up to $\simeq$ equivalence, the t-corridor length function of $\Pi \rtimes_\phi \mathbb{Z}$ depends only on the semidirect product (i.e. although it depends on the form of the finite presentation, it does not depend on the choice of $\mathcal{A}$ and $\hat{\phi}$).*

14.4. **Passing to Powers.** The purpose of this subsection is to prove Proposition 14.2.

Let $(\mathcal{A} \cup \{t\})^{\pm 1}$ be as above. Identifying $\Gamma_p = \Pi \rtimes_{\phi^p} \mathbb{Z}$ with the subgroup $\Pi \rtimes p\mathbb{Z}$ of Γ, we take generators $\mathcal{A} \cup \{\tau\}$ where $\tau = t^p$ in Γ. To each word $w \in (\mathcal{A}^{\pm} \cup \{t^{\pm 1}\})^*$ that equals $1 \in \Gamma$ we associate a word

[24] A familiar problem in this type of argument arises from degeneracies that threaten the planarity of the diagram; such problems are removed by surgery [13]. In the current setting these surgeries take place only in the regions between the t-corridors and therefore do not affect our discussion.

w_p in the free group on $\mathcal{A} \cup \{\tau\}$ according to the following scheme. First we draw a path on the integer lattice in $\mathbb{R}^2$ that begins at the origin and proceeds up one space as we read t, down one as we read t^{-1} and moves one space to the right as we read a letter from $\mathcal{A}^{\pm}$. We shall modify w by replacing certain open segments of this path that lie in the vertical intervals $[mp, (m+1)p]$; these segments are of two types, called *bumps* and *steps*.

If both endpoints of the subpath are at height mp and none of its edge are at height $(m+1)p$, then the segment is called an *up-bump*. If the initial endpoint is at height mp, the terminus at height $(m+1)p$ and all other vertices are at heights in $(mp, (m+1)p)$, then the segment is called an *up-step*. A *down-bump* and *down-step* are defined similarly.

When we have replaced all steps and bumps from the path defined by w, the horizontal segments of the resulting path will all run at heights divisible by p.

To this end, we write $w = u_1 v_1 u_2 v_2 \dots$ where u_1 is the first non-trivial prefix of w whose exponent sum in t is $0 \mod p$ and v_1 is the (possibly empty) subword before the next $t^{\pm 1}$, then u_2 is the first non-trivial prefix of w whose exponent sum in t is $0 \mod p$, and so on. Each u_i labels either a bump or a step.

If u_i labels a bump then we replace it by the reduced word $U_i \in F(A)$ that is equal in Γ to u_i. If $u_i = t^{\varepsilon} u_i'$, $\varepsilon = \pm 1$, is a step, then we replace it by the unique reduced word $t^{\varepsilon p} U_i$ with $U_i \in F(A)$ and $t^{\varepsilon} U_i = u_i$ in Γ.

Let $\tilde{w}_p \in (\mathcal{A}^{\pm} \cup \{t^{\pm 1}\})^*$ be the word obtained from w by the above process and let $w_p \in (\mathcal{A}^{\pm} \cup \{t^{\pm p}\})^*$ be the word obtained from $\tilde{w}_p$ by (starting from the left) replacing sub-words labelled $t^{\pm p}$ by $\tau^{\pm p}$ and then freely reducing.

As usual, in the following lemma $L = \max\{|\phi(a)| : a \in \mathcal{A}\}$.

Lemma 14.4. *$w = \tilde{w}_p = w_p$ in Γ and $|w_p| \le |\tilde{w}_p| \le L^{p-1}|w|$.*

Proof. The bound on $|\tilde{w}_p|$ comes from the following observation. For a bump labelled u_i, one can pass from u_i to U_i by deleting all letters $t^{\pm 1}$ from u_i and replacing each occurrence of $a \in \mathcal{A}$ in u_i, say $u_i = \alpha a \beta$, by the freely reduced word in $F(A)$ representing $\phi^r(a)$, where $-r$ is the exponent sum of t in α. Similarly, if a step is labelled $u_i = t^{\varepsilon} u_i'$, then U_i is obtained by deleting all t from u_i' and replacing each occurrence of $a \in \mathcal{A}$ in u_i, say $u_i' = \alpha a \beta$, by the freely reduced word in $F(A)$ representing $\phi^{\varepsilon(p-r)}(a)$, where εr is the exponent sum of t in α. $\square$

The replacement scheme described in the preceding proof corresponds to the construction of a singular-disc diagram $A(w)$ exhibiting the equality $w = \tilde{w}_p$ in Γ. Specifically, for each bump or step, one draws the vertical line joining each vertex to the height where it will be pushed, one labels it by the appropriate power of t, and then one fills-in the resulting line of rectangles with 2-cells whose boundary labels have the form $t^{-1}at\phi^{-1}(a)$. (Starting from this specific planar embedding one will in general have to flip some of the components of the interior in order to get an embedded diagram $A(w)$ with boundary cycle $\tilde{w}_p w_p^{-1}$.)

Lemma 14.5. *$A(w)$ is a union of t-corridors; each has at most one of its ends on the boundary arc labelled $\tilde{w}_p$, and the length of a t-corridor in $A(w)$ is at most $L^{p-1}\max|u_i|$, where the u_i are the sub-words of w labelling bumps and steps.*

Proof. The diagram $A(w)$ consists of a string of disc diagrams, one for each bump or step. A t-corridor in a disc corresponding to a bump labelled u_i has both of its ends on the arc labelled u_i, while a t-corridor in a disc corresponding to a step labelled tu_i' may have one end on the corresponding arc labelled t^p in $\tilde{w}_p$ and one on the arc labelled u_i' or (if the change in height along u_i' is not monotone) both ends on the arc labelled u_i'. In all cases, the label on the bottom side of the corridor is a concatenation of less than $|u_i|$ words of the form $\phi^r(a)$ with $a \in \mathcal{A}$ and $|r| \leq p-1$. □

Proof of Proposition 14.2. As we discussed immediately before subsection 5.1, the set of diagrams for Γ_p is, after p-refinement, a subset of the diagrams over Γ, and hence the corridor length function of the latter $\preceq$-dominates that of the former. (There are some constants to take account of here, such as a factor of p in length coming from the p-refinement, and an L^{p-1} needed to estimate the area of a t-corridor in terms of the corresponding τ-corridor, but these are trivial matters.) Thus the true content of the proposition is that the corridor length function of Γ is $\preceq$-bounded above by that of the Γ_p.

For each freely-reduced word $W \in (\mathcal{A}^{\pm} \cup \{t^{\pm p}\})^*$ that is null-homotopic in Γ_p we fix a van Kampen diagram $\Delta(W)$ whose τ-corridors have length at most $\Lambda(|W|)$. Then, for each freely-reduced $w \in (\mathcal{A}^{\pm} \cup \{t^{\pm 1}\})^*$ that is null-homotopic in Γ we define a van Kampen diagram $\Delta_p(w)$ as follows. First, we replace $\Delta(w_p)$ by its p-refinement (which has boundary label $\tilde{w}_p$). We then attach to this the singular-disc diagram $A(w)$ along the portion of its boundary labelled $\tilde{w}_p$.

We claim that the length of each t-corridor in $\Delta_p(w)$ is at most

$$L^{p-1}\,(2+\Lambda(L^{p-1}|w|)).$$

It follows from Lemma 14.5 that each of the t-corridors in $\Delta_p(w)$ is either contained in the annular diagram $A(w)$, or else is a layer in the p-refinement of a τ-corridor from $\Delta(w_p)$, possibly augmented on each end by a t-corridor in $A(w)$. (The fact that there are no t-corridors in $A(w)$ with both ends on the boundary arc labelled $\tilde{w}_p$ is crucial here.)

The length of a t-corridor in $A(w)$ is at most $L^{p-1}|v|$. The length of a τ-corridor from $\Delta(w_p)$ is at most $\Lambda(|w_p|) \le \Lambda(L^{p-1}|w|)$, and the length of each layer in its refinement is therefore at most $L^{p-1}\,\Lambda(L^{p-1}|w|)$. □

Appendix A. On a Result of Brinkmann

The following theorem is the main result in Peter Brinkmann's paper [11]. It plays a vital role in the first proof that the conjugacy problem is solvable for free-by-cyclic groups [5] (our Corollary B).

Theorem A.1. [11, Theorem 0.1] *Let $\phi : F \to F$ be an automorphism of a finitely generated free group. Then there exists a constant $K \ge 1$ such that for any pair of exponents N, i satisfying $0 \le i \le N$, the following two statements hold:*

(1) If w is a cyclic word in F, then

$$\|\phi^i(w)\| \le K\Big(\|w\| + \|\phi^N(w)\|\Big),$$

where $\|w\|$ is the length of the cyclic reduction of w with respect to some word metric on F.

(2) If w is a word in F, then

$$|\phi^i(w)| \le K\Big(|w| + |\phi^N(w)|\Big),$$

where $|w|$ is the word length of w.

The purpose of this appendix is to explain how to extract Theorem A.1 from our proof of Theorem A. We regard words and cyclic words in F_n as, respectively, based and unbased loops in the graph R with one vertex and n edges; the assertions of Theorem A.1 are then statements about how the lengths of the tightened images of such loops grow when one applies the obvious topological realisation $\overline{\phi}$ of ϕ. As in the previous subsection, these assertions will follow if we can establish the corresponding bounds with $\overline{\phi} : R \to R$ replaced by a topological (IRTT) representative $f : G \to G$ of a power of ϕ satisfying Assumption 5.7.

Remark A.2. *The proof given below shows that the constant K of Theorem 4.1 suffices for Theorem A.1. Brinkmann* [11] *states that (his constant) K can be computed effectively, but we do not see how to prove this. Indeed, given his approach (and ours), this assertion would seem to require an effective construction of an improved relative train track representative for ϕ, and a proof that such a construction exists does not seem to be available at the moment.*

The following lemma allows a proof of the assertions in (1) and (2) to be undertaken simultaneously.

Lemma A.3. *If σ is a nontrivial loop in G, then for some $j \geq 1$, the loop $f_\#^j(\sigma)$ admits a splitting at a vertex.*

Proof. According to [2, Lemma 4.1.2, p.554], σ admits a splitting $\sigma = \sigma_1$, where σ_1 is a path, but we argue further to arrange for this splitting to be at a vertex.

We divide the argument into a number of cases, depending on the largest i so that the stratum H_i contains an edge of σ_1. If this H_i is a zero stratum, $f_\#(\sigma_1) \subset G_{i-1}$ and an obvious induction applies. If H_i parabolic, then we apply [2, Lemma 4.1.4] to the circuit σ to obtain a splitting into paths, at least one of which is a basic path, and so has a vertex at one end. If H_i is an exponential stratum, then there is a positive integer K so that the number of i-illegal turns in $f_\#^k(\sigma_1)$ is the same for all $k \geq K$. In this case, since all Nielsen paths of exponential weight are edge-paths and all periodic paths are Nielsen, [2, Lemma 4.2.6] implies that $f_\#^K(\sigma_1)$ admits a splitting into sub-paths which are either r-legal or pre-Nielsen paths. If all sub-paths of $f_\#^K(\sigma_1)$ are pre-Nielsen paths, then $f_\#^{K+1}(\sigma_1)$ is a Nielsen path, and we ensured in [8, Section 1] that all Nielsen paths are edge-paths.

Suppose, then, that $f_\#^K(\sigma_1)$ contains an r-legal path ρ of weight r in its splitting. Then an iterate $f_\#^i(\rho)$ of ρ contains a displayed edge ε of weight r, and the path $f_\#^{K+i}(\sigma_1)$ splits immediately on either side of ε. Since σ has weight i, the splitting of $f_\#^{K+i}(\sigma_1)$ induces a splitting of $f_\#^{K+i}(\sigma)$ at a vertex, as required. □

In order to prove the statements (1) and (2), we analyze the van Kampen diagram Δ over the mapping torus of $f : G \to G$ that has boundary label $t^{-k}\sigma t^k f_\#^k(\sigma)^{-1}$. This is a simple stack of corridors as consider in Subsection 3.2.

In the restricted setting of stack diagrams, many of the difficulties that had to be overcome in the proof of Theorem A do not arise (there

are no singularities, for example), but there remain difficulties that one does not encounter in the context of positive automorphisms.

The number of edges in $\partial\Delta$ not labelled t is the quantity that determines the upper bound we seek, $n := |\sigma| + |f^N(\sigma)|)$. We must bound the length of each corridor in Δ linearly in terms of n. Theorem 4.1 provides a bound in terms of $|\partial\Delta|$, so we must argue is that in the context of stack diagrams, one can dispose of the contribution of the t-edges to this bound. In order to do so, we make an exhaustive list of those places in the proof of Theorem 4.1 where t-edges were accounted for, and we explain why, in each case, they are not required in the setting of simple stack diagrams.

(1) The t-edges contributed to the bound on the size of $S_0(2)$ and $S_0(3a)$ in Section 7, but these sets do not arise in stack diagrams.

(2) The t-edges were required in determining the sets $\mathrm{down}_1(\mathcal{T})$ used to bound the lengths of teams (see Definition 11.20). But $\mathrm{down}_1(\mathcal{T})$ was used only to bound the lengths of those teams whose narrow past had several components at some time in the past, and this cannot happen in a stack diagram.

(3) The t-edges entered the definition of $\mathrm{down}_2(\mathcal{T})$, which was used to bound the number of edges in $Q(\mathcal{T})$ before time $t_2(\mathcal{T})$ (see Definition 11.28). But there are no such edges in a stack of corridors, so we do not have to worry about double-counting, and an improved bound on the lengths of teams can be derived directly from the Pincer Lemma, noting that there are less than $2|\partial\Delta|$ adjacencies of colours.

(4) In the bonus scheme, the set ∂^e is used to bound the size of the interval of time $[\tau_0(e), \tau_2(e)]$, but in a stack of corridors it is clear that $\tau_0(e) = \tau_2(e)$, so the edges ∂^e are not required.

(5) Likewise, when bounding the size of the bonuses provided by rascals, we do not need to use the edges $\mathrm{down}_2(e)$ if our diagram is simply a stack of corridors

(6) A final use of t-edges is hidden in our references to [7] in the implementation of the Bonus scheme, specifically the bound on the sum of the lengths of blocks satisfying condition (iv) of the 'tautologous tetrad'. This is unnecessary in stack diagrams because there are no singularities and no edges that are cancelled by edges from outside the future of S_0, so the paths π_l and π_r travel forwards in time until they hit the boundary and $\sum |\mathrm{bdy}(\mathfrak{B})| < n$ bounds the size of the sum of all such blocks. □

References

[1] J. Alonso, Inégalités isopérimétriques et quasi-isométries, *C. R. Acad. Sci. Paris*, **311** (1990), 761-764.

[2] M. Bestvina, M. Feighn and M. Handel, The Tits alternative for $Out(F_n)$ I: Dynamics of exponentially growing automorphisms, *Ann. of Math. (2)*, **151** (2000), 517–623.
[3] M. Bestvina, M. Feighn and M. Handel, The Tits alternative for $\mathrm{Out}(F_n)$ II: A Kolchin type theorem, *Ann. of Math. (2)*, **161** (2005), 1–59.
[4] M. Bestvina and M. Handel, Train tracks and automorphisms of free groups, *Ann. of Math. (2)*, **135** (1992), 1–51.
[5] O. Bogopolski, A. Martino, O. Maslakova and E. Ventura, Free-by-cyclic groups have solvable conjugacy problem, preprint.
[6] M.R. Bridson, The geometry of the word problem, in *Invitations to geometry and topology* (M.R. Bridson and S.M. Salamon, eds), Oxford University Press, 2002.
[7] M.R. Bridson and D.P. Groves, The quadratic isoperimetric inequality for mapping tori of free group automorphisms I: Positive automorphisms, preprint at http://arxiv.org/math.GR/0211459.
[8] M. R. Bridson and D. Groves, Free-group automorphisms, train tracks, and the beaded decomposition, preprint at http://arxiv.org/math.GR/0507589.
[9] M.R. Bridson and D. Groves, The growth of conjugacy classes under free group automorphisms, in preparation.
[10] M.R. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, Springer-Verlag, Berlin, 1999.
[11] P. Brinkmann, Dynamics of free group automorphisms, preprint.
[12] D. Cooper, Automorphisms of free groups have finitely generated fixed point sets, *J. Algebra*, **111** (1987), 453–456.
[13] R.C. Lyndon and P.E. Schupp, Combinatorial group theory, Springer-Verlag, Berlin, 1977.
[14] A.Yu. Ol'shanskii and M.V. Sapir, Groups with small Dehn functions and bipartite chord diagrams, *GAFA*, to appear.
[15] S. Schleimer, Polynomial time word problems, preprint.

Martin R. Bridson, Mathematics Department, 180 Queen's Gate, London, SW7 2BZ, U.K.
E-mail address: m.bridson@ic.ac.uk

Daniel Groves, Department of Mathematics, California Institute of Technology, Pasadena, CA, 91125, USA
E-mail address: groves@caltech.edu