# Relative expansions for free group endomorphisms

Jean Pierre Mutanguha*

May 27, 2024

**Abstract**

This survey is based on a minicourse I taught at UC-Riverside in March 2023. The goal is to show the relation between the dynamics of a free group endomorphism and the geometry of its mapping torus. As a new result, we characterize the Dehn functions of ascending HNN extensions of free groups: they are linear, quadratic, or exponential.

The minicourse consisted of four lectures, and this survey's four sections respect that structure (for the most part). The main reference was my thesis [Mut21].

## Contents



* *Email:* jpmutang@princeton.edu, *Web address:* https://mutanguha.com
Department of Mathematics, Princeton University, Princeton, NJ, USA

# 1 Outline

## 1.1 Motivation: surface bundles over the circle

The story starts with Bill Thurston's hyperbolization theorem. Suppose $\Sigma$ is a closed surface with negative Euler characteristic $\chi(\Sigma)$, $F\colon \Sigma \to \Sigma$ a homeomorphism, and $f = [F]$ the isotopy class of $F$. The (topological) mapping torus $M_f = \Sigma \times [0,1]/{\sim_F}$ of $f$ is the quotient space defined using the equivalence relation $(s,1) \sim_F (F(s),0)$ for $s \in \Sigma$; the homeomorphism type of the closed 3-manifold $M_f$ does not depend on the representative $F \in f$ used in the equivalence relation. Thurston's remarkable theorem related the dynamics of the mapping class $f$ to the geometry of its mapping torus $M_f$:

**Theorem 1.1** (cf. [Thu82, Thm. 5.6])**.** *The following are equivalent:*

1. *$M_f$ admits an $\mathbb{H}^3$-structure, i.e. there is a homeomorphism from the universal cover of $M_f$ to $\mathbb{H}^3$ that conjugates the deck transformations to isometries;*
2. *$\pi_1(M_f)$ has no $\mathbb{Z}^2$-subgroup; and*
3. *$f$ has no periodic homotopy classes of essential closed curves in $\Sigma$, i.e. for all $n \geq 1$ and $\pi_1$-injective map $\sigma\colon \mathbb{S}^1 \to \Sigma$, $F^n \circ \sigma$ is not homotopic to $\sigma$.* □

The mapping class $f$ is atoroidal if it satisfies condition (3). The theorem gives an equivalence of: (1) a geometric property of $M_f$; (2) an algebraic property of $\pi_1(M_f)$; and (3) a dynamical property of $f$. The real content (and most difficult part) of this theorem is the implication (3⇒1). It requires a deeper understanding of the dynamics of atoroidal mapping classes. The following theorem is the initial step in this direction and forms part of the Nielsen–Thurston classification [Thu88, Thm. 4] for mapping classes:

**Theorem 1.2.** *The mapping class $f$ is atoroidal if and only if it contains a pseudo-Anosov homeomorphism of $\Sigma$.* □

We refer the reader to case (ii) of [Thu88, Thm. 4] for the definition of pseudo-Anosov homeomorphisms. For our purposes, we mention that a pseudo-Anosov homeomorphism quantifies the *growth* of homotopy classes of essential closed curves under iteration of its mapping class. This played a crucial role in the proof of Theorem 1.1. Some authors will call a mapping class pseudo-Anosov if it contains a pseudo-Anosov homeomorphism.

## 1.2 Motivation: free-by-cyclic groups

Bringing the story closer to our goal, Peter Brinkmann proved a (coarse) hyperbolization theorem analogous to Theorem 1.1. Suppose $\mathbb{F}$ is a finitely generated free group, $\Phi\colon \mathbb{F} \to \mathbb{F}$ an automorphism, and $\phi = [\Phi]$ the outer automorphism for $\Phi$. The (algebraic) mapping torus $\mathbb{F} \rtimes_\phi \mathbb{Z}$ of $\phi$ is the free-by-cyclic group given by the *relative presentation*:

$$\mathbb{F} \rtimes_\phi \mathbb{Z} = \langle \mathbb{F}, t \mid txt^{-1} = \Phi(x),\ \forall x \in \mathbb{F} \rangle;$$

as before, the isomorphism type of the free-by-cyclic group $\mathbb{F} \rtimes_\phi \mathbb{Z}$ does not depend on the representative $\Phi \in \phi$ used in the presentation. Brinkmann's theorem related the dynamics of an outer automorphism to the geometry of its mapping torus:

**Theorem 1.3** (cf. [Bri00, Thm. 1.2])**.** *The following are equivalent:*

1. $\mathbb{F} \rtimes_\phi \mathbb{Z}$ *is hyperbolic;*
2. $\mathbb{F} \rtimes_\phi \mathbb{Z}$ *has no* $\mathbb{Z}^2$*-subgroup; and*
3. $\phi$ *has no periodic conjugacy class of nontrivial elements in* $\mathbb{F}$*, i.e. for all* $n \geq 1$ *and nontrivial* $x \in \mathbb{F}$*,* $\Phi^n(x)$ *is not conjugate to* $x$ *in* $\mathbb{F}$*.* □

See Section 4 for a definition of hyperbolic groups. The outer automorphism $\phi$ is atoroidal if condition (3) holds. As with Thurston's theorem, Brinkmann's theorem gives an equivalence of geometric, algebraic, and dynamical properties and the hard part is the implication (3⇒1). Brinkmann's contribution was the following intermediate step:

**Theorem 1.4** (cf. [Bri00, Thm. 1.1])**.** *If* $\phi$ *is atoroidal, then there is a constant* $L \geq 1$ *such that* $2 \cdot \|\Phi^L(x)\| \leq \max(\|x\|, \|\Phi^{2L}(x)\|)$ *for all* $x \in \mathbb{F}$*.* □

Here, $\|x\|$ is the word-length of the shortest representative of the conjugacy class of $x$ in $\mathbb{F}$ with respect to some arbitrarily chosen basis of $\mathbb{F}$. An outer automorphism $\phi$ is hyperbolic if it satisfies the conclusion of the theorem; hyperbolic outer automorphisms play the same quantitative role as pseudo-Anosov homeomorphisms. Theorem 1.3 follows from a special case of Bestvina–Feighn's combination theorem [BF92]: hyperbolicity of an outer automorphism implies hyperbolicity of its mapping torus.

### 1.3 Objective: ascending HNN extensions of free groups

To conclude our story, we extended Brinkmann's theorem to the case when $\Phi \colon \mathbb{F} \to \mathbb{F}$ is an injective but not necessarily surjective endomorphism. The outer class $\phi = [\Phi]$ is the set of endomorphisms of $\mathbb{F}$ obtained by post-composing $\Phi$ with an inner automorphism of $\mathbb{F}$. The mapping torus $\mathbb{F}*_\phi$ of $\phi$ is the ascending HNN extension of $\mathbb{F}$ given by the same relative presentation as before:

$$\mathbb{F}*_\phi = \langle \mathbb{F}, t \mid txt^{-1} = \Phi(x),\ \forall x \in \mathbb{F} \rangle;$$

the only difference is that $\mathbb{F}$ is not a normal subgroup of the mapping torus if $\Phi$ is not surjective. If $\mathbb{F} = \mathbb{Z}$, then $\Phi$ is multiplication by a nonzero integer $d$ and its mapping torus, denoted $BS(1, d)$, is an example of a *Baumslag–Solitar group*. In this survey, we sketch an equivalence between the dynamics of $\phi$ and the geometry of $\mathbb{F}*_\phi$.

**Theorem 4.4** (cf. [Mut21, Thm. 5.2.7])**.** *The following are equivalent:*

1. *$\mathbb{F}*_\phi$ is hyperbolic;*

2. *$\mathbb{F}*_\phi$ has no $BS(1,d)$-subgroup for all $d \geq 1$; and*

3. *for all $n, d \geq 1$ and nontrivial $x \in \mathbb{F}$, $\Phi^n(x)$ is not conjugate to $x^d$ in $\mathbb{F}$.*

Condition (3) is stronger than the previous atoroidal condition. To prove the difficult implication (3⇒1), we will establish some quantitative properties — hyperbolicity of $\phi$ is one of them but is not sufficient when $\Phi$ is not surjective. The implication will then follow from Bestvina–Feighn's combination theorem.

Misha Gromov proved the equivalence between hyperbolicity of a finitely presented group and linearity of its Dehn function (see Section 4 for definitions). Bridson–Groves show that $\mathbb{F} \rtimes_\phi \mathbb{Z}$ has a linear or quadratic Dehn function [BG10]. Ilya Kapovich observed that the Dehn function of $\mathbb{F}*_\phi$ is exponential if it contains a $BS(1,d)$-subgroup for some $d \geq 2$ and asked for all the possible Dehn functions [Kap00, Cor. 5.7 & Prob. 6.5]. As a new result, we answer Kapovich's question using relative hyperbolicity:

**Corollary 4.9.** *$\mathbb{F}*_\phi$ has a linear, quadratic, or exponential Dehn function.*

**Acknowledgments.** I thank: Matthew Durham and Thomas Koberda for organizing the workshop and inviting me to give a minicourse; Kailey Perry for giving the introductory talk for the minicourse; and Jagerynn Verano for helpful comments on a preliminary version.

## 2 Expansions of free splittings

Throughout, $\mathbb{F}$ will denote a finitely generated nontrivial free group, $\Phi\colon \mathbb{F} \to \mathbb{F}$ an injective endomorphism, and $\phi = [\Phi]$ its outer class. The hyperbolization theorems necessitate a deeper understanding of the dynamics of $\phi$. When $\Phi$ is surjective, the theory of train tracks gives a robust framework for studying the dynamics of the outer automorphism $\phi$. For instance, Brinkmann used *improved relative train tracks* to prove Theorem 1.4.

We mostly care about the case when $\Phi$ is not surjective. Our current goal is to find optimal representatives that exhibit the dynamics of $\phi$. Rather than try to directly extend the theory of improved relative train tracks to the nonsurjective case, we will develop our optimal representatives "from scratch" using nonsurjectivity to our advantage. We will use Stallings maps, bounded cancellation, and free splittings.

### 2.1 Stallings maps

A graph is a 1-dimensional CW-complex and the volume $\text{vol}(\Gamma)$ of a finite graph $\Gamma$ is the number of edges in $\Gamma$. A core graph is a graph whose components are not contractible and have no proper deformation retracts. The core of a graph is the smallest deformation retract of the union of the graphs noncontractible components. A cellular map is a continuous function between graphs that maps vertices to vertices and is locally injective or constant on edges. An immersion is a cellular map that is locally injective everywhere. An expansion is an immersion $g\colon \Gamma \to \Gamma$ such that the combinatorial length of the edge-paths $g^n(e)$ is unbounded as $n \to \infty$ for every edge $e$ in $\Gamma$.

Let $\Gamma$ be a connected finite core graph, $\star \in \Gamma$ a chosen vertex (basepoint), and identify $\pi_1(\Gamma, \star) \cong \mathbb{F}$. Pick a nontrivial subgroup $A \le \mathbb{F}$, and let $c\colon (\widehat{\Gamma}, \hat{\star}) \to (\Gamma, \star)$ be the based cover corresponding to $A$. The *Stallings based map* $\hat{s}\colon (\widehat{S}, \hat{\star}) \to (\Gamma, \star)$ for $A$ over $(\Gamma, \star)$ is the restriction of $c$ to the smallest deformation retract of $(\widehat{\Gamma}, \hat{\star})$ containing $\hat{\star}$. The Stallings map $s\colon S \to \Gamma$ for the conjugacy class $[A]$ over $\Gamma$ is the restriction of $\hat{c}$ to the core of $\widehat{\Gamma}$.

For $n \ge 1$, let $s_n\colon S(\phi^n) \to \Gamma$ denote the Stallings map for $[\Phi^n(\mathbb{F})]$ over $\Gamma$.

**Example 2.1.** Suppose $\mathbb{F} = \mathbb{Z}$, $\mathbb{S}^1$ is a rose (i.e. a graph with one vertex) with one edge, and $\Phi$ is multiplication by $d \ne 0$. For $n \ge 1$, $s_n\colon S(\phi^n) \to \mathbb{S}^1$ is a cover with $|d|^n$ sheets; as a CW-complex, $S(\phi^n)$ has $|d|^n$ vertices and edges. Note that $\mathrm{vol}(S(\phi^n)) \to \infty$ as $n \to \infty$ if and only if $\Phi$ is not surjective.

When $\Phi$ is surjective, the Stallings maps $s_n$ are the identity map and not really useful for anything. On the other hand, the previous example illustrates that they get arbitrarily complicated (as measured with volume) when $\mathbb{F}$ is cyclic and $\Phi$ is not surjective. The following lemma shows that the dichotomy holds even when $\mathbb{F}$ is not cyclic:

**Lemma 2.1.** *$\Phi$ is not surjective if and only if $\mathrm{vol}(S(\phi^n)) \to \infty$ as $n \to \infty$.*

*Proof.* If $\Phi$ is surjective, then $s_n$ is the identity $\mathrm{id}\colon \Gamma \to \Gamma$ and $\mathrm{vol}(S(\phi^n)) = \mathrm{vol}(\Gamma)$ is constant in $n \ge 1$. Conversely, suppose there exists a constant $V \ge 1$ such that $\mathrm{vol}(S(\phi^n)) \le V$ for infinitely many $n \ge 1$. There are only finitely many combinatorially distinct immersions $\Gamma' \to \Gamma$ (mapping edges to edges) with $\mathrm{vol}(\Gamma') \le V$. So, for some $n' > n \ge 1$, there is a cellular isomorphism $\iota\colon S(\phi^{n'}) \to S(\phi^n)$ such that $s_{n'} = s_n \circ \iota$. This can be extended to an isomorphism $\hat{\iota}\colon \widehat{\Gamma}_{n'} \to \widehat{\Gamma}_n$ of covers of $\Gamma$. Correspondence between covers and conjugacy classes of subgroups implies $[\Phi^{n'}(\mathbb{F})] = [\Phi^n(\mathbb{F})]$. In other words, $x\Phi^n(\mathbb{F})x^{-1} = \Phi^{n'}(\mathbb{F}) \le \Phi^n(\mathbb{F})$ for some $x \in \mathbb{F}$. We leave it as an exercise that $x\Phi^n(\mathbb{F})x^{-1} = \Phi^n(\mathbb{F})$ and hence $\Phi$ is surjective. □

The exercise at the end of the proof is a special case of the following:

**Exercise 2.2.** *Let $\Psi\colon \mathbb{F} \to \mathbb{F}$ be an automorphism and $A \le \mathbb{F}$ a finitely generated subgroup. If $\psi(A) \le A$, then $\psi(A) = A$.*

*Hint.* Use *subgroup separability* (also known as the LERF-property or Hall's theorem): for any finitely generated subgroup $A \le \mathbb{F}$ and element $b \in \mathbb{F} \setminus A$, there is a finite index subgroup $A' \le \mathbb{F}$ that contains $A$ but not $b$. Moreover, subgroup separability can be proven using the Stallings based map for $A$ over a rose [Sta83, §6]. □

### 2.2 Bounded cancellation

Let $\Gamma$ be a connected finite core graph. A cellular map $g\colon \Gamma \to \Gamma$ represents $\phi$ if there is an isomorphism/marking $\pi_1(\Gamma) \cong \mathbb{F}$ such that $g$ induces $\phi$. The outer class $\phi$ is reducible if it can be represented by a cellular map with an invariant nonempty proper core subgraph; otherwise, it is irreducible.

**Example 2.2.** Suppose $\mathbb{F} = \langle a, b \rangle$ and $\Phi\colon \mathbb{F} \to \mathbb{F}$ maps $a \mapsto a$ and $b \mapsto ba^{-1}b^{-1}$. Let $R$ be the rose with two petals, and pick the marking $\pi_1(R) \cong \mathbb{F}$ that identifies the basis $\{a, b\} \subset \mathbb{F}$ with oriented petals of $R$. We will abuse notation and refer to the oriented petals by $a, b$ respectively (see Fig. 1); the petals with opposite orientations are $\bar{a}, \bar{b}$. The cellular map $f\colon R \to R$ that maps $a$ to $a$ and $b$ to the edge-path $b\bar{a}\bar{b}$ represents $\phi$. The oriented petal $a$ determines an $f$-invariant proper core subgroup of $R$, and so $\phi$ is reducible.

Now suppose $\Psi\colon \mathbb{F} \to \mathbb{F}$ maps $a \mapsto ab$ and $b \mapsto ba$, and let $\psi = [\Psi]$ be the outer class. It takes more work to show that an outer endomorphism is irreducible. As $\mathbb{F}$ has rank 2, any nonempty proper core subgraph of an $\mathbb{F}$-marked graph has circle components; therefore, to prove that $\psi$ is irreducible, it suffices to show that $\Psi^n(x)$ and $x^d$ are not conjugate in $\mathbb{F}$ for all $n, d \geq 1$ and nontrivial $x \in \mathbb{F}$. For the case $d = 1$, note that the cellular map $g\colon R \to R$ that maps $a$ to $ab$ and $b$ to $ba$ is an expansion representing $\psi$ (see Fig. 1). The case $d \geq 2$ will follow from Example 3.2 and the converse of Proposition 3.2.

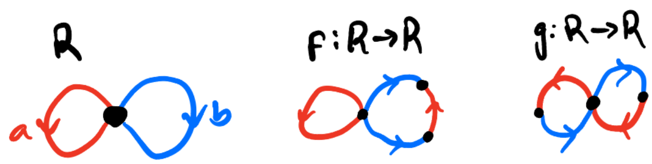


Figure 1: A marked rose and two cellular maps.

Patrick Reynolds' thesis [Rey11] shows that $\phi$ is (uniquely) represented by an expansion if it is irreducible and $\Phi$ is not surjective; we will give a new proof that is comparatively elementary as it only uses the classical bounded cancellation lemma and Stallings maps.

**Lemma 2.3** (bounded cancellation)**.** *If $f\colon \Gamma' \to \Gamma$ is a cellular map of finite graphs and $\tilde{f}\colon \widetilde{\Gamma}' \to \widetilde{\Gamma}$ a lift to the universal covers, then there is a constant $C \geq 0$ such that the $\tilde{f}$-image of the geodesic $[p, q]$ is in the $C$-neighborhood of $[\tilde{f}(p), \tilde{f}(q)]$ for all $p, q \in \widetilde{\Gamma}'$.*

The minimal such constant $C$ is the cancellation constant for the cellular map $f$ and denoted $C(f)$. The following proof is due to Bestvina–Feighn–Handel [BFH97, Lem. 3.1].

*Sketch of proof.* The map $f$ is a composition of edge collapses and subdivisions, $n$ *folds* ($n \geq 0$), and an immersion that maps edges to edges [Sta83, §3.3]. The lemma holds for: edge collapses and subdivision with $C = 0$; folds with $C = 1$; and immersions with $C = 0$. So the lemma holds for $f$ with $C = n$. □

A point in a graph is bivalent if it has a neighborhood homeomorphic to an open interval of $\mathbb{R}$. A point in a core graph is a branch point if it is not bivalent. Let $\Gamma$ be a core graph with no component homeomorphic to $\mathbb{S}^1$. A natural arc of $\Gamma$ is a maximal connected subset that contains no branch point. The natural structure on $\Gamma$ is the CW-complex whose vertices are the branch points and edges are the natural arcs (i.e. forget the bivalent vertices). A cellular map $g \colon \Gamma \to \Gamma$ is natural if it is a cellular map with respect to the natural structure, i.e. it maps branch points to branch points and is locally injective or constant near bivalent points.

**Theorem 2.4** (cf. [Rey11, Cor. 3.23])**.** *If $\phi$ is irreducible and $\Phi$ is not surjective, then some expansion $g \colon \Gamma \to \Gamma$ represents $\phi$.*

*Proof* (cf. [Mut21, Prop. 3.4.1])*.* Assume $\operatorname{rank}(\mathbb{F}) \geq 2$. Let $g_0 \colon \Gamma_0 \to \Gamma_0$ be a cellular map representing $\phi$. The map $g_0$ is $K$-Lispchitz (for some $K \geq 1$) with cancellation constant $C = C(g_0) \geq 0$. Let $S(\phi^n) \to \Gamma_0$ be the Stallings map for $[\Phi^n(\mathbb{F})]$ over $\Gamma_0$. The lift of $g_0$ to the cover $\widehat{\Gamma}_n$ corresponding to $[\Phi^n(\mathbb{F})]$ and the deformation retraction $\widehat{\Gamma}_n \to S(\phi^n)$ induce a $K$-Lipschitz map $g_n \colon S(\phi^n) \to S(\phi^n)$ with $C(g_n) \leq C$. By bounded cancellation, the map $g_n$ maps branch points to the $C$-neighborhood of branch points. After replacing $g_n$ with a homotopic map, we may assume it is a natural map with Lipschitz constant $K + C$ and cancellation constant $C(g_n) \leq 2C$.

By nonsurjectivity of $\Phi$ and Lemma 2.1, we can pick $N \gg 1$ such that $S(\phi^N)$ has a natural arc with more than $2C(K + C)^M$ edges, where $M = 3\operatorname{rank}(\mathbb{F}) - 4$. A natural arc $\alpha$ of $S(\phi^N)$ is *long* if $g_N^m(\alpha)$ covers a natural arc with more than $2C(K + C)^M$ edges for some $m \geq 0$; otherwise, it is *short*. Since $S(\phi^N)$ has at most $M + 1$ natural arcs and $g_N$ is $(K + C)$-Lipschitz, long natural arcs have more than $2C$ edges. The $g_N$-invariant short subgraph (i.e. closure of union of short natural arcs) is a proper subgraph. By irreducibility of $\phi$, the components of the short subgraph are contractible — note that $g_N$ represents $\phi$ via the isomorphism $\Phi^N \colon \mathbb{F} \to \Phi^N(\mathbb{F})$.

Let $\Gamma$ be the graph obtained by collapsing in $S(\phi^N)$ each component of the short subgraph and $g \colon \Gamma \to \Gamma$ the map induced by $g_N$ represents $\phi$. Since the natural arcs of $\Gamma$ have more than $2C \geq C(g_N) \geq C(g)$ edges, there are no folds in natural map $g$, and $C(g) = 0$. By injectivity of $\Phi$, the subgraphs on which $g$ is constant are contractible; after repeatedly collapsing these subgraphs, we may assume $g$ is an immersion. An edge $e$ of $\Gamma$ is *non-expanding* if its iterates $g^m(e)$ are edges for all $m \geq 1$. The $g$-invariant non-expanding subgraph (i.e. closure of the union of non-expanding edges) is proper as $\Phi$ is not surjective. By irreducibility of $\phi$ again, the components of the non-expanding subgraph are contractible. Collapse the non-expanding subgraph to ensure $g$ is an expansion. □

By bounded cancellation, an expansion representing $\phi$ is unique if it exists. Expansions do not always exist: expansions cannot represent outer automorphisms; more generally, an expansion represents $\phi$ if and only if there are no $\phi$-periodic (conjugacy classes of) nontrivial free factors of $\mathbb{F}$ — a precise statement is given in the next subsection.

### 2.3 Free splittings

We introduce our final tool: free splittings! A graph pair $(\Gamma, G)$ is a finite core graph $\Gamma$ with a proper core subgraph $G \subset \Gamma$ — we call $G$ the peripheral subgraph. For a graph pair $(\Gamma, G)$, edges of $G$ are peripheral and the remaining edges of $\Gamma$ are *nonperipheral*; the *relative length* of an edge-path in $\Gamma$ is the number of nonperipheral edges in the path.

A relative cellular map $(\Gamma, G) \to (\Gamma', G')$ is a cellular map $\Gamma \to \Gamma'$ that maps $G$ to $G'$. The *relative universal cover* of $(\Gamma, G)$ is the result of collapsing in $\widetilde{\Gamma}$ each component of $\widetilde{G}$, where $\widetilde{\Gamma}$ is the universal cover of $\Gamma$ and $\widetilde{G} \subset \widetilde{\Gamma}$ the lift of $G$. A relative immersion is a relative cellular map whose lifts to the universal covers induce immersions on the relative universal covers. Finally, a relative expansion is a relative immersion $g\colon (\Gamma, G) \to (\Gamma, G)$ such that the relative length of $g^n(e)$ is unbounded as $n \to \infty$ for every nonperipheral edge $e$ in $\Gamma$. When $G$ is empty, these relative notions are exactly the usual ones.

**Example 2.3.** Suppose $\mathbb{F} = \langle a, b\rangle$ and $\Phi\colon \mathbb{F} \to \mathbb{F}$ maps $a \mapsto a$ and $b \mapsto ba^{-1}b^{-1}$. Let $B$ be the barbell graph consisting of two (oriented) loops $l, r$ and a separating edge $m$, and $G \subset B$ the proper core subgraph determined by $l, r$ (see Fig. 2). Pick the marking that identifies $a, b$ with $l, mr\bar{m}$ respectively. The relative celular map $h\colon (B, G) \to (B, G)$ that maps $l \mapsto l$, $m \mapsto mr\bar{m}$, and $r \mapsto \bar{l}$ is a relative expansion representing $\phi$.

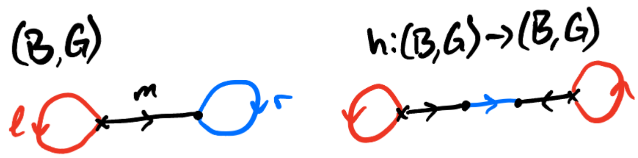


Figure 2: A marked barbell and a relative expansion.

For a relative cellular map $g\colon (\Gamma, G) \to (\Gamma, G)$, the $g$-stable part of $G$ is the union of components of $G$ that intersect $g^N(G)$, where $N$ is the number of components in $G$. The restriction of $g$ to $G$ is the peripheral restriction; it is an almost homotopy equivalence if its restriction to the $g$-stable part is a homotopy equivalence.

Let $\Gamma$ be a connected finite core graph. A free splitting (of $\mathbb{F}$) is a graph pair $(\Gamma, G)$ with a marking $\pi_1(\Gamma) \cong \mathbb{F}$. A nontrivial conjugacy class $[x]$ is peripheral in a free splitting $(\Gamma, G)$ if the immersed loop in $\Gamma$ representing $[x]$ is in $G$; otherwise, it is *nonperipheral*.

*Remark.* In [Mut21], a free splitting is defined as the relative universal cover of a marked graph pair. Relative cellular maps are then the lifts of cellular maps of pairs to the relative universal covers; relative immersions/expansions are defined on the relative covers as well. The advantage of this point of view is that relative immersions/expansions are honest immersions/expansions, which makes proofs clearer. The downside is that relative universal covers are infinite objects and hard to illustrate.

If $\Phi$ is surjective, then no relative expansion represents $\phi$. The converse, a generalization of Reynold's Theorem 2.4, is the heart of [Mut21] and this survey:

**Theorem 2.5** (cf. [Mut21, Thm. 3.4.5])**.** *If $\Phi$ is not surjective, then $\phi$ is represented by a relative expansion $g\colon (\Gamma, G) \to (\Gamma, G)$ whose peripheral restriction is an almost homotopy equivalence.*

*Sketch of proof.* Construct a $(K+C)$-Lipschitz natural map $g_N\colon S(\phi^N) \to S(\phi^N)$ with cancellation constant $C(g_N) \le 2C$ and a $g_N$-invariant short proper subgraph $G_N \subset S(\phi^N)$ as in our proof of Theorem 2.4. This time, $\Gamma$ is obtained by collapsing in $S(\phi^N)$ each contractible component of $G_N$ and deformation retracting the noncontractible components to their cores. Let $G \subset \Gamma$ be the image of $\text{core}(G_N)$, and call this process "relative collapsing $G_N$" in $S(\phi^N)$. Then $(\Gamma, G)$ is a free splitting, and the induced relative natural map $g\colon (\Gamma, G) \to (\Gamma, G)$ represents $\phi$ with *relative cancellation constant* $C_G(g) = 0$, where $C_G(g)$ is the cancellation constant of the induced maps on the relative universal cover.

For convenience, forget the bivalent vertices in $\Gamma$. In the base case, the peripheral restriction of $g$ is an almost homotopy equivalence. By injectivity of $\Phi$, $\text{core}(G') = \text{core}(G)$ for subgraphs $G' \supset G$ on which $g$ is *relatively constant.* After repeatedly relatively collapsing these subgraphs, we may assume $g$ is a relative immersion. As $\Phi$ is not surjective, some edge must be *relatively expanding.* Relatively collapse the non-relatively-expanding edges to ensure $g$ is a relative expansion (whose peripheral restriction is an almost homotopy equivalence). This concludes the base case.

By induction, we may suppose the peripheral restriction $g_G$ of $g$ is a relative expansion on a graph pair $(G, H)$ whose peripheral restriction is an almost homotopy equivalence. For a simplification of the remaining steps, you may pretend $H$ is empty — the steps in the general relative setting are nearly identical. There are two cases to consider: either some $g^m(\Gamma) \subset G$ for some $m \ge 1$; or $g$ is a relative immersion (after repeatedly relatively collapse subgraphs on which $g$ is relatively constant if necessary).
*Case 1*: the Stallings map for $[\Phi^m(\mathbb{F})]$ over $(G, H)$ determines a free splitting with a relative expansion representing $\phi$ whose peripheral restriction is an almost homotopy equivalence.
*Case 2*: use the relative expansion $g_G$ and bounded cancellation to promote $g$ to a relative expansion whose peripheral restriction is an almost homotopy equivalence. □

If a free splitting $(\Gamma, G)$ admits a relative expansion $g$ whose peripheral restriction is an almost homotopy equivalence, then the $g$-stable part of $G$ determines the unique (up to conjugacy) free factor system consisting of maximal $\phi$-periodic nontrivial free factors of $\mathbb{F}$; thus any two such free splittings determine the same peripheral conjugacy classes in $\mathbb{F}$. By bounded cancellation, there is an essentially unique relative expansion representing $\phi$ whose peripheral restriction is an almost homotopy equivalence [Mut21, Cor. 3.4.7] (when $\Phi$ is not surjective), and we will call it the canonical relative expansion for $\phi$.

## 2.4 Side quest: stable images

Most of this survey focuses on the outer class $\phi$ of the injective endomorphism $\Phi\colon \mathbb{F} \to \mathbb{F}$. Let us take a brief excursion into studying $\Phi$.

The stable image of $\Phi$ is the intersection $\Phi^\infty(\mathbb{F}) = \bigcap_{n\geq 1} \Phi^n(\mathbb{F})$. The stable image is $\Phi$-fixed (setwise), i.e. $\Phi(\Phi^\infty(\mathbb{F})) = \Phi^\infty(\mathbb{F})$; in particular, the restriction of $\Phi$ to the stable image is an automorphism, and Edward Turner proved the stable image is a free factor of $\mathbb{F}$ [Tur96, Thm. 1] (see also proof of Corollary 2.8). It can be used to reduce questions about injective endomorphisms to the corresponding questions about automorphisms. For algorithmic reductions, we would need to compute the stable image.

Let $(\Gamma, \star)$ be a connected finite graph $\Gamma$ with a chosen vertex (basepoint) $\star \in \Gamma$; it is a *pointed core graph* if there is no proper deformation retract containing the basepoint. A *based cellular map* $g\colon (\Gamma, \star) \to (\Gamma, \star)$ is a cellular map that fixes the basepoint $\star$; it *represents* $\Phi$ if there is a marking $\pi_1(\Gamma, \star) \cong \mathbb{F}$ such that $g$ induces $\Phi$. One can similarly define *pointed free splittings* $(\Gamma, G, \star)$ and *based relative expansions* $g\colon (\Gamma, G, \star) \to (\Gamma, G, \star)$.

**Example 2.4.** Suppose $\mathbb{F} = \langle a, b\rangle$ and $\Phi'\colon \mathbb{F} \to \mathbb{F}$ maps $a \mapsto b^{-2}ab^2$ and $b \mapsto b^{-1}a^{-1}b$. Let $B$ be the marked barbell from Example 2.3, and subdivide the separating (oriented) edge $m$ to get two edges $m_1, m_2$ and $m = m_1m_2$. Construct a pointed core graph $(B', 0)$ by identifying $1 \in [0, 1]$ with the midpoint of $m$ (see Fig. 3). The deformation retraction $B' \to B$ will induce a marking $\pi_1(B', 0) \cong \mathbb{F}$. The relative expansion $h$ from Example 2.3 extends to a based relative expansion $h'$ on $(B', G, 0)$ that represents $\Phi'$. As $0 \notin G$, the stable image of $\Phi'$ is trivial. On the other hand, the endomorphism $\Phi$ from Example 2.3 has the stable image $\langle a\rangle$. The endomorphisms $\Phi, \Phi'$ represent the same outer class $\phi$.

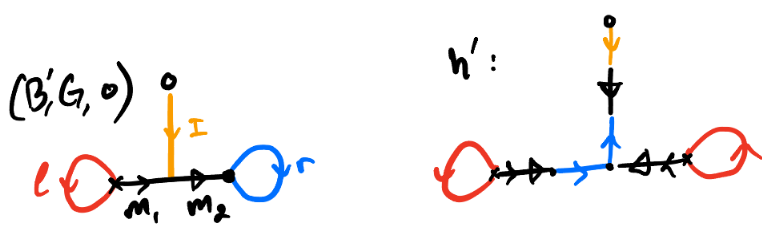


Figure 3: A marked pointed core graph and a based relative expansion.

We now show that relative expansions always extend to based relative expansions:

**Theorem 2.6.** *If $\Phi$ is not surjective, then it is represented by a based relative expansion $g\colon (\Gamma, G, \star) \to (\Gamma, G, \star)$ whose peripheral restriction is an almost homotopy equivalence.*

By bounded cancellation again, we call this the *canonical based relative expansion* for $\Phi$.

*Sketch of proof.* Let $g'\colon (\Gamma', G) \to (\Gamma', G)$ be the canonical relative expansion for $\phi$. Construct $\Gamma_0$ by identify $1 \in [0, 1]$ with an arbitrary point in $\Gamma'$ and choose the basepoint

$\star_0 = 0 \in [0,1] \subset \Gamma_0$. We can now naively extend the marking $\pi_1(\Gamma') \cong \mathbb{F}$ to $\pi_1(\Gamma_0, \star_0) \cong \mathbb{F}$ and the relative expansion $g'$ to a based relative cellular map $g_0 \colon (\Gamma_0, G, \star_0) \to (\Gamma_0, G, \star_0)$ that represents $\Phi$.

Let $S(\Phi^n)_\star \to (\Gamma_0, \star_0)$ be the Stallings based map for $\Phi^n(\mathbb{F})$ over $(\Gamma_0, \star_0)$. Analogous to the proof of Theorem 2.4, construct a Lipschitz based natural map $g_N \colon S(\Phi^N)_\star \to S(\Phi^N)_\star$ with cancellation constant $C(g_N) \leq 2C$ and a $g_N$-invariant relatively short proper subgraph $G_N \subset S(\Phi^N)$. One can then check that relatively collapsing $G_N$ should produce the required based relative expansion $g \colon (\Gamma, G, \star) \to (\Gamma, G, \star)$. In essence, this proof shows that we could have carefully extended:

- the core graph $\Gamma'$ to a pointed core graph $(\Gamma, \star)$ with $\mathrm{core}(\Gamma) = \Gamma'$;
- the marking $\pi_1(\Gamma') \cong \mathbb{F}$ to $\pi_1(\Gamma, \star) \cong \mathbb{F}$; and
- the relative expansion $g'$ to a based relative expansion $g$ on $(\Gamma, G, \star)$. □

Note that the proofs of Theorems 2.4 to 2.6 are constructive:

**Corollary 2.7.** *There is an algorithm that finds the canonical based relative expansion for nonsurjective* $\Phi$. □

**Corollary 2.8.** *There is an algorithm that finds a basis for the stable image* $\Phi^\infty(\mathbb{F})$.

*Proof.* If $\Phi$ is an automorphism, then $\Phi^\infty(\mathbb{F}) = \mathbb{F}$, and any basis of $\mathbb{F}$ will do. Otherwise, compute the canonical based relative expansion on $(\Gamma, G, \star)$ for $\Phi$. If the basepoint $\star$ is not in $G$, then $\Phi^\infty(\mathbb{F})$ is trivial, and the algorithm returns an empty set. Otherwise, $\Phi^\infty(\mathbb{F})$ is identified with $\pi_1(G, \star)$ via the marking $\pi_1(\Gamma, \star) \cong \mathbb{F}$. In particular, $\Phi^\infty(\mathbb{F})$ is a free factor of $\mathbb{F}$, and the algorithm returns any basis for $\pi_1(G, \star) \leq \pi_1(\Gamma, \star)$. □

# 3 Pullbacks in free groups

In this section, we will introduce a second condition on $\phi$ needed for hyperbolization.

## 3.1 Pullbacks of immersions

For a finitely generated nontrivial subgroup $A \leq \mathbb{F}$, $\mathrm{rr}(A) = \mathrm{rank}(A) - 1$ is the reduced rank of $A$. Hanna Neumann proved that a nontrivial intersection of two finitely generated subgroups $A, B \leq \mathbb{F}$ satisfies $\mathrm{rr}(A \cap B) \leq 2\,\mathrm{rr}(A)\,\mathrm{rr}(B)$ (and conjectured that the "2" in the upper bound could be removed) [Neu57].

Let $g_i \colon \Gamma_i \to \Gamma$ be immersions of core graphs for $i = 1, 2$. The pullback of $(g_1, g_2)$ is the immersion map $g \colon \Gamma_1 \times_\Gamma \Gamma_2 \to \Gamma$, where $\Gamma_1 \times_\Gamma \Gamma_2$ is the core of the graph:

$$\Gamma_{12} = \{(x_1, x_2) \in \Gamma_1 \times \Gamma_2 \ : \ g_1(x_1) = g_2(x_2)\};$$

and $g \colon (x_1, x_2) \mapsto g_i(x_i)$ — the subspace $\Gamma_{12} \subset \Gamma_1 \times \Gamma_2$ is homeomorphic to a graph, and we endow it with a CW-structure so that $g$ maps edges to edges.

**Example 3.1.** Let $\Gamma_1 = \Gamma_2 = \mathbb{S}^1$ be the rose with one edge and $g_i \colon \Gamma_i \to \mathbb{S}^1$ the cover with $d_i$ sheets. The pullback of $(g_1, g_2)$ is the disjoint union of $\gcd(d_1, d_2)$ connected covers of $\mathbb{S}^1$, each with $\operatorname{lcm}(d_1, d_2)$ sheets.

Suppose $\Gamma$ is a core graph with $\pi_1(\Gamma) \cong \mathbb{F}$ and $A_1, A_2 \leq \mathbb{F}$ are finitely generated nontrivial subgroups. Let $s_i \colon S_i \to \Gamma$ $(i = 1, 2)$ be the Stallings map for $[A_i]$ over $\Gamma$. Stallings observed that pullbacks represent intersections [Sta83, Thm. 5.5]: if $A_1 \cap A_2$ is not trivial, then some component of the pullback of $(s_1, s_2)$ is the Stallings map for $[A_1 \cap A_2]$ over $\Gamma$. Gersten used this observation and an Euler characteristic argument to give a topological proof of Neumann's inequality [Sta83, §7.7].

If a component of the pullback represents an intersection, what about the rest of the pullback? Let $A_1 \backslash \mathbb{F} / A_2$ be the set of *double cosets* for $(A_1, A_2)$ and $\mathcal{O}(A_1, A_2)$ be the subset consisting of double cosets $A_1 x A_2$ such that $A_1 \cap x A_2 x^{-1}$ is not trivial. There is a bijective correspondence between $\mathcal{O}(A_1, A_2)$ and the components of $s \colon S_1 \times_\Gamma S_2 \to \Gamma$, where each component is the Stallings map for $[A_1 \cap x A_2 x^{-1}]$ with $A_1 x A_2 \in \mathcal{O}(A_1, A_2)$. Extending Gersten's observation, Walter Neumann proved $-\chi(S_1 \times_\Gamma S_2) \leq 2\chi(S_1)\chi(S_2)$ and conjectured that the "2" could be removed [Neu90]; also known as the "strengthened Hanna Neumann conjecture", this was independently proven by Igor Mineyev [Min12] and Joel Friedman [Fri15], but it is not needed for our purposes.

## 3.2 Pullback stability

Pick a connected finite core graph $\Gamma$ with $\pi_1(\Gamma) \cong \mathbb{F}$ and let $s_n \colon S(\phi^n) \to \Gamma$ be the Stallings maps for $[\Phi^n(\mathbb{F})]$ over $\Gamma$ for $n \geq 1$. Set $\Lambda_n = S(\phi^n) \times_\Gamma S(\phi^n)$ and $\mathcal{O}_n = \mathcal{O}(\Phi^n(\mathbb{F}), \Phi^n(\mathbb{F}))$. By W. Neumann's inequality, $-\chi(\Lambda_n) \leq 2 \operatorname{rr}(\mathbb{F})^2$. Denote the pullback of $(s_n, s_n)$ by $\varphi_n \colon \Lambda_n \to \Gamma$. As $s_{n+1}$ factors through $s_n$, we have $\varphi_{n+1} = \varphi_n \circ \psi_{n+1}$ for some immersion $\psi_{n+1} \colon \Lambda_{n+1} \to \Lambda_n$.

**Example 3.2.** Let $\mathbb{F} = \langle a, b \rangle$, $\Psi \colon \mathbb{F} \to \mathbb{F}$ be given by $a \mapsto ab$ and $b \mapsto ba$, $\psi = [\Psi]$ the outer class, and $R$ the rose with the marking $\pi_1(R) \cong \mathbb{F}$ identifying $a, b$ with oriented petals of $R$. Then the pullback $\Lambda_n = S(\psi^n) \times_R S(\psi^n) \to R$ is a disjoint union of immersions of a rose and two circles (see Fig. 4 for $n = 1, 2$). The core graph $\Lambda_n$ is homeomorphic to $\Lambda_m$ for all $m, n \geq 1$. Although the CW-structure of $\Lambda_n$ varies with $n \geq 1$, the topology does not.

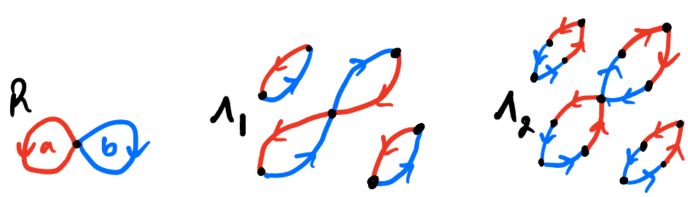


Figure 4: A marked rose and two pullbacks.

Let $\widehat{\mathcal{O}}_n \subset \mathcal{O}_n$ consist of the double cosets $\Phi^n(\mathbb{F})x\Phi^n(\mathbb{F})$ such that $x \notin \Phi(\mathbb{F})$ and $\widehat{\Lambda}_n \subset \Lambda_n$ be the union of components corresponding to $\widehat{\mathcal{O}}_n$. We can now make a few of observations: the immersion $\psi_{n+1}$ restricts to an immersion $\widehat{\Lambda}_{n+1} \to \widehat{\Lambda}_n$; if $\widehat{\Lambda}_n$ is empty, then so is $\widehat{\Lambda}_{n+1}$; and if $\widehat{\Lambda}_n$ is a (possibly empty) disjoint union of circles, then so is $\widehat{\Lambda}_{n+1}$. By W. Neumann's inequality, we conclude:

**Lemma 3.1** (cf. [Mut21, Lem. 5.1.4])**.** *For $n \geq 2\,\mathrm{rr}(\mathbb{F})^2$, components of $\widehat{\Lambda}_n$ are circles.* □

We say "pullbacks for $\phi$ stabilize" if some $\widehat{\Lambda}_n$ is empty — this is the second condition needed for hyperbolization! The next proposition gives a (necessary and) sufficient condition for pullbacks to stabilize:

**Proposition 3.2** (cf. [Mut21, Prop. 5.1.5])**.** *If pullbacks for $\phi$ do not stabilize, then $\Phi^m(x)$ and $x^d$ are conjugate in $\mathbb{F}$ for some $m, d \geq 2$ and nontrivial $x \in \mathbb{F}$.*

One can show directly that the converse holds, and we leave that as an elementary (but challenging) exercise; alternatively, the converse is Theorem 4.8(5⇒4) below.

*Sketch of proof.* Suppose $g\colon \Gamma \to \Gamma$ represents $\phi$ and pullbacks for $\phi$ do not stabilize, i.e. $\widehat{\Lambda}_n \neq \emptyset$ for all $n \geq 1$. Then $\Phi$ is not surjective and $\widehat{\Lambda}_n$ has circle components for $n \geq 2\,\mathrm{rr}(\mathbb{F})^2$; therefore, a component of $\widehat{\Lambda}_n$ is a pair of immersed loops $(\sigma, \sigma')$ in $\Gamma$. Pick $r \geq 2\,\mathrm{rr}(\mathbb{F})^2$, and consider the immersions $\psi_n \cdots \psi_{r+1}\colon \widehat{\Lambda}_n \to \widehat{\Lambda}_r$ for $n > r$. By the pigeonhole principle, we can choose components $(\sigma_n, \sigma_n')$ of $\widehat{\Lambda}_n$ for $n \geq r$ such that $(\sigma_r, \sigma_r')$ is in the image of a component $(\sigma_n, \sigma_n')$ for $n > r$; in particular, the loop $g^{n-r}(\sigma_n)$ is homotopic to a power of $\sigma_r$ for $n > r$.

Let $g$ be an expansion. The following argument is due to Kapovich [Kap00, Prop. 3.7]. Note that $g^{n-r}(\sigma_n)$ wraps around $\sigma_r$ for $n > r$. Since $g$ is an expansion, we can pick $N \gg r$ so that $g^{N-r}(e)$ is longer than a large power of $\sigma_r$, say $\sigma_r^{100}$ for all edges $e$ in $\Gamma$. By the pigeonhole principle again, some edge $e$ is contained in loops $\sigma_n, \sigma_{n+m}$, where $N \leq n < n+m \leq N + \mathrm{vol}(\Gamma)$. Since $g$ is an immersion, $g^{n-r}(e)$ covers $\sigma_r^{100}$, and $g^m(\sigma_r^{100})$ wraps around $\sigma_r$. If the arbitrarily large power $\sigma_r^{100}$ was chosen properly in terms of $\sigma_r$, then a final pigeonhole principle argument implies $g^m(\sigma_r)$ is homotopic to a power $\sigma_r^d$ with $|d| \geq 2$. Thus $\sigma_r$ represents a nontrivial conjugacy class $[x]$ in $\mathbb{F}$ such that $[\Phi^m(x)] = [x^d]$. Double $m$ and square $d$ to ensure $m, d \geq 2$.

In general, let $g$ be the canonical relative expansion. As the peripheral restriction of $g$ is an almost homotopy equivalence, the immersed loops $\sigma_n$ ($n \geq r$) are not peripheral. So they contain nonperipheral edges and a relative version of Kapovich's argument applies. □

This proof invokes the axiom of (countable) choice; however, a careful application of the pigeonhole principle can strengthen the proposition and make the proof constructive.

**Proposition 3.3.** *For a computable positive integer $N(\phi)$, if $\widehat{\Lambda}_{N(\phi)}$ is not empty, then $\Phi^m(x)$ and $x^d$ are conjugate in $\mathbb{F}$ for some $m, d \geq 2$ and nontrivial $x \in \mathbb{F}$.* □

As a corollary, pullbacks for $\phi$ stabilize if and only if $\widehat{\Lambda}_{N(\phi)}$ is empty.

# 4 Geometry of mapping tori

We are ready to discuss the geometry of $\mathbb{F}*_\phi$. For $\delta \geq 0$, a geodesic space is *$\delta$-hyperbolic* if any geodesic triangle is in the $\delta$-neighborhood of any pair of its sides; e.g. the real hyperbolic $n$-spaces $\mathbb{H}^n$ ($n \geq 2$) are $\ln(1+\sqrt{2})$-hyperbolic. A finitely generated group is hyperbolic if its Cayley graph with respect to a finite generating set is $\delta$-hyperbolic for some $\delta \geq 0$.

Now suppose $G$ is the fundamental group of a finite 2-dimensional CW-complex $X$ and the 1-skeleton of the universal cover $\widetilde{X}$ is not contractible. A *simple loop* in $\widetilde{X}$ is an embedding $\gamma \colon \mathbb{S}^1 \to \widetilde{X}$ that maps 1-cells to 1-cells, where $\mathbb{S}^1$ is a circle with some CW-structure; the *length* of $\gamma$, $\ell(\gamma)$, is the combinatorial length of $\mathbb{S}^1$. A *(reduced) disk diagram* for $\gamma$ is a continuous extension $\delta \colon \mathbb{D} \to X$ of $\gamma$ that homeomorphically maps $n$-cells to $n$-cells ($n = 0, 1, 2$), where $\mathbb{D}$ is the closed disk with some CW-structure and $\partial\mathbb{D} = \mathbb{S}^1$; the *area* of $\gamma$, $\text{area}(\gamma)$, is the minimal number of 2-cells in $\mathbb{D}$ over all disk diagrams for $\gamma$. The Dehn function of $G$ is the function $\mathbb{N} \to \mathbb{N}$ given by:

$$n \mapsto \max\{\text{area}\,\gamma : \gamma \text{ is a simple loop in } \widetilde{X} \text{ with } \ell(\gamma) \leq n\}.$$

We say $G$ has a linear (quadratic, exponential resp.) Dehn function if its Dehn function is bounded above and below by linear (quadratic, exponential resp.) functions; $G$ has a linear Dehn function if and only if it is hyperbolic.

Here is the main observation needed to complete the hyperbolization theorem: hyperbolicity and pullback stability for $\phi$ imply hyperbolicity of $\mathbb{F}*_\phi$; more generally, pullbacks for $\phi$ stabilize if and only if $\mathbb{F}*_\phi$ has a linear or quadratic Dehn function. These observations will follow from Bestvina–Feighn's combination theorem and its relativization.

## 4.1 Bestvina–Feighn's combination theorem

Our exposition on Bestvina–Feighn's combination theorem has been simplified and specialized to the mapping torus setting. The reader will find the most general version in [BF92].

We already used canonical relative expansions to give a sufficient condition for pullback stability. We can use them again to extend Theorem 1.4:

**Proposition 4.1** (cf. [Mut21, Prop. 5.2.6])**.** *If $\phi$ is atoroidal, then $\phi$ is hyperbolic.*

The converse is a nice exercise.

*Sketch of proof.* Theorem 1.4 covers the case when $\Phi$ is surjective. So we may assume $\Phi$ is not surjective. Let $g \colon (\Gamma, G) \to (\Gamma, G)$ be the canonical relative expansion for $\phi$. There is a constant $L_0 \geq 1$ such that $2 \cdot \|x\|_G \leq \|\Phi^{L_0}(x)\|_G$ for all nonperipheral $x \in \mathbb{F}$, where $\|x\|_G$ is the relative length of the immersed loop in $\Gamma$ representing $[x]$. The peripheral restriction is an almost homotopy equivalence and, by Theorem 1.4 again, there is constant $L_1 \geq 1$ such that $2 \cdot \|\Phi^{L_1}(x)\| \leq \max(\|x\|, \|\Phi^{2L_1}(x)\|)$ for all peripheral $x \in \mathbb{F}$ since $\phi$ is atoroidal. With care, we can deduce that $\phi$ is hyperbolic. $\square$

Let $\Gamma$ be a connected finite core graph with marking $\pi_1(\Gamma) \cong \mathbb{F}$ and $g \colon \Gamma \to \Gamma$ a cellular map representing $\phi$. The topological mapping torus $M_g$ of $g$ has an induced marking $\pi_1(M_g) \cong \mathbb{F}*_\phi$. Consider the universal cover $\widetilde{M}_g$ of $M_g$ and equivariantly collapse the $\widetilde{\Gamma}$-cross-sections to get the *Bass–Serre tree* $T_\phi$ for $\mathbb{F}*_\phi$. Orient the axis for the *stable letter* $t \in \mathbb{F}*_\phi$ in $T_\phi$ so that $t$ acts positively on it; the induced orientations on its edges equivariantly extend to an orientation on $T_\phi$. Note that every vertex of $T_\phi$ has exactly one incoming edge with respect to this *$t$-orientation* (see Fig. 5.). Along the axis for $t$, there is a unique vertex $\star$ whose stabilizer is $\mathbb{F} \le \mathbb{F}*_\phi$. When $\Phi$ is surjective, $T_\phi$ is a line.

For $L \ge 1$, a conjugacy in $(\mathbb{F}, \phi)$ of length $2L$ is a pair $([x], [\rho_x])$ of a conjugacy class of some $x \in \mathbb{F}*_\phi$ and an orbit in $T_\phi$ of some immersed edge-path $\rho_x$ of length $2L$ fixed (pointwise) by $x$ — there is a natural one-to-one correspondence between conjugating $x$ and translating $\rho_x$. The conjugacy is unidirectional if $\rho_x$ is monotone (with respect to the $t$-orientation on $T_\phi$); it is strictly bidirectional if $\rho_x$ decomposes into two monotone pieces of lengths $L$ (see Fig. 5). When $\Phi$ is surjective, all conjugacies in $(\mathbb{F}, \phi)$ are unidirectional. Let $(v_i)_{i=-L}^{L}$ be the consecutive vertices of $\rho_x$, $x_i \in \mathbb{F}$ the conjugate of $x$ corresponding to translating $v_i$ to $\star$, and $\|x_i\|$ the length of the conjugacy class of $x_i$ in $\mathbb{F}$. A conjugacy $([x], [\rho_x])$ flares if $2 \cdot \|x_0\| \le \max(\|x_{-L}\|, \|x_L\|)$. We say "$(\mathbb{F}, \phi)$ satisfies the conjugacies flare condition" if there is an $L \ge 1$ such that all conjugacies in $(\mathbb{F}, \phi)$ of length $2L$ flare.

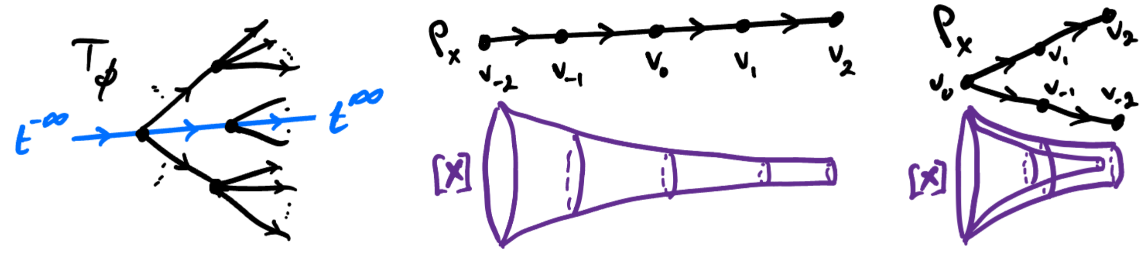


Figure 5: A Bass–Serre tree and two conjugacies of length 4: flaring unidirectional and non-flaring strictly bidirectional. The class $[x]$ is represented by some map $\mathbb{S}^1 \times [-2, 2] \to M_g$.

**Proposition 4.2.** *$\phi$ is hyperbolic and its pullbacks stabilize if and only if $(\mathbb{F}, \phi)$ satisfies the conjugacies flare condition.*

*Sketch of proof.* Any conjugacy is a concatenation of a unidirectional conjugacy with a strictly bidirectional conjugacy. Hyperbolicity of $\phi$ is equivalent to the flaring of unidirectional conjugacies. Pullback stability for $\phi$ is equivalent to a uniform bound on the length of strictly bidirectional conjugacies. Together, they imply conjugacies flare in $(\mathbb{F}, \phi)$ — sufficiently long conjugacies will be almost unidirectional. Conversely, if pullbacks for $\phi$ do not stabilize, then, by the proof of Proposition 3.2, there are integers $m, d_\pm \ge 2$ and a strictly birectional conjugacy $([x], [\rho_x])$ with length $2r \gg 2m$ such that $[\Phi^m(x_{\pm r})] = [x_{\pm r}^{d_\pm}]$ in $\mathbb{F}$; this conjugacy does not flare (see Fig. 5). $\square$

**Theorem 4.3** (combination [BF92]). *If $(\mathbb{F}, \phi)$ satisfies the conjugacies flare condition, then $\mathbb{F}*_\phi$ is hyperbolic.* □

Technically, Bestvina–Feighn assume "annuli flare" rather than conjugacies flare. The annuli flare condition is more difficult to define, yet not more enlightening. We refer the interested reader to the proofs of [Mut21, Lem. 5.2.3-5.2.4, Thm. 5.3.7] for an idea on how to upgrade conjugacy flaring to annuli flaring (via Proposition 4.2). Alternatively, the next subsection has another proof of the next theorem's main implication (3⇒1) that is independent of the current subsection. We are now ready to prove hyperbolization:

**Theorem 4.4** (cf. [Mut21, Thm. 5.2.7]). *The following are equivalent:*

1. *$\mathbb{F}*_\phi$ is hyperbolic;*
2. *$\mathbb{F}*_\phi$ has no $BS(1,d)$-subgroup for all $d \geq 1$;*
3. *for all $n, d \geq 1$ and nontrivial $x \in \mathbb{F}$, $\Phi^n(x)$ is not conjugate to $x^d$ in $\mathbb{F}$; and*
4. *$\phi$ is hyperbolic and its pullbacks stabilize.*

*Proof.* The implications $(1 \Rightarrow 2 \Rightarrow 3)$ are standard; $(3 \Rightarrow 4)$ is Propositions 3.2 and 4.1; and $(4 \Rightarrow 1)$ is Proposition 4.2 and Theorem 4.3. □

## 4.2 Mj–Reeves' relative combination theorem

Beyond hyperbolicity, we can still study the Dehn function of $\mathbb{F}*_\phi$. To this end, we will prove a relative version of Theorem 4.4 when $\Phi$ is not surjective — this is a new result!

Suppose $g\colon (\Gamma, G) \to (\Gamma, G)$ is the canonical relative expansion for $\phi$. A $\phi$-peripheral free-by-cyclic subgroup of $\mathbb{F}*_\phi$ is a subgroup corresponding to a component of the mapping torus of the peripheral restriction $g|_G$.

The free splitting $(\Gamma, G)$ lets us represent the sequence $([x_i])_{i=-L}^{L}$ for a conjugacy in $(\mathbb{F}, \phi)$ into a sequence $(\sigma_i)_{i=-L}^{L}$ of immersed loops in $\Gamma$. A conjugacy in $(\mathbb{F}, \phi)$ flares rel. $G$ if $2 \cdot \|\sigma_0\|_G \leq \max(\|\sigma_{-L}\|_G, \|\sigma_L\|_G)$. We say "$(\mathbb{F}, \phi)$ satisfies the conjugacies flare rel. $G$ condition" if there is an $L \geq 1$ such that all conjugacies in $(\mathbb{F}, \phi)$ of length $2L$ flare rel. $G$.

**Proposition 4.5.** *Suppose $\Phi$ is not surjective and $g\colon (\Gamma, G) \to (\Gamma, G)$ is the canonical relative expansion for $\phi$. $(\mathbb{F}, \phi)$ satisfies the conjugacies flare rel. $G$ condition if and only if pullbacks for $\phi$ stabilize.*

*Sketch of proof.* As $g$ is a relative expansion, we automatically have flaring rel. $G$ of unidirectional conjugacies in $(\mathbb{F}, \phi)$; therefore, conjugacy flaring rel. $G$ in $(\mathbb{F}, \phi)$ is equivalent to the existence of a uniform bound on the length of strictly bidirectional conjugacies, which is equivalent to pullback stability for $\phi$. □

Let $\widehat{g}\colon \widehat{\Gamma} \to \widehat{\Gamma}$ be the $\Phi$-equivariant "lift" of $g$ to the relative universal cover of $(\Gamma, G)$ — $\Phi$-equivariant means $\widehat{g}(x \cdot p) = \Phi(x) \cdot \widehat{g}(p)$ for all $x \in \mathbb{F}$ and $p \in \widehat{\Gamma}$. For $L \geq 1$, a hallway in $(\widehat{\Gamma}, g)$ of length $2L$ is a pair $((\rho_i)_{i=-L}^{L}, \rho)$, where $\rho_i$ is an immersed edge-path in $\widehat{\Gamma}$ and $\rho = (v_i)_{i=-L}^{L}$ is an immersed edge-path in $T_\phi$ of length $2L$. The hallway is unidirectional if $\rho$ is monotone (with respect to the $t$-orientation on $T_\phi$); it is strictly bidirectional if $\rho$ decomposes into two monotone pieces of lengths $L$.

The girth of a hallway $((\rho_i)_{i=-L}^{L}, \rho)$ is the combinatorial length $|\rho_0|$ of the edge-path $\rho_0$. The hallway flares if $2 \cdot |\rho_0| \leq \max(|\rho_{-L}|, |\rho_L|)$. The hallway is $r$-thin ($r \geq 0$) if the combinatorial distance between the initial (terminal resp.) vertices of $\rho_i$ and $g_j(\rho_j)$ is at most $r$ whenever $|i - j| = 1$ and $(v_i, v_j)$ is the $t$-orientation between the vertices — here, $g_j\colon \widehat{\Gamma} \to \widehat{\Gamma}$ is a lift of $g$ that depends on $(v_i, v_j)$; for instance, $g_j = \widehat{g}$ if $(v_i, v_j)$ is on the axis of $t$. The hallway is $\widetilde{G}$-bounded if it is 0-thin and the initial and terminal vertices of $\rho_i$ correspond to lifts of $G$ for $-L \leq i \leq L$.

We say "$(\widehat{\Gamma}, g)$ satisfies the $\widetilde{G}$-bounded hallways strictly flare condition" if there is an $L \geq 1$ such that all $\widetilde{G}$-bounded hallways in $(\widehat{\Gamma}, g)$ of length $2L$ flare. We say "$(\widehat{\Gamma}, g)$ satisfies the hallways flare condition" if there are $L \geq 1$ and $H\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that all $r$-thin hallways in $(\widehat{\Gamma}, g)$ of length $2L$ and girth at least $H(r)$ flare.

**Theorem 4.6** (relative combination [MR08, Thm. 4.5])**.** *Suppose $\Phi$ is not surjective, $g\colon (\Gamma, G) \to (\Gamma, G)$ is the canonical relative expansion for $\phi$, $\widehat{g}\colon \widehat{\Gamma} \to \widehat{\Gamma}$ is the $\Phi$-equivariant lift of $g$ to the relative universal cover of $(\Gamma, G)$. If $(\widehat{\Gamma}, g)$ satisfies the hallways flare and $\widetilde{G}$-bounded hallways strictly flare conditions, then $\mathbb{F}*_\phi$ is hyperbolic relative to the family of $\phi$-peripheral free-by-cyclic subgroups.* □

Besides being a generalization of hyperbolicity, the exact definition of (strong) relative hyperbolicity [Far98] is not crucial; what will matter is that the Dehn function of a relatively hyperbolic group is the Dehn function of its peripheral subgroups. To apply Mj–Reeves' relative combination theorem, we need to relate it to the conjugacies flare condition.

**Proposition 4.7.** *Suppose $\Phi$ is not surjective, $g\colon (\Gamma, G) \to (\Gamma, G)$ is the canonical relative expansion for $\phi$, and $\widehat{g}\colon \widehat{\Gamma} \to \widehat{\Gamma}$ is the $\Phi$-equivariant lift of $g$ to the relative universal cover of $(\Gamma, G)$. If $(\mathbb{F}, \phi)$ satisfies the conjugacies flare rel. $G$ condition, then $(\widehat{\Gamma}, g)$ satisfies the hallways flare and $\widetilde{G}$-bounded hallways strictly flare conditions.*

*Sketch of proof.* Let $((\rho_i)_{i=-L}^{L}, \rho)$ be a $\widetilde{G}$-bounded hallway in $(\widehat{\Gamma}, g)$. Each $\rho_i$ concatenates with a translate of itself to form a fundamental domain of an axis of some element $x_i \in \mathbb{F}$. In particular, it determines a conjugacy in $(\mathbb{F}, \phi)$ that flares (rel. $G$) if and only if the original $\widetilde{G}$-bounded hallway flares; therefore, the conjugacies flare condition implies the $\widetilde{G}$-bounded hallways strictly flare condition.

For the hallways flare condition, we first argue that strictly bidirectional $r$-thin hallways with large enough (with respect to $r$) girth have uniformly bounded (independent of $r$) length. If $|\rho_0|$ is large enough, then $g_{-1}(\rho_{-1}), \rho_0, g_1(\rho_1)$ coincide (modulo initial/terminal segments of length at most $r$). Thus the (relative) pullback of $(g, g)$ contains a long (rel. $G$) immersed loop that is not in the diagonal, and we get a strictly bidirectional conjugacy in $(\mathbb{F}, \phi)$ of length 2. The same argument shows that if $|\rho_0|$ is large enough, then there is a strictly bidirectional conjugacy in $(\mathbb{F}, \phi)$ with the same length as the initial strictly bidirectional hallway. But the former has a uniformly bounded length by conjugacy flaring.

The previous paragraph implies that sufficiently long $r$-thin hallways with large enough girth are almost unidirectional. As $g$ is a relative expansion, one can verify the hallways flare condition, starting with unidirectional hallways. A variation of this argument for annuli appears in [Mut21, Thm. 5.3.7]. We leave the details to the interested reader. □

We can now prove relative hyperbolization:

**Theorem 4.8.** *Suppose $\Phi$ is not surjective. The following are equivalent:*

1. *$\mathbb{F}*_\phi$ is hyperbolic relative to the family of $\phi$-peripheral free-by-cyclic subgroups;*
2. *$\mathbb{F}*_\phi$ has linear or quadratic Dehn function;*
3. *$\mathbb{F}*_\phi$ has no $BS(1,d)$-subgroup for all $d \geq 2$;*
4. *for all $n, d \geq 2$ and nontrivial $x \in \mathbb{F}$, $\Phi^n(x)$ is not conjugate to $x^d$ in $\mathbb{F}$; and*
5. *pullbacks for $\phi$ stabilize.*

*Proof.* The implication $(3 \Rightarrow 4)$ is an exercise; $(4 \Rightarrow 5)$ is Proposition 3.2; and $(5 \Rightarrow 1)$ is Propositions 4.5 and 4.7, and Theorem 4.6. $(1 \Rightarrow 2)$: As free-by-cyclic groups have linear or quadratic Dehn functions [BG10], so does $\mathbb{F}*_\phi$ [Far98, Thm. 3.8]. $(2 \Rightarrow 3)$: If $\mathbb{F}*_\phi$ has a $BS(1,d)$ for some $d \geq 2$, then it has an exponential Dehn function [Kap00, Cor. 5.7]. □

**Corollary 4.9.** *$\mathbb{F}*_\phi$ has a linear, quadratic, or exponential Dehn function.* □

*Second proof of Theorem 4.4:* 3⇒1*.* By Theorem 1.3(3⇒1), assume $\Phi$ is not surjective. By Theorem 4.8(4⇒1), $\mathbb{F}*_\phi$ is hyperbolic relative to the family of $\phi$-peripheral free-by-cyclic subgroups. These free-by-cyclic subgroups are hyperbolic by Theorem 1.3(3⇒1); therefore, $\mathbb{F}*_\phi$ is hyperbolic [Far98, Thm. 3.8]. □

## References


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

[BFH97] Mladen Bestvina, Mark Feighn, and Michael Handel. Laminations, trees, and irreducible automorphisms of free groups. *Geom. Funct. Anal.*, 7(2):215–244, 1997.

[BG10] Martin R. Bridson and Daniel Groves. The quadratic isoperimetric inequality for mapping tori of free group automorphisms. *Mem. Amer. Math. Soc.*, 203(955):xii+152, 2010.

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

[Far98] Benson Farb. Relatively hyperbolic groups. *Geom. Funct. Anal.*, 8(5):810–840, 1998.

[Fri15] Joel Friedman. Sheaves on graphs, their homological invariants, and a proof of the Hanna Neumann conjecture: with an appendix by Warren Dicks. *Mem. Amer. Math. Soc.*, 233(1100):xii+106, 2015.

[Kap00] Ilya Kapovich. Mapping tori of endomorphisms of free groups. *Comm. Algebra*, 28(6):2895–2917, 2000.

[Min12] Igor Mineyev. Submultiplicativity and the Hanna Neumann conjecture. *Ann. of Math. (2)*, 175(1):393–414, 2012.

[MR08] Mahan Mj and Lawrence Reeves. A combination theorem for strong relative hyperbolicity. *Geom. Topol.*, 12(3):1777–1798, 2008.

[Mut21] Jean Pierre Mutanguha. The dynamics and geometry of free group endomorphisms. *Adv. Math.*, 384:Paper No. 107714, 60, 2021.

[Neu57] Hanna Neumann. On the intersection of finitely generated free groups. Addendum. *Publ. Math. Debrecen*, 5:128, 1957.

[Neu90] Walter D. Neumann. On intersections of finitely generated subgroups of free groups. In *Groups—Canberra 1989*, volume 1456 of *Lecture Notes in Math.*, pages 161–170. Springer, Berlin, 1990.

[Rey11] Patrick R. Reynolds. *Dynamics of irreducible endomorphisms of* $F_N$. PhD thesis, University of Illinois at Urbana-Champaign, 2011.

[Sta83] John R. Stallings. Topology of finite graphs. *Invent. Math.*, 71(3):551–565, 1983.

[Thu82] William P. Thurston. Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. *Bull. Amer. Math. Soc. (N.S.)*, 6(3):357–381, 1982.

[Thu88] William P. Thurston. On the geometry and dynamics of diffeomorphisms of surfaces. *Bull. Amer. Math. Soc. (N.S.)*, 19(2):417–431, 1988.

[Tur96] Edward C. Turner. Test words for automorphisms of free groups. *Bull. London Math. Soc.*, 28(3):255–263, 1996.