Efficient Representatives for Automorphisms of Free Products

D. J. COLLINS & E. C. TURNER

The Scott conjecture for automorphisms of free groups says that if ϕ is an automorphism of a free group of rank n, then the subgroup of elements fixed by ϕ is a free group of rank at most n. This conjecture was recently settled in the positive in the brilliant paper of Bestvina and Handel [BH], which applied the Perron–Frobenius theory of nonnegative matrices [Se], the folding techniques introduced by Stallings [St], and motivations from the train-track theory of homeomorphisms of surfaces to the study of self-homotopy equivalences of graphs. In this paper we prove the generalization of the Scott conjecture to an arbitrary group G represented as a free product of freely indecomposible factors. Our proof is patterned on that of Bestvina and Handel: we consider 2-complexes $\mathfrak{X}$ whose fundamental groups are isomorphic to G and *efficient* self-homotopy equivalences $f: \mathfrak{X} \to \mathfrak{X}$ (see 2.10) that generalize the relative train-track maps of [BH]. In Section 1, we define topological maps of graphs of complexes, describe the way in which they model automorphisms of free products, and discuss the simplification operations on them. In Section 2 we define and prove the existence of efficient representatives for general automorphisms (see 2.12) and establish their most important properties. In Section 3, we apply the analysis of Section 2 to construct a "core" $\mathfrak{F}$ of the covering space $\hat{\mathfrak{X}}$ of $\mathfrak{X}$ corresponding to the subgroup Fix(ϕ) (see 3.8) and prove that its fundamental group has Kuroš rank at most that of the group (3.11). This was the approach employed in [GT] (resp. [CT]) in proving the finite rank of Fix(ϕ) in the free (resp. free product) case.

1. Basic Objects and Constructions

In order to generalize the Scott conjecture, we need general notions of the (absolute) rank of a group and the (relative) rank of a subgroup (groups will always be assumed to be countable). If $G = \bigstar_{i=1}^{m} G_i$ is represented as a free product of freely indecomposable factors, then it was shown in [Ku] that the set of factors (and in particular the number of factors) is well-defined up to isomorphism. The Kuroš subgroup theorem states that if H is a subgroup

Received January 26, 1993. Revision received October 20, 1993.
The second author wishes to thank the SERC for travel support during this research.
Michigan Math. J. 41 (1994).

of G then $H \cong F_s \star_{j=1}^{t} H_j$, where the H_j are intersections of H with conjugates of the factor groups G_i and F_s is a free group of rank s which meets no such conjugate (either of s or t could be ∞). If the representation of G as a free product is changed by an automorphism then these intersections change, but by [BL] the number of factors (and also their isomorphism types) is invariant. The following terms are therefore well-defined.

1.1. DEFINITION. If G is a group and $G = \star_{i=1}^{m} G_i$ where each factor G_i is freely indecomposable, then m is the *Kuroš rank of* G—denoted K-rank(G). If H is a subgroup of G such that $H \cong F_s \star_{j=1}^{t} H_j$ as above, then $s+t$ is the *Kuroš subgroup rank of* H *in* G—denoted K-rank(G, H).

Comment: Clearly K-rank(G, H) $\leq$ K-rank(H) in general. The two are equal if $s = 0$ and the factor groups H_j have the property that none of their subgroups decompose nontrivially as free products—for example, if G is a free product of finite groups. Also, K-rank(G, H) is invariant under abstract group isomorphism.

Our main theorem is the following.

THEOREM 3.12. *If* $\phi: G \to G$ *is an automorphism, then*

$$\text{K-rank}(G, \text{Fix}(\phi)) \leq \text{K-rank}(G).$$

1.2. GRAPHS OF COMPLEXES. The multiplication table complex C_H of the group H is the standard 2-complex associated to the multiplication table presentation of H: thus C_H has a single vertex, an edge for each pair $\{x, x^{-1}\}$ in H, and a 2-cell c with $\partial c = xyz$ for each relation $xyz = 1$ in G. A *graph of complexes* $\mathfrak{X}$ is the union of a graph X with a family $\{C_j\}$ of 2-complexes called the *factor complexes,* where C_j is the multiplication table complex C_{G_j} of a freely indecomposable group $G_j \neq \mathbf{Z}$ with an edge E_{C_j} called its *stem* that joins a vertex v_{C_j} of X to C_{G_j}. (We use multiplication table complexes for uniqueness of representation of group elements by paths: if finite complexes were required, say, then other choices could be made. Stems are included for convenience in keeping track of "countries" in the covering spaces of Section 3 and can be safely ignored elsewhere.) The union of the factor complexes (including stems) is denoted by $\mathcal{C}(\mathfrak{X})$. The edges of the underlying graph X are *real* edges and the edges of the factor complexes are *infinitesimal* edges. An *infinitesimal* x *in the complex* C is a loop at v_C of the form $x = E_C e \bar{E}_C$, where e is an edge of the multiplication complex. (The term infinitesimal is used because, when lengths are assigned to edges in Section 2, the edges of the factor complexes will always have length zero.) Clearly the fundamental group $\pi_1(\mathfrak{X})$ is isomorphic to the free product of the free group $\pi_1(X)$ with the groups G_j. The *rosebush* $\mathfrak{R}_G$ for the free product $G = F_s \star_{j=1}^{t} G_j$ is the graph of complexes whose underlying graph X is a bouquet of s circles with vertex v_G together with t free edges to which the stems for the C_{G_j} are attached: clearly $\pi_1(\mathfrak{R}_G, v_G) \cong G$. Graphs of complexes are the topological models for free products which we use in a manner analogous to the familiar way in which graphs model free groups.

The following proposition is clear.

1.3. PROPOSITION. *If $\mathfrak{X}$ is a graph of complexes whose underlying graph X has rank s and whose factor complexes correspond to freely indecomposable groups $\{H_1, \ldots, H_t\}$, then for any vertex $v \in \mathfrak{X}$,*

$$\pi_1(\mathfrak{X}, v) \cong F_s \bigstar_{j=1}^{t} H_j.$$

Furthermore, K-rank$(\pi_1(\mathfrak{X}, v)) = s + t$.

1.4. TOPOLOGICAL MAPS AND REALIZATIONS OF AUTOMORPHISMS. A map

$$f: \mathfrak{X} \to \mathfrak{X}'$$

between graphs of complexes is *topological* if it is a continuous map satisfying the following conditions.

(1) f carries vertices to vertices.
(2) The restriction of f to each factor complex C_i is a homeomorphism onto a complex C_j which maps edges to edges and stems to stems.
(3) Every real edge of $\mathfrak{X}$ can be subdivided as $[z_0, z_1, \ldots, z_r]$ in such a way that $f|_{[z_i, z_{i+1}]}$ maps the subinterval $[z_i, z_{i+1}]$ linearly across a real edge or an infinitesimal. Furthermore, no two successive subintervals are mapped to inverse real edges or to infinitesimals in the same factor complex.

Note that a topological map f is locally injective on real edges. Standard techniques show that any continuous map between graphs of complexes is homotopic to an essentially unique topological map. This process will be called *tightening* in general and *real tightening* if real edges are involved. (For uniqueness, it is important that the factor complexes are multiplication complexes.) A map is called a *train-track map* [BH, p. 8] if every iterate f^k of f is topological. Train-track maps have the property that if each edge can be assigned a length in such a way that f has the effect of increasing the length of each edge by the same factor λ, then f has the same effect on the lengths of legal paths.

If $f: \mathfrak{X} \to \mathfrak{X}$ is topological homotopy equivalence and μ is a path in $\mathfrak{X}$ from v to $f(v)$, then the *path-induced automorphism* $\pi_1(f, \mu): \pi_1(\mathfrak{X}, v) \to \pi_1(\mathfrak{X}, v)$ is defined as

$$\pi_1(f, \mu)([\alpha]) = [\mu \circ f(\alpha) \circ \bar{\mu}].$$

(In general, the inverse of a path σ is denoted by $\bar{\sigma}$.) If μ_v is the constant path at a fixed point v, then we call $\pi_1(f, \mu_v)$ a *point-induced automorphism* and write $\pi_1(f, \mu_v) = \pi_1(f, v)$.

1.5. DEFINITION. If $\phi: G \to G$ is an automorphism of G, then a *representative of ϕ* is a topological self-homotopy equivalence f of a graph of complexes $\mathfrak{X}$, a path μ in $\mathfrak{X}$, and a homotopy equivalence $\tau: \mathcal{R}_G \to \mathfrak{X}$ with $\tau(v_G) = \mu(0) = v$ so that the following diagram commutes:

$$\begin{array}{ccc} G \cong \pi_1(\mathcal{R}_G, v_G) & \xrightarrow{\pi_1(\tau)} & \pi_1(\mathfrak{X}, v) \\ \phi \downarrow & & \downarrow \pi_1(f,\mu) \\ G \cong \pi_1(\mathcal{R}_G, v_G) & \xrightarrow{\pi_1(\tau)} & \pi_1(\mathfrak{X}, v). \end{array}$$

A *rosebush representative* is one for which $\mathfrak{X} \cong \mathcal{R}_G$ and τ is the identity.

1.6. PROPOSITION. *A topological map f of a graph of complexes $\mathfrak{X}$ is a homotopy equivalence if and only if $\pi_1(f, \mu)$ is an isomorphism for any μ. Every automorphism $\phi \in \mathrm{Aut}(G)$ has rosebush representatives.*

Proof. Clearly, if f is a homotopy equivalence then $\pi_1(f, \mu)$ is an isomorphism for any μ. It is easy to reduce the converse to the case in which μ is the constant path at a fixed point v_0. So assume that $\pi_1(f, v_0)$ is an isomorphism.

Case 1: The underlying graph X is a tree T (perhaps infinite and with infinitely many factor complexes).

The fundamental group $G = \bigstar\, G_i$ is a free product of the fundamental groups of the factor complexes, none isomorphic to $\mathbf{Z}$, and $\pi_1(f, v_0)(G_i)$ is a conjugate of $G_{\sigma(i)}$ for some permutation $\sigma(i)$ [Fo]. If $f(T) \subset T$, then f is clearly a homotopy equivalence (since then f is well-defined on $\mathfrak{X}/T$). In general, let $f_{\bar{\sigma}}$ be some homotopy equivalence of $\mathfrak{X}$ which carries the factor complex $C_{\sigma(i)}$ to C_i by the inverse of the homeomorphism $f|_{C_i}$ (with e.g. $f(T) \subset T$), and let $f' = f \circ f_{\bar{\sigma}}$. Thus f is a homotopy equivalence if and only if f' is, and f' fixes all the factor complexes.
Claim: The map f' is a homotopy equivalence.
Proof of claim: The induced map $\pi_1(f', v_0)$ is an isomorphism and f' fixes all the vertices at which factor complexes are attached. By a homotopy supported near the remaining vertices we can arrange that it fixes all vertices. After tightening, then, the image of f' on each edge (and so each path) is determined by an element of $\bigstar\, G_i$, the intermediate tree segments being uniquely determined, and the automorphism $\pi_1(f', v_0)$ has the effect of conjugating each G_i by the label on the path from v_0 to the vertex at which C_i is attached. The inverse automorphism has the same form—the homotopy inverse for f' fixes the factor complexes and is defined on edges to give the conjugating elements for this inverse.

Case 2: $\mathfrak{X}$ is any graph of complexes.

Let K be the normal subgroup of G generated by the factor complexes; that is, K is the kernel of the map on π_1 induced by the map that collapses each factor complex to its vertex in X. Let $\tilde{\mathfrak{X}}_K$ be the covering space of $\mathfrak{X}$ corresponding to K. Since K is invariant under f, there is a lift $\tilde{f}_K: \tilde{\mathfrak{X}}_K \to \tilde{\mathfrak{X}}_K$ covering f and (see 3.2(iii)) v_0 is covered by a fixed point $\tilde{v}_0$ of $\tilde{f}_K$. Furthermore, $\tilde{f}_K$ induces an isomorphism on $\pi_1(\tilde{\mathfrak{X}}_K, \tilde{v}_0)$. ($K$ is the normal subgroup generated by the factors of G that are not isomorphic to $\mathbf{Z}$. The argument of step 1 of Theorem 2.11 below shows that if a free factor is invariant under an isomorphism of a finitely generated free product, then the restriction is an isomorphism.)

The covering projection map is an isomorphism on all homotopy groups beyond the first, so $\tilde{f}_K$ induces an isomorphism on π_k if and only if f does. But by Case 1, $\tilde{f}_K$ is a homotopy equivalence. It follows that f induces an isomorphism on all homotopy groups and, since $\mathfrak{X}$ is a CW complex, it follows from the Whitehead theorem [Sp, p. 405] that f is a homotopy equivalence. This completes the proof of the first statement of the proposition.

Rosebush representatives can be composed, so it suffices to show that generators of Aut(G) can be induced by topological maps. It is routine to show that all the Fouxe-Rabinovitch generators [Fo] have rosebush representatives—in fact, these representatives can be taken with trivial paths μ. □

Just as in the case of free groups, varying the paths μ between the basepoint v and its image $f(v)$ corresponds to varying the automorphism through its outer automorphism class. Thus a representative f corresponds to an outer automorphism $\mathcal{O}$ of G.

1.7. MATRICES M AND PF-INDICES $\Lambda(M)$. If $M = (m_{ij})$ is an $n \times n$ matrix of nonnegative integers, define $\Gamma(M)$ to be the directed graph with vertices $\{1, \ldots, n\}$ and an edge joining i to j if $m_{ji} > 0$. This partially orders the indices $\{1, \ldots, n\}$ by the condition

$$i \preccurlyeq j \Leftrightarrow \text{there is a directed path in } \Gamma(M) \text{ from } i \text{ to } j.$$

This condition is equivalent to the existence of some power $M^k = (m_{ij}^{(k)})$ of M so that the image of the ith basis vector under M^k has a nonzero value in the jth position—i.e., $m_{ji}^{(k)} > 0$. The relation $\preccurlyeq$ determines equivalence classes of indices; namely,

$$i \equiv j \Leftrightarrow \{k \mid i \preccurlyeq k\} = \{k \mid j \preccurlyeq k\}.$$

M is said to be *irreducible* if $\Gamma(M)$ has one equivalence class.

Comment: In [BH, §5] i and j are defined to be equivalent if $i \preccurlyeq j$ and $j \preccurlyeq i$, but their proofs suggest that they are working with the above definition. The two notions agree except for classes determining zero diagonal blocks (see discussion following 1.8). For example, the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

has two equivalence classes relative to the definition given above but three relative to that in [BH].

1.8. THEOREM. *Suppose that $M = (m_{ij})$ is an irreducible $n \times n$ matrix of nonnegative integers.*

(i) *There is a unique positive eigenvector* $\mathbf{w}$ *of norm* 1 *for M which has associated eigenvalue $\lambda(M) \geq 1$.*

(ii) *If* $\mathbf{v}$ *is a positive vector, $\mu > 0$, and $M\mathbf{v} \leq \mu\mathbf{v}$ with at least one component inequality strict, then $\lambda(M) < \mu$.*

(iii) *If M_1 is an irreducible square matrix obtained from M by reducing some entries or by deleting pairs of rows and columns, then $\lambda(M_1) < \lambda(M)$.*

(iv) *For any i, j, $m_{ij} \leq \lambda^n$.*

The eigenvalue λ will be called the *PF eigenvalue* of M and an associated eigenvector a *PF eigenvector.*

Proof. Statements (i) and (ii) are the Perron–Frobenius theorem, a proof of which is in [Se].

(iii) Suppose that M_1 results from reducing some entries of M and that $\mathbf{v}$ is a PF eigenvector for M. Then $M_1\mathbf{v} \leq \lambda(M)\mathbf{v}$ and at least one component inequality must be strict. Hence, by (ii), $\lambda(M_1) < \lambda(M)$.

Now suppose that some pairs of rows and columns of M are deleted. Without loss of generality we may assume that M_1 is the leading $r \times r$ submatrix of M. If $\mathbf{v}$ is a PF eigenvector for M and $\mathbf{w}$ consists of the first r entries of $\mathbf{v}$, then $M_1\mathbf{v} \leq \lambda(M)\mathbf{v}$. Since M is irreducible, some entry in the upper right-hand block of M is nonzero and therefore in the inequality $M_1\mathbf{v} \leq \lambda(M)\mathbf{v}$ at least one component inequality is strict. Again the result follows from (ii).

(iv) Write $\mathbf{w} = (b_i)$. From $M\mathbf{w} = \lambda\mathbf{w}$ we have $m_{ij}b_j \leq \lambda b_i$. Choose k such that $m_{ji}^{(k)} \neq 0$. From $M^k\mathbf{w} = \lambda^k\mathbf{w}$ we obtain $m_{ji}^{(k)}b_i \leq \lambda^k b_j$. Combining these and cancelling b_ib_j gives $m_{ij}m_{ji}^{(k)} \leq \lambda^{k+1}$.

Since M is irreducible, the directed graph $\Gamma(M)$ is path-connected. In particular, there exists a path of length at most $n-1$ from the vertex i to the vertex j and so we can choose $k \leq n-1$. (This argument is due to Aidan Schofield.) □

If M is an $n \times n$ matrix, then simultaneous permutation of rows and columns can be performed so that $i < j \Rightarrow i \preccurlyeq j$. This puts M into the form

$$M = \begin{pmatrix} M_1 & ? & ? & ? \\ 0 & M_2 & ? & ? \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_k \end{pmatrix}$$

where the matrices M_r are either zero matrices or irreducible matrices, each of which corresponds to an equivalence class under the relation $\preccurlyeq$. The M_r are determined up to row–column permutation, so in particular the set of PF values of the diagonal blocks is well-defined.

1.9. DEFINITION. The *PF index* $\Lambda(M) = \{\lambda_{i_1} \geq \lambda_{i_2} \geq \cdots \geq \lambda_{i_k}\}$ *of* M is the nonincreasing list of PF values $\lambda_i > 1$ of the matrices M_j. The set of PF indices is ordered lexicographically (with zeroes added at the end if necessary).

1.10. THEOREM. *For any PF index Λ_0 and integer n, there are only finitely many $\Lambda(M) \leq \Lambda_0$ associated with nonnegative $n \times n$ integral matrices M.*

Proof. This is a consequence of 1.8(iv). Since all the diagonal blocks M_r of M have size bounded by n, their entries are all bounded by λ^n where λ is the largest entry in Λ. □

1.11. STRATIFICATION OF $\mathfrak{X}$ AND $\Lambda(f)$ FOR TOPOLOGICAL $f: \mathfrak{X} \to \mathfrak{X}$. Suppose that $f: \mathfrak{X} \to \mathfrak{X}$ is topological and that $\{E_1, \ldots, E_n\}$ is the set of real edges of $\mathfrak{X}$. If m_{ij} is the number of times $f(E_j)$ covers E_i (regardless of orientation), then the *transition matrix* $M(f) = (m_{ij})$ and the PF index of f is $\Lambda(f) = \Lambda(M(f))$. The real edges of $\mathfrak{X}$ are grouped into equivalence classes by the relation $\leqslant$ and these classes can be linearly ordered. The rth *stratum* $\mathcal{H}_r$ is the union of the closed edges in the rth equivalence class (with $\mathcal{H}_0 = \mathcal{C}(\mathfrak{X})$) and the rth *subgraph of complexes* $\mathfrak{X}_r$ is $\bigcup_{j=0}^{r} \mathcal{H}_j$. Thus f determines a stratification of $\mathfrak{X}$:

$$\mathcal{C}(\mathfrak{X}) = \mathfrak{X}_0 \subset \mathfrak{X}_1 \subset \cdots \subset \mathfrak{X}_m = \mathfrak{X}.$$

The stratum $\mathcal{H}_r$ is called *growing, level,* or *descending* according as the corresponding diagonal matrix is irreducible with PF value > 1, is irreducible with PF value $= 1$, or is a zero matrix.

1.12. OPERATIONS ON TOPOLOGICAL REPRESENTATIVES f AND EFFECTS ON $\Lambda(f)$. The basic strategy for proving that efficient representatives for automorphisms exist is to start with a topological representative $f: \mathfrak{X} \to \mathfrak{X}$ of $\phi \in \mathrm{Aut}(G)$ and to improve it, the measure of improvement being the PF index $\Lambda(f)$. The basic operations (following [BH]) are the following:

(1) pruning (collapsing invariant and pretrivial forests);
(2) adding a prevertex (BH subdivision);
(3) valence-1 homotopy (including pruning);
(4) valence-2 homotopy (including pruning);
(5) core subdivision;
(6) folding; and
(7) adding isolated fixed points to the vertex set.

Comments on the Basic Operations: These operations are all analyzed in [BH] for the case of ordinary graphs, and their results carry over to the setting of graphs of complexes with the following qualifications. The forests (which are unions of whole edges) to be collapsed in *pruning*—either maximal forests that are invariant under f or that are mapped to the vertex set by some iterate of f [BH, paragraph before 1.5]—are contained in the underlying graph X: in particular, stems are not collapsed. *Adding a prevertex* (i.e., a point whose image is a vertex of X) is exactly the same as in [BH, Lemma 1.10] (where it is called subdivision). Note that all prevertices lie in X. *Valence-1 and -2 homotopies* [BH, Lemma 1.11 & Lemma 1.13] are also the same with the understanding that the vertices in question have degree 1 or 2 in X and have no factor complexes attached. *Core subdivision* can also be defined as in [BH, proof of Theorem 5.12]. Core subdivision is applied to a particular stratum $\mathcal{H}_r$: each edge $E \in \mathcal{H}_r$ is subdivided (perhaps) into three subedges such that (a) each point in the initial and terminal subedges is mapped to $\mathcal{H}_{r-1}$ by some iterate of f and (b) these "germs" are maximal with respect to this property.

The basic operation of *folding* on a graph of complexes is exactly as in [BH], but its use in the context of illegal turns in irreducible Nielsen paths requires comment. For us, a *turn at v* is a pair $\langle E_1, E_2\rangle$ where E_1 and E_2 are real edges and the terminal vertex of E_1 is the initial vertex of E_2. The turn $\langle E_1, E_2\rangle$ is *degenerate* if $E_2 = \bar{E}_1$. We call $\langle E_1, E_2\rangle$ a *real* turn if $f(E_1) = \dots E_1'$ and $f(E_2) = E_2' \dots$, where the Es are real edges; in these circumstances, the image turn is $f(\langle E_1, E_2\rangle) = \langle E_1', E_2'\rangle$. We shall be concerned only with real turns all of whose iterated images under f are also real; such a turn is said to be *illegal* if some image under f is degenerate and *legal* otherwise. With this notion of illegal turn, the Bestvina–Handel discussion of folding carries over (in particular, the issue of full as opposed to partial folds in the analog of [BH, Lemma 3.7]).

An operation that replaces f with f' will be called *safe* if $\Lambda(f') \le \Lambda(f)$ and *beneficial* if $\Lambda(f') < \Lambda(f)$. The following theorem summarizes the effects of the above operations.

1.13. THEOREM. *All of the basic operations are safe with one exception: a valence-2 homotopy applied to a valence-2 vertex v which is in the interior of a growing stratum $\mathcal{H}_r$. The following are beneficial: valence-1 homotopies, pruning that involves any edges in growing strata, and folding when real tightening is necessary.*

Proof. The basic operations are analyzed by Bestvina and Handel in [BH] in the free case, and their proofs carry over to the general case with the above comments. Pruning is discussed in [BH] in the paragraph preceding the statement of Theorem 1.7, adding a prevertex in Lemmas 1.10 and 5.1, valence-1 homotopy in Lemmas 1.11 and 5.2, valence-2 homotopy in Lemmas 1.13 and 5.4, and folding in Lemmas 1.5 and 5.3. Core subdivision or adding an isolated fixed point to the vertex set may add level or descending strata, but the growing strata and their PF values are unchanged. □

2. Efficient Representatives

The goal of this section is to define and prove the existence of *efficient* representatives $f\colon \mathcal{X} \to \mathcal{X}$ (see 2.10). These representatives generalize the relative train-track maps of Bestvina and Handel for the case of free groups and play the same role in the proof of the Scott conjecture. The key step in both analyses is the understanding of Nielsen paths (see 2.1) and the definition of a measure of r-length of paths in $\mathcal{X}_r$ for growing strata $\mathcal{H}_r$ (see 2.9).

2.1. DEFINITION. A path ρ between fixed points of f is a *Nielsen path* (NP) if $f(\rho)$ is homotopic to ρ relative to its endpoints, denoted $f(\rho) \sim \rho$. An *indivisible Nielsen path* (INP) is a real Nielsen path that is not a nontrivial product of real Nielsen paths or an infinitesimal Nielsen path that lies in one factor complex. A Nielsen path ρ has *height* r if ρ lies in $\mathcal{X}_r$ and contains edges of $\mathcal{H}_r$.

Two technical problems arise in the general case due to the presence of infinitesimals. Suppose, for example, that $\rho = \sigma E y$ is an INP, with $f(E) = \cdots Ex$; then $f(y) = y'$ and $y = xy'$. But there may be many infinitesimals z such that $z = xz'$ ($z' = f(z)$), producing perhaps infinitely many INPs of the form $\sigma E z$. Furthermore, if $\rho = \rho_1 x \rho_2$ (where ρ_1 and ρ_2 are INPs and x is a fixed infinitesimal), then $\rho = (\rho_1 x)\rho_2 = \rho_1(x\rho_2)$ are distinct factorizations into INPs. This motivates the following definition.

2.2. Definition. A Nielsen path is *real-ended* if it begins and ends with real edges. The topological representative $f: \mathcal{X} \to \mathcal{X}$ is *neat* if any INP ρ' can be written as $\rho' = x\rho y$ for fixed infinitesimals x and y and a real-ended INP ρ.

2.3. Proposition. *Suppose $f: \mathcal{X} \to \mathcal{X}$ is a topological representative. Then:*

(i) *f is homotopic to a neat representative; and*
(ii) *if f is neat then any Nielsen path can be factored uniquely into real-ended and infinitesimal INPs.*

Proof. (i) Consider first the case indicated above (in which the E appearing in $f(\rho)$ is in the image of E); namely $\rho = \sigma E y$ is a Nielsen path with $f(E) = \cdots Ex$, $f(y) = y'$, and $y = xy'$. First, "untighten" f by subdividing E into E_1E_2; map E_1 to the image of E followed by y' and E_2 to y'^{-1}. Then fold $\bar{E}_2$ against y (which has the same image as $\bar{E}_2$) to get E_1 mapped to $\cdots E_1 y^{-1}xy'$. But $y^{-1}xy' = 1$, so tightening gives the desired map. This is described in Figure 1, where the image of an edge is indicated (in parentheses) below the edge and $\boxed{C}$ denotes the factor complex.

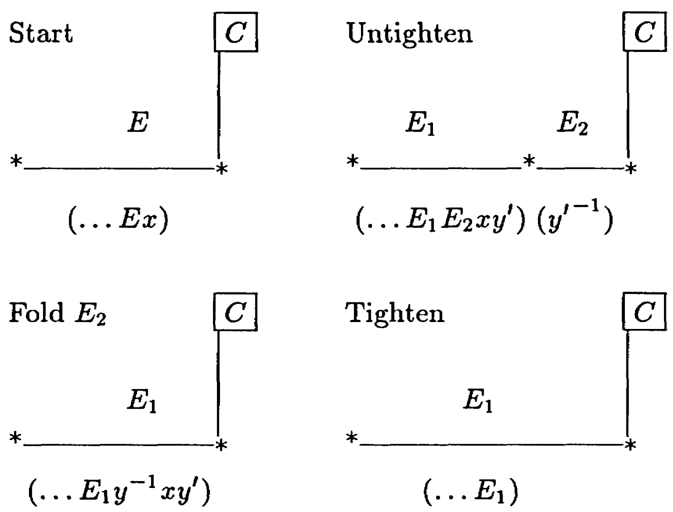

Figure 1 Neatening a topological representative

The remaining possibility is that the E appearing in $f(\rho)$ is in the image of σ, and so $\rho=\sigma Ey$, $f(\sigma)=\sigma E\bar{\alpha}$, and $f(E)=\alpha x=\alpha yy'^{-1}$ for some real α. First subdivide and untighten E as $E=E_1\cdot E_2\mapsto \alpha y\cdot y'^{-1}$ and then subdivide σ as $\sigma_1\cdot\sigma_2\mapsto\sigma_1\sigma_2E_1\cdot E_2\bar{\alpha}$. Fold $\bar{E}_2$ against y^{-1} (as above), producing a real-ended INP $\sigma_1\sigma_2E_1$ where

$$\sigma_1\mapsto\sigma_1\sigma_2E_1,$$

$$\sigma_2\mapsto y'^{-1}\bar{\alpha},$$

$$E_1\mapsto\alpha y',$$

Now fold E_1 against $\bar{\sigma}_2$ to get the INP σ_1. The effect of these steps is to replace the original INP ρ by the shorter σ_1 without increasing the number of edges in the graph (the original E has vanished but σ has been subdivided). A finite number of these steps, together with those described by Figure 1, produces a neat representative.

(ii) If f' is neat, then any Nielsen path can be uniquely factored into real-ended INPs and infinitesimal bits. Each such bit is a fixed element in a free product under an automorphism that permutes or preserves the factors and so must be a product of fixed elements of the factors. □

The above *neatening* process can be applied to any topological representative to produce a neat map and also preserves all properties of interest.

2.4. PROPOSITION. *A neat map has only finitely many real-ended INPs.*

2.5. DEFINITION. If f is neat and $\mathcal{H}_r$ is a growing stratum, then $N(f,r)$ is the number of real-ended INPs of height r and

$$N(f)=\sum_{\text{growing }\mathcal{H}_r}N(f,r).$$

Comment: Neatness is a convenience here rather than an esential condition: for maps that aren't neat, one could define $N(f,r)$ to be the number of equivalence classes of INPs of height r, where two such classes are considered equivalent if they differ only in infinitesimals at the beginning and end.

Proof of Proposition 2.4. Since f is topological, it has only finitely many fixed points other than whole fixed edges, so it suffices to show that there are finitely many INPs joining a particular pair. If $\{\rho_0,\rho_1,\dots\}$ is an infinite set of distinct INPs joining x_0 to x_1, then $\{\sigma_i=\rho_i\rho_0^{-1}: i\ge 1\}$ is an infinite set of distinct elements of the free part of $\mathrm{Fix}(\pi_1(f,x_0))$, which is known to be finitely generated [CT]. Suppose that $\{\lambda_1,\dots,\lambda_k\}$ is a set of generators of the free part of $\mathrm{Fix}(\pi_1(f,x_0))$. Then each λ_i is homotopic to a product of INPs and any word in the λ_is is a product of at least as many INPs as the length of the word. But this could only account for as many as $\binom{2k}{2}$ products of two INPs, contradicting the infinity of the list $\{\sigma_i: i\ge 1\}$. □

The proofs of our main theorems require that we begin with a representative of minimum Λ value, the existence of which is complicated (as in [BH]) by the fact that valence-2 homotopies are not always safe. In an earlier version of this paper, we "proved" that for any representative of an automorphism ϕ there is a *geometrically bounded* representative—that is, one for which the graph of complexes $\mathfrak{X}$ has no vertices of degree 1 or 2—whose Λ value is no larger. Since there are only finitely many geometrically bounded representatives of any automorphism, this would prove that there is one with minimum value of Λ. We are indebted to Warren Dicks and Enrique Ventura for convincing us that in our "proof" the quotes are necessary. We therefore mimic the method of [BH] by considering representatives f such that $\Lambda(f)$ "looks like" it belongs to a geometrically bounded representative.

2.6. DEFINITION. The representative f is *bounded* if it has no more than $3n-3$ strata (the maximum possible among geometrically bounded representatives) and if each $\lambda_r \in \Lambda(f)$ is the PF value associated with some matrix of size $(3n-3)\times(3n-3)$. $\Lambda_{\min}(\phi) = \min\{\Lambda(g) \mid g$ is a bounded representative of $\phi\}$.

2.7. PROPOSITION. *If f is a bounded representative of ϕ such that $\Lambda(f) = \Lambda_{\min}(\phi)$ and if $\tilde{f}$ is obtained from f by safe moves, then $\Lambda(\tilde{f}) = \Lambda_{\min}(\phi)$. There are finitely many bounded representatives g of ϕ such that $\Lambda(g)$ is less than any given bound.*

Proof. The proofs of Lemmas 5.1–5.6 of [BH] apply with all operations generalized as in Section 1. This proof leaves open the possibility that there are nongeometrically bounded representatives with strictly smaller Λ values. It would be interesting to know whether this possibility can occur. □

We are now in a position to define the central notions of this section: that of a relative train-track map (2.8) and of an efficient representative (2.10) for an automorphism of a free product.

2.8. DEFINITION. The topological representative $f: \mathfrak{X} \to \mathfrak{X}$ is a *relative train-track* (RTT) *map* if the following conditions hold for each growing stratum $\mathcal{H}_r$.

(i) *f preserves r-germs* if, for each edge $E \in \mathcal{H}_r$, $f(E)$ begins and ends with edges in $\mathcal{H}_r$. (Note: this means that turns with both edges in $\mathcal{H}_r$ are legal or illegal as discussed in 1.12.)

(ii) f is *injective on r-connecting paths* if, for each nontrivial path $\alpha \subseteq \mathfrak{X}_{r-1}$ joining points of $\mathcal{H}_r \cap \mathfrak{X}_{r-1}$, $[f(\alpha)]$ is nontrivial.

(iii) f is *r-legal* if σ r-legal implies that $f(\sigma)$ is r-legal.

2.9. MEASURING LENGTHS OF PATHS IN $\mathfrak{X}_r$. If $f: \mathfrak{X} \to \mathfrak{X}$ is an RTT map and $\mathcal{H}_r$ is a growing stratum, then (as in [BH]) an eigenvector for $M(f)_r$ determines a length $L_r(E)$ for each edge E of the growing stratum $\mathcal{H}_r$. For an r-legal path σ composed of whole edges, define

$$L_r(\sigma) = \sum_{E \subset \sigma \cap \mathcal{H}_r} L_r(E).$$

Since f is topological, $L_r(f(E)) = \lambda_r L_r(E)$. It follows from 2.8(ii) and 2.8(iii) that on any r-legal σ composed of complete edges, $L_r(f(\sigma)) = \lambda_r L_r(\sigma)$. Suppose that $\mu_i \in \mathcal{X}_{r-1}$ and $\lambda_i \in \mathcal{H}_r$; then

$$\sigma = \mu_0 \lambda_1 \mu_1 \dots \lambda_k \mu_k \Rightarrow [f(\sigma)] = [f(\mu_0)][f(\lambda_1)][f(\mu_1)] \dots [f(\lambda_k)][f(\mu_k)]$$

and the $[f(\mu_i)]$ are nontrivial. If $\sigma \in \mathcal{H}_r$ is any path, then $\sigma = \gamma_1 \bar{\sigma} \gamma_2$ for germs γ_1 and γ_2. Let $\tilde{L}_r(\sigma) = L_r(\bar{\sigma})$ and define

$$L_r(\sigma) = \lim_{m \to \infty} \left(\frac{\tilde{L}_r(f^m(\sigma))}{\lambda_r^m} \right).$$

(Note that the sequence is increasing and bounded by the length of the path extending σ that includes the whole edges in which γ_1 and γ_2 are contained.) It follows that, for all r-legal paths $\sigma \in \mathcal{X}_r$, $L_r(f(\sigma)) = \lambda_r L_r(\sigma)$.

This defines a measure of length for arbitrary paths in $\mathcal{X}_r$, since any path can be factored into a product of legal pieces; the r-length of $\mathcal{X}$ is

$$L_r(\mathcal{X}) = \sum_{E \subset \mathcal{H}_r} L_r(E).$$

By a *metric RTT map* we mean an RTT map f together with length functions L_r (corresponding to eigenvalues λ_r) as above.

2.10. DEFINITION. The map f is *stable* if $\Lambda(f) = \Lambda_{\min}(\phi)$ and $N(f)$ is the least value among all representatives with PF index $\Lambda_{\min}(\phi)$. We say f is *efficient* if it is a neat, stable relative train-track map with the additional property that, for each growing stratum $\mathcal{H}_r$ and INP ρ_r of height r, $L_r(\rho_r) = 2L_r(\mathcal{X})$.

2.11. THEOREM. *If $f: \mathcal{X} \to \mathcal{X}$ is efficient, then for each growing stratum $\mathcal{H}_r$ there exists at most one real INP ρ_r of height r.*

2.12. THEOREM. *For any automorphism ϕ, there exist efficient representatives.*

Proof of Theorem 2.11.

Step 1: It suffices to prove the theorem for the case in which $\mathcal{H}_r$ is the topmost stratum.

If there is a real INP ρ of height r, then f maps the component of $\mathcal{X}_r$ containing ρ back to itself, and irreducibility of the rth stratum shows that there is only one component $\hat{\mathcal{X}}_r$ of $\mathcal{X}_r$ containing real edges (the remainder being complexes at isolated vertices of X). To establish step 1 it suffices to show that the restriction $\hat{f}_r$ of f to the invariant subcomplex $\hat{\mathcal{X}}_r$ is a homotopy equivalence.

If x_0 is an endpoint of ρ then clearly $\pi_1(\hat{f}_r, x_0)$ is an injection. Let G and $\hat{G}$ be the fundamental groups of $\mathcal{X}$ and $\hat{\mathcal{X}}_r$, respectively, and let $\alpha = \pi_1(f, x_0)$.

Without loss of generality we can assume that $G \supsetneq \hat{G}$. If $\alpha|_{\hat{G}}$ is not a surjection, then

$$G \supsetneq \hat{G} \supsetneq \alpha(\hat{G})$$

is a strictly decreasing sequence. But since α is an automorphism of G, an application of the Kuroš subgroup theorem shows that $\alpha(\hat{G})$ is a free factor of $\hat{G}$, which contradicts that fact that $\hat{G}$ and $\alpha(\hat{G})$ have the same K-rank. (The application of the Kuroš subgroup theorem is as follows. Write $H = \hat{G}$ and $K = \alpha(\hat{G})$, so that $G \supseteq H \supseteq K$ and K is a free factor of G; we show that K is a free factor of H. H is represented as a subgroup of the free product $G = K \star L$, where L is any complementary free factor in a way that contains a free factor $H \cap uKu^{-1}$ for u a representative of the double coset HK [Co, pp. 175–180]. Taking $u = 1$ gives K as a free factor of H.) Thus $\pi_1(\hat{f}_r, x_0)$ is an isomorphism. But since $\hat{f}_r$ is a topological map of a graph of complexes, it follows (see 1.6) that it is a homotopy equivalence. This finishes step 1.

Suppose a choice of length functions L_r has been made so that f is a metric RTT map. Arguing as in [BH, Lemmas 3.4 & 5.11], each real INP ρ of height r contains one illegal turn in $\mathcal{H}_r$ and $\rho = \alpha\bar{\beta}$, with the illegal turn between α and $\bar{\beta}$. Furthermore $f(\alpha) = \alpha\tau$, $f(\beta) = \beta\tau$, and "folding the INP ρ" to produce a representative f' is defined as in [BH, p. 21]—namely, folding the initial edge of $\bar{\alpha}$ against that of β (or appropriate germs in either case). If the L_r length of the folded edge is l and ρ' is the resulting INP in $\mathcal{X}'$, then $L_r(\rho') = L_r(\rho) - 2l$ and $L_r(\mathcal{X}) = L_r(\mathcal{X}) - l$.

Each fold is either a partial fold, in which each of the edges is subdivided, or a full fold. Full folds are of two types: those that fold a full edge against another full edge (full–full) and those that fold a full edge against a partial edge (full–partial). Partial and full–partial folds preserve both 2.8(i) and 2.8(ii). That 2.8(i) is preserved is straightforward (note that no real tightening could occur because Λ is minimal). Since partial and full–partial folds do not identify points in $\mathcal{H}_r \cap \mathcal{X}_{r-1}$, any pretrivial path would have to be a loop, an impossibility since f is a homotopy equivalence. Thus 2.8(ii) is also preserved by partial and full–partial folds. Finally, the proof of [BH, Lemma 5.9] generalizes to show that since $\Lambda = \Lambda_{\min}$, 2.8(iii) follows from 2.8(i). Thus, for efficient maps, partial and full–partial folds produce efficient maps.

Full–full folds preserve 2.8(i), since the strata and the maximal initial and terminal germs that would become new edges after a core subdivision are exactly the same after a full–full fold as before. But full–full folds may destroy 2.8(ii) by introducing inessential connecting paths. Restoring 2.8(ii) by folding out inessential connecting paths of height $s < r$ as in [BH, Lemma 5.14] (and step 1 of the proof of 2.12 below) produces an RTT map f^* whose rth stratum is the same as that of f'. Thus, in all cases, folding the INP ρ produces a metric RTT—f^* in the case of full–full folds, and f' otherwise. Each triple $(\mathcal{X}, f, \rho)$, where ρ is an INP in a growing stratum of the metric

RTT map f, thereby generates an infinite chain of such triples, denoted by $\mathcal{W}(\mathcal{X}, f, \rho)$ by iterating this process.

Step 2: Since f is stable, each fold encountered in the construction of $\mathcal{W}(\mathcal{X}, f, \rho)$ is a full fold.

The argument, as in [BH, Lemma 5.17], is that otherwise one could subdivide and collapse an initial segment of τ to produce a representative in which the INP corresponding to ρ lies in a nongrowing stratum, thus reducing $N(f)$ and contradicting the stability of f.

Step 3: In $\mathcal{W}(\mathcal{X}, f, \rho)$, only finitely many ratios $L_r(\rho)/L_r(\mathcal{X})$ are encountered.

For any particular representative $f: \mathcal{X} \to \mathcal{X}$, there are only finitely many ratios for *all* INPs, since there are only finitely many INPs. In fact, the ratios achieved will depend only on the definition of f on the real edges, since they depend only on the placement of the fixed points (determined by how each edge covers itself) and the measure of length (determined by $M(f)_r$). Since all folds in the construction of $\mathcal{W}(\mathcal{X}, f, \rho)$ are full, the number of edges is bounded. The argument of [BH, Lemma 3.7] then shows that finitely many f can arise.

Step 4: If no full–full folds are encountered in $\mathcal{W}(\mathcal{X}, f, \rho)$, then $L_r(\rho) = 2L_r(\mathcal{X})$.

Each full–partial fold produces a neat, stable relative train-track map (without tightening); thus, if $(\mathcal{X}', f', \rho')$ is obtained from $(\mathcal{X}, f, \rho)$ by folding a path of length l, then $L_r(\mathcal{X}') = L_r(\mathcal{X}) - 2l$ and $L_r(\rho') = L_r(\rho) - l$. If $L_r(\rho) \neq 2L_r(\mathcal{X})$ then the ratio moves strictly away from 2 (the segment from the point $(L_r(\mathcal{X}), L_r(\rho))$ to $(L_r(\mathcal{X}'), L_r(\rho'))$ in the xy plane is parallel to the line $y = 2x$ and both points are off the line). So if $L_r(\rho) \neq 2L_r(\mathcal{X})$, we obtain infinitely many ratios, contradicting step 3.

Comment: It is unclear to us whether this ratio condition will hold in general. If f were an efficient map with an INP ρ and a real edge E not in ρ that could be "unfolded" to give another RTT map, then the ratio condition would fail for this new RTT.

Step 5: If $\mathcal{H}_r$ is growing, there is at most one INP of height r.

If there are two distinct INPs ρ and ρ' of height r, stability implies that they remain distinct through the infinite sequence $\mathcal{W}(\mathcal{X}, f, \rho)$. Since full–full folds reduce the number of edges in $\mathcal{H}_r$, only finitely many can occur and we can find a stage $(\mathcal{X}^{(k)}, f^{(k)}, \rho^{(k)})$ in the sequence $\mathcal{W}(\mathcal{X}, f, \rho)$ beyond which none are encountered. Step 4 then guarantees that the ratios are constantly equal to 2 beyond this point, and the rest of the argument as in [BH, Lemma 3.9] finishes the proof. □

Proof of Theorem 2.12. We begin with a neat stable topological representative $f: \mathcal{X} \to \mathcal{X}$, the existence of which is guaranteed by Proposition 2.6, and perform operations (steps 1 and 2 below) that preserve stability and neatness and which lead in finitely many steps to a relative train-track map. To obtain the ratio condition we fold the INP ρ_r (which is unique by 2.11), moving through the representatives in $\mathcal{W}(\mathfrak{X}, f, \rho)$ until we get beyond all full–full folds (see 2.12, step 5), and repeat step 2 below if necessary to produce the desired efficient representative.

Step 1: There are finitely many r-connecting paths for growing strata $\mathcal{H}_r$ whose images $f(\alpha)$ are homotopically trivial—call them *pretrivial*. As in [BH, Lemma 5.14], they can be folded out, starting with the top stratum. Folding out connecting paths at height r will not introduce any new ones at height r or higher, and each such folding reduces the number of points in $\mathcal{H}_r \cap \mathfrak{X}_{r-1}$. This process will therefore terminate in a stable representative which is injective on r-connecting paths for all growing strata $\mathcal{H}_r$.

Step 2: Perform a core subdivision of each growing stratum, in any order, to produce a representative which preserves r-germs for all growing strata $\mathcal{H}_r$. Core subdivision preserves injectivity on connecting paths. This is obvious on all strata other than the cored stratum $\mathcal{H}_r$; for $\mathcal{H}_r$, we must consider connecting paths that may begin or end with edges in a new stratum that were formerly germs. But if γ_1 and γ_2 are germs and $\alpha = \gamma_1\alpha'\gamma_2$ is a pretrivial path, then $f(\gamma_1) = f(\bar{\gamma}_2)$ and α' is pretrivial. So, if there are any pretrivial r-connecting paths after the coring, there were before. Finally, the proof of [BH, Lemma 5.9] generalizes to show that since $\Lambda = \Lambda_{\min}$; 2.8(iii) follows from 2.8(i). □

3. The Scott Conjecture for Free Products

The problem is to show that for $f: \mathfrak{X} \to \mathfrak{X}$ and a path μ in $\mathfrak{X}$, the Kuroš rank of the fixed-point subgroup of the path-induced automorphism $\pi_1(f, \mu)$ is less than or equal to the Kuroš rank of $\mathfrak{X}$. We begin (as in [BH]) with an analysis that reduces the general case to that in which $\pi_1(f, \mu) = \pi_1(f, v)$, a point-induced automorphism. Let $\eta: \tilde{\mathfrak{X}} \to \mathfrak{X}$ be the universal covering of $\mathfrak{X}$ with covering translation group $\mathbf{T}$. Associated with each lift $\tilde{v}$ of a point v in $\mathfrak{X}$, there is the standard isomorphism

$$\Theta_{\tilde{v}}: \mathbf{T} \to \pi_1(\mathfrak{X}, v) \quad \text{given by } \Theta_{\tilde{v}}(t) = \eta[\tilde{v}, t(\tilde{v})],$$

where $[\tilde{v}, t(\tilde{v})]$ is any path in $\tilde{\mathfrak{X}}$ from $\tilde{v}$ to $t(\tilde{v})$.

3.1. Definition. For any lift $\tilde{f}$ of f, define $\zeta_{\tilde{f}}: \mathbf{T} \to \mathbf{T}$ so that the following diagram commutes:

$$\begin{array}{ccc} \tilde{\mathfrak{X}} & \xrightarrow{\tilde{f}} & \tilde{\mathfrak{X}} \\ {\scriptstyle t}\downarrow & & \downarrow{\scriptstyle \zeta_{\tilde{f}}(t)} \\ \tilde{\mathfrak{X}} & \xrightarrow{\tilde{f}} & \tilde{\mathfrak{X}}. \end{array}$$

3.2. Proposition. *Suppose that $\tilde{f}$ is a lift of f and that $\tilde{v}$ is a lift of v.*

(i) *If* $\mu=\eta[\tilde v,\tilde f(\tilde v)]$, *then the following diagram commutes:*

$$\begin{array}{ccc} \mathbf{T} & \xrightarrow{\Theta_{\tilde v}} & \pi_1(\mathfrak{X},v) \\ \zeta_{\tilde f}\downarrow & & \downarrow \pi_1(f,\mu) \\ \mathbf{T} & \xrightarrow{\Theta_{\tilde v}} & \pi_1(\mathfrak{X},v). \end{array}$$

(ii) *Any* $\pi_1(f,\mu)$ *appears in such a diagram for that lift* $\tilde f$ *of* f *such that* $\tilde f(\tilde v)=\tilde\mu(1)$ *for* $\tilde\mu$ *the lift of* μ *starting at* $\tilde v$.
(iii) $\zeta_{\tilde f}$ *determines a point-induced automorphism under* $\Theta_{\tilde v_0}$ *if and only if* $\tilde f(\tilde v_0)=\tilde v_0$.
(iv) *If* $\tilde v_0$ *is a fixed point for* $\tilde f$ *and* $\alpha=\eta[\tilde v,\tilde v_0]$, *then* $f(\alpha)=\bar\mu\alpha$.
(v) $\mathrm{Fix}(\zeta_{\tilde f})=\{t\mid t\tilde f=\tilde f t\}$.

Proof. Suppose that $\zeta_{\tilde f}(t)=t'$. Then $\tilde f t=t'\tilde f$, and we have

$$\begin{aligned}\pi_1(f,\mu)(\Theta_{\tilde v}(t)) &= \mu\circ f(\eta[\tilde v,t(\tilde v)])\circ\bar\mu\\ &= \eta[\tilde v,\tilde f(\tilde v)]\circ\eta\tilde f[\tilde v,t(\tilde v)]\circ\overline{\eta[t'\tilde v,t'\tilde f(\tilde v)]}\\ &= \eta([\tilde v,\tilde f(\tilde v)]\circ[\tilde f(\tilde v),t'\tilde f(\tilde v)]\circ[t'\tilde f(\tilde v),t'\tilde v])\\ &= \eta([\tilde v,t'\tilde v])\\ &= \Theta_{\tilde v}(t')\\ &= \Theta_{\tilde v}(\zeta_{\tilde f}(t)).\end{aligned}$$

This verifies (i). If $\tilde\alpha=[\tilde v,\tilde v_0]$, then $\tilde f(\tilde\alpha)=[\tilde f(\tilde v),\tilde v_0]=[\tilde f(\tilde v),\tilde v]\circ[\tilde v,\tilde v_0]$, establishing (iv). The remaining statements follow easily from the definitions. □

Thus the path-induced automorphisms $\pi_1(f,\mu)$ fall into sets of conjugate automorphisms (with isomorphic fixed point subgroups) according to the lifts $\tilde f$. Furthermore, the class determined by $\tilde f$ contains a point-induced automorphism if and only if $\tilde f$ has a fixed point.

3.3. PROPOSITION. *If* K-rank$(\mathbf{T},\mathrm{Fix}(\zeta_{\tilde f}))\ge 2$, *then* $\tilde f$ *has a fixed point.*

Proof. Suppose that $\tilde f$ is fixed-point free. We use the strategy of [GT] and [CT] to show that the edges of an appropriate graph can be directed in a way that guarantees rank at most 1. As described in [CT], the universal cover $\tilde{\mathfrak{X}}$ of $\mathfrak{X}$ is composed of countries (which cover factor complexes) and real edges (which cover real edges of $\mathfrak{X}$). Each real edge $\tilde E$ of $\tilde{\mathfrak{X}}$ can be directed by choosing an interior point $\tilde v\in\tilde E$ in the edge and using the direction of the first (partial) edge in the reduced path $[\tilde v,\tilde f(\tilde v)]$; this is well-defined since $\tilde{\mathfrak{X}}$ is simply connected and $f(\tilde v)\neq\tilde v$.

Claim:
(i) *At each vertex that is not in a country, exactly one edge is directed out.*
(ii) *Among the edges incident to a particular country, at most one is directed out.*
(iii) *There is at most one country with no outwardly directed edges.*

Proof of Claim: The condition determining directions on edges works equally well at vertices which are not in countries, verifying (i). If an edge incident to a country is directed away from the country, then—since $\tilde{f}$ carries countries to countries—paths determining directions for all vertices in the given country will leave the country on the unique edge on the way to the image country, verifying (ii). A country has no outwardly directed edges if and only if it is setwise fixed by $\tilde{f}$: if there is more than one of them, choose two such that the real-ended path between them encounters the fewest countries. Since $\tilde{f}$ must permute these countries in an order-preserving way (paths are homotopically unique) there are no outwardly directed edges. But then the path is a sequence of real edges which is stretched over itself by $\tilde{f}$, and must have a fixed point by the Brouwer fixed-point theorem. This establishes (iii), completing the proof of the claim.

Now consider $\hat{\mathfrak{X}} = \tilde{\mathfrak{X}}/\mathrm{Fix}(\zeta_{\tilde{f}})$, the covering space of $\mathfrak{X}$ with fundamental group $\mathrm{Fix}(\zeta_{\tilde{f}})$. Since $t\tilde{f} = \tilde{f}t$ for all $t \in \mathrm{Fix}(\zeta_{\tilde{f}})$, $\hat{\mathfrak{X}}$ inherits edge directions from $\tilde{\mathfrak{X}}$, satisfying properties (i)–(iii), and (iii) guarantees that at most one country in $\hat{\mathfrak{X}}$ is essential (i.e., has nontrivial fundamental group). Collapse each country of $\hat{\mathfrak{X}}$ to a point to form a graph $\hat{X}$. If there are no essential countries, then $\pi_1(\hat{\mathfrak{X}}) = \pi_1(\hat{X})$ and the edge directions induced on X imply that $\pi_1(\hat{X})$ is trivial or free of rank 1 [GT]. If there is an essential country C then $\pi_1(\hat{\mathfrak{X}}) \cong \pi_1(\hat{X}) \star \pi_1(C)$ and $\pi_1(C)$ corresponds to a conjugate of a subgroup of one of the factor groups. But in this case the edge directions on $\hat{X}$ have a sink—the collapsed C—implying that the fundamental group of $\hat{X}$ is trivial. In either case, the proposition follows. □

A restatement of Proposition 3.3 is the following.

3.4. COROLLARY. *If $\phi \in \mathrm{Aut}(G)$ has a realization $f: \mathfrak{X} \to \mathfrak{X}$ with a lift $\tilde{f}$ which is fixed-point free, then*

$$\text{K-rank}(G: \mathrm{Fix}(\phi)) \leq 1 \leq \text{K-rank}(G).$$

It remains therefore to deal with the situation where, for every realization f of ϕ, every lift $\tilde{f}$ to the universal cover has a fixed point. For this case we employ the theory of Section 2.

Let $f: \mathfrak{X} \to \mathfrak{X}$ be an efficient representative of ϕ with the associated stratification

$$\mathcal{C} = \mathfrak{X}_0 \subset \mathfrak{X}_1 \subset \cdots \subset \mathfrak{X}_m = \mathfrak{X}.$$

We assume additionally that all the isolated fixed points of f are vertices; it is easy to check that this does not affect efficiency. The main construction of this section is that of the complex $\mathfrak{F}$ below, which is the analog of the graph Σ in [BH] and of the covering spaces D_ϕ of [GT] and $\tilde{K}$ of [CT]. The complex $\mathfrak{F}$ (or, more precisely, its component $\mathfrak{F}^v$ containing v) is a "core" of the covering space of $\mathfrak{X}$ corresponding to the subgroup $\mathrm{Fix}(\pi_1(f, v))$, in the sense that it is a subcomplex of the covering space that carries the full fundamental group. The construction uses the classification of INPs described in 3.5–3.7. (Recall that, since efficient maps are neat, all real INPs are real-ended and have real-ended images under f.)

3.5. LEMMA. *INPs ρ_r of height r are classified as follows.*
Type G: *$\mathcal{H}_r$ is growing and ρ_r is the unique INP of height r.*
Type L: *$\mathcal{H}_r$ is level and composed of a single edge E, and*
(i) *$\rho_r^{\pm 1} = E\alpha$ for some $\alpha \subseteq \mathcal{X}_{r-1}$ (α may be trivial), or*
(ii) *$\rho_r^{\pm 1} = E\beta\bar{E}$ for some $\beta \subseteq \mathcal{X}_{r-1}$ but not of type* (iii), *or*
(iii) *$\rho_r^{\pm 1} = E\gamma x\bar{\gamma}\bar{E}$ for some infinitesimal x and $\gamma \subseteq \mathcal{X}_{r-1}$.*

3.6. LEMMA. *If $\mathcal{H}_r$ is a level stratum and there exist INPs of height r but none of type* L(i), *then either*
(i) *all INPs of height r are of type* L(ii) *for β a power of a fixed β_0, or*
(ii) *all INPs of height r are of type* L(iii) *for a fixed $\gamma = \gamma_0$ and $x \in K$, K a factor subcomplex.*

3.7. DEFINITION. A *stratum of type G or L(i)* is one containing INPs of type G or L(i) respectively. A *stratum of type L(ii) or L(iii)* is one which contains INPs of type L(ii) or L(iii), respectively, but which is not of type L(i). Strata without INPs are not considered in this classification.

Proof of Lemma 3.5. If $\mathcal{H}_r$ is growing, this follows from 2.11.
If $\mathcal{H}_r$ is level, then f permutes the edges of $\mathcal{H}_r$ (perhaps covering lower strata) and so $f(\sigma)$ contains the same number of edges of $\mathcal{H}_r$ as does σ. Moreover, since $f(\sigma) \sim \sigma$, the sequence of edges in $\mathcal{H}_r$ that occurs in $f(\sigma)$ must be identical with that of σ without cancellation of these edges in reducing $f(\sigma)$. So the matrix M_r is the identity and the irreducibility of M_r shows that $\mathcal{H}_r$ contains just one edge. Irreducibility and the fact that all isolated fixed points of f are vertices of $\mathcal{X}$ implies that $f(E) = E\tau$ (replacing E with $\bar{E}$ if necessary): irreducibility rules out all but the indicated three possibilities. (This generalizes [BH, Prop. 6.3].) □

Proof of Lemma 3.6. All INPs of height r are of the form $E\beta\bar{E}$ for loops β at the terminal vertex v of E, and $f(E) = E\tau$ (see 3.5). Thus

$$\{\beta \mid E\beta\bar{E} \text{ is an INP}\} = \{\beta \mid f(\beta) = \bar{\tau}\beta\tau\} = \mathrm{Fix}(\pi_1(f, \bar{\tau})).$$

Thus we must show that K-rank(Fix($\pi_1(f, \bar{\tau})$)) ≤ 1. If not, then by 3.3 the lift $\tilde{f}$ has a fixed point $\tilde{v}_0$ and by 3.2(iv) there is a path α with $f(\alpha) = \bar{\tau}\alpha$. But then $E\alpha$ is an INP, contradicting the hypothesis. □

3.8. THE COMPLEX $\mathfrak{F}$. $\mathfrak{F}$ is a graph of complexes, and the map $p: \mathfrak{F} \to \mathcal{X}$ is topological except that its restriction to a factor complex is an embedding rather than a homeomorphism.

Vertices of $\mathcal{X}$: The *primary vertices* of $\mathcal{X}$ are fixed points of f which are added as needed (in parentheses) with edges and fixed factor complexes. The *secondary vertices* are those of the form $\hat{v}_r$ added with edges of type L(iii).

Edges of $\mathfrak{X}$: Each stratum $\mathcal{H}_r$ of $\mathfrak{X}$ which is of type G, L(i), L(ii), or L(iii) contributes one edge ϵ_r to $\mathfrak{F}$.

Type G: There is a unique INP ρ_r of height r such that ϵ_r joins the fixed endpoints (vertices of $\mathfrak{F}$) of ρ_r and $p(\epsilon_r) = \rho_r$.

Type L(i): Choose one particular INP $\rho_r = E\alpha_0$ (as in L(i) of 3.5) and proceed as for Type G.

Type L(ii): Choose the generating INP $\rho_r = E\beta_0\bar{E}$ (as in (i) of 3.6) and proceed as above.

Type L(iii): ϵ_r is a free edge whose initial point is the initial point of E (a vertex of $\mathfrak{F}$) and whose endpoint is a new vertex $\hat{v}_r$: $p(\epsilon_r) = E\gamma_0$ (as in (ii) of 3.6).

Complexes of $\mathfrak{X}$:

Fixed factor complexes: If C is a factor complex at a fixed vertex v, then $\mathfrak{F}$ includes a factor complex K at v corresponding to $\{$infinitesimals $k \in C \mid \pi_1(f, v)(k) = k\} = \operatorname{Fix}(\pi_1(f, v)) \cap \pi_1(C, v)$. p maps K into C in the natural way.

Conjugate-fixed factor complexes: For each stratum $\mathcal{H}_r$ of type L(iii), $\mathfrak{F}$ includes a factor complex K_r, at the vertex $\hat{v}_r$, corresponding to the factor subcomplex K of 3.6(ii)—that is, corresponding to $\{$infinitesimals $k \mid E\gamma_0 k\bar{\gamma}_0\bar{E}$ is an INP$\}$. Then p maps K_r into the factor complex C containing K in the natural way.

Then $\mathfrak{F}$ is filtered as

$$\mathfrak{F}_0 \subset \mathfrak{F}_1 \subset \cdots \subset \mathfrak{F}_m,$$

with $\mathfrak{F}_0$ composed of the fixed factor complexes, and each level of the filtration is either equal to the preceding or obtained by adding an edge ϵ_r (resp. an edge ϵ_r with a factor complex $\mathcal{K}_r$) to the preceding level according as the stratum $\mathcal{H}_r$ is of type G, L(i), or L(ii) (resp. L(iii)).

The proof of the main result of this section, Proposition 3.11, precedes by induction on the following notion of rank (or Kuroš rank) which is a variant of the negative Euler characteristic defined in [BH].

3.9. Definition. If $\mathcal{Z}$ is a connected graph of complexes with underlying graph Z and q factor complexes, then the *Kuroš rank* of $\mathcal{Z}$ is

$$\text{K-rank}(\mathcal{Z}) = \text{K-rank}(\pi_1(\mathcal{Z})) = \operatorname{rank}(\pi_1(Z)) + q.$$

For a graph of complexes $\mathcal{Z}$ with noncontractible components $\mathcal{Z}_1, \mathcal{Z}_2, \ldots, \mathcal{Z}_p$, the *reduced Kuroš rank* is

$$\tilde{\text{K}}\text{-rank}(\mathcal{Z}) = 1 + \sum_{k=1}^{p} (\text{K-rank}(\mathcal{Z}_k) - 1).$$

3.10. Lemma. *Reduced Kuroš rank is monotone in the sense that for graphs of complexes* $\mathcal{Z}_1 \subseteq \mathcal{Z}_2$,

$$\tilde{\text{K}}\text{-rank}(\mathcal{Z}_1) \leq \tilde{\text{K}}\text{-rank}(\mathcal{Z}_2).$$

The inequality is strict if and only if either

(i) *there is a path in $\mathcal{Z}_2$ meeting $\mathcal{Z}_2 \backslash \mathcal{Z}_1$ whose endpoints lie in noncontractible components of $\mathcal{Z}_1$, or*

(ii) *there is an infinitesimal complex in $\mathcal{Z}_2 \backslash \mathcal{Z}_1$ which is attached at a vertex of a noncontractible component of $\mathcal{Z}_1$.*

Proof. $\mathcal{Z}_2$ is obtained from $\mathcal{Z}_1$ by a sequence of steps, each of which either adds an edge (and perhaps a vertex) or adds an edge with a factor complex. When an edge is introduced the reduced Kuroš rank either remains unchanged (this can happen in several ways) or increased by 1 if the edge introduced has its endpoints in noncontractible components. If a factor complex is attached, the reduced Kuroš rank will increase unless the complex is attached to a contractible component. □

3.11. PROPOSITION. *With the notation as above and $\mathfrak{X}$ connected,*

$$\tilde{\mathrm{K}}\text{-rank}(\mathfrak{F}) \leq \tilde{\mathrm{K}}\text{-rank}(\mathfrak{X}) = \mathrm{K}\text{-rank}(\mathfrak{X}).$$

Furthermore, if v is a primary vertex of $\mathfrak{F}$ lying in the component $\mathfrak{F}^v$, then the mapping $p_v = p|_{\mathfrak{F}^v}: \mathfrak{F}^v \to \mathfrak{X}$ induces an isomorphism $\pi_1(\mathfrak{F}^v, v) \to \mathrm{Fix}(\pi_1(f, v))$ and

$$\mathrm{K}\text{-rank}(\pi_1(\mathfrak{X}, v), \mathrm{Fix}(\pi_1(f, v))) = \tilde{\mathrm{K}}\text{-rank}(\mathfrak{F}^v) = \mathrm{K}\text{-rank}(\mathfrak{F}^v).$$

Proof. We prove the following by induction on r.

(a) $\tilde{\mathrm{K}}\text{-rank}(\mathfrak{F}_r) \leq \tilde{\mathrm{K}}\text{-rank}(\mathfrak{X}_r)$.

(b) If ν is a nontrivial path between primary vertices in $\mathfrak{F}_r$, then $p(\nu)$ is a nontrivial Nielsen path in $\mathfrak{X}_r$.

(c) If σ is a Nielsen path in $\mathfrak{X}_r$, then there exists a path ν in $\mathfrak{F}_r$ such that $p(\nu) = \sigma$.

Observe that the inequality of the proposition follows from (a) and that (b) and (c) imply that $\pi_1(\mathfrak{F}, v)$ and $\mathrm{Fix}(\pi_1(f, v))$ are isomorphic via $\pi_1(p)$.

For $r = 0$, (a) is trivial since $\mathfrak{F}_0$ consists of the union of the factor subcomplexes corresponding to $\mathrm{Fix}(f|_C)$ as C ranges over the complexes in $\mathfrak{C}_0 = \mathfrak{X}_0$. To verify (b), it suffices to observe that any nontrivial path x in $\mathfrak{F}_0$ is of the form $x = x_1 x_2 \ldots x_q$ (where each x_i is an infinitesimal in a single factor complex) and that $p(x) = x$. To verify (c), note that any Nielsen path in $\mathfrak{X}_0$ is an infinitesimal path of the type just described.

Suppose $r > 0$ and that $\tilde{\mathrm{K}}\text{-rank}(\mathfrak{F}_r) = \tilde{\mathrm{K}}\text{-rank}(\mathfrak{F}_{r-1}) + 1$ (the only case in which (a) is not immediate). As noted in 3.10, this means that either the endpoints of ϵ_r lie in noncontractible components of $\mathfrak{F}_{r-1}$, or that $\mathfrak{F}_r$ contains an infinitesimal complex not present in $\mathfrak{F}_{r-1}$ which is attached to a noncontractible component of $\mathfrak{F}_{r-1}$. In either case, it can be checked that $p(\epsilon_r)$ will have its endpoints in noncontractible components of $\mathfrak{X}_r$, so that $\tilde{\mathrm{K}}\text{-rank}(\mathfrak{X}_r) > \tilde{\mathrm{K}}\text{-rank}(\mathfrak{X}_{r-1})$, verifying (a).

To prove (b) and (c), we recall that a real Nielsen path can be decomposed uniquely into a product of INPs. If there is no INP of height r then there is nothing to prove, so suppose that ρ_r has height r and consider growing and level strata separately. If $\mathcal{H}_r$ is growing then ρ_r is the unique INP of height r and $p(\epsilon_r)=\rho_r$, whence (c) follows. To verify (b), observe that for any $\nu\in\mathfrak{F}_r$, $p(\nu)$ is of the form $\beta_0\rho_r^{\pm1}\beta_1\rho_r^{\pm1}\dots\rho_r^{\pm1}\beta_q$; the fact that $p(\nu_r)$ is nontrivial follows from the fact that ρ_r has edges in $\mathcal{H}_r$; the β_js are nontrivial by the induction hypothesis.

If $\mathcal{H}_r$ is level and $\mathcal{H}_r$ is contained in $\mathrm{Fix}(f)$ a similar argument applies.

We finally consider the case of a level $\mathcal{H}_r$ which contains a single edge E with $f(E)=E\tau$, τ nontrivial. An easy cancellation argument [BH, proof of Prop. 6.3] shows that if ν is a path in $\mathfrak{F}_r$ not in $\mathfrak{F}_{r-1}$ then no occurrence of E can be cancelled in reducing $p(\nu)$, proving (b). To prove (c), we consider the three types of level strata separately. If $\mathcal{H}_r$ is of type L(i), then all INPs of height r are of the form $E\alpha$ or $E\beta\bar{E}$ for $\alpha,\beta\subseteq\mathcal{X}_{r-1}$. Then $E\alpha=(E\alpha_0)(\bar{\alpha}_0\alpha)=p(\epsilon_r)(\bar{\alpha}_0\alpha)$ and $E\beta\bar{E}=(E\alpha_0)(\bar{\alpha}_0\beta\alpha_0)(\bar{\alpha}_0\bar{E})=p(\epsilon_r)(\bar{\alpha}_0\beta\alpha_0)p(\bar{\epsilon}_r)$, and $(\bar{\alpha}_0\alpha)$ and $(\bar{\alpha}_0\beta\alpha_0)$ are NPs which by induction are images of $p|_{\mathfrak{F}_{r-1}}$. This verifies (c) for type-L(i) strata. For type-L(ii) strata, (c) is immediate because all INPs of height r are powers of ρ_r. For type L(iii) strata, all INPs have the form of 3.5(iii) and are images of loops $\epsilon_r x\bar{\epsilon}_r$ for infinitesimals $x\in\mathcal{K}_r$. This completes the induction, verifying (a), (b), and (c).

It remains to prove that K-rank$(\pi_1(\mathcal{X},v),\mathrm{Fix}(\pi_1(f,v))=\tilde{\mathrm{K}}$-rank$(\mathfrak{F}^v)$. From the discussion in the introduction, it is enough to show that for any path γ from v to v_C, $\mathrm{Fix}(\pi_1(f,v))\cap\gamma\pi_1(C,v_C)\bar{\gamma}$ is conjugate to the image under $\pi_1(p)$ of some $\mu\pi_1(K,v_K)\bar{\mu}$ where μ is a path in $\mathfrak{F}$ from v to v_K. An easy inductive argument, using (c), verifies this. (It should be observed that without the above arguments, we only know that K-rank$(\pi_1(\mathcal{X},v),\mathrm{Fix}(\pi_1(f,v))\le\tilde{\mathrm{K}}$-rank$(\mathfrak{F}^v)$. However, for our application this inequality actually suffices.) $\square$

Note: Two fixed points of any map f are said to be *Nielsen equivalent* if they are joined by a path which is endpoint homotopic to its image under f. It is clear that two primary vertices of $\mathfrak{F}_r$ lie in the same component of $\mathfrak{F}_r$ if and only if they are Nielsen equivalent in $\mathcal{X}_r$.

Our main result—the Scott conjecture for free products—now follows easily.

3.12. THEOREM. *Let $G=\bigstar_{i=1}^m G_i$ be the free product of m freely indecomposable factors and let ϕ be an automorphism of G. Then the Kuroš subgroup rank of* $\mathrm{Fix}(\phi)$ *is at most the Kuroš rank of G.*

Proof. Either K-rank$(G,\mathrm{Fix}(\phi))\le 1$ or there is an efficient representation $f:\mathcal{X}\to\mathcal{X}$ of ϕ such that

$$\text{K-rank}(G,\mathrm{Fix}(\phi))=\text{K-rank}(\pi_1(\mathcal{X},v),\mathrm{Fix}(\pi_1(f,v)).$$

Let $\mathfrak{F}$ be the complex constructed above and $\mathfrak{F}^v$ the component of $\mathfrak{F}$ containing the primary vertex v. Then

$$\begin{aligned}\text{K-rank}(\pi_1(\mathcal{X},v),\mathrm{Fix}(\pi_1(f,v)))&=\tilde{\text{K}}\text{-rank}(\mathfrak{F}^v)\le\tilde{\text{K}}\text{-rank}(\mathfrak{F})\\&\le\tilde{\text{K}}\text{-rank}(\mathcal{X})=\text{K-rank}(G).\quad\square\end{aligned}$$

References

[BL] R. Baer and F. Levi, *Freie Produkte und ihre Untergruppen,* Compositio Math. 3 (1936), 391–398.

[BH] M. Bestvina and M. Handel, *Train tracks and automorphisms of free groups,* Ann. of Math. (2) 135 (1992), 1–51.

[Co] D. E. Cohen, *Combinatorial group theory: a topological approach,* London Math Soc. Stud. Texts, 14, Cambridge Univ. Press, Cambridge, 1989.

[CT] D. J. Collins and E. C. Turner, *Free product fixed points,* J. London Math. Soc. (2) 38 (1988), 67–76.

[Fo] D. I. Fouxe-Rabinovitch, *Über die Automorphismengruppen der freien Produkte, I,* Mat. Sb. (N.S.) 8 (1940), 265–276; *II,* Mat. Sb. (N.S.) 9 (1941), 183–220.

[GT] R. Z. Goldstein and E. C. Turner, *Fixed subgroups of homomorphisms of free groups,* Bull. London Math. Soc. 18 (1986), 468–470.

[Ku] A. G. Kuroš, *Die Untergruppen der freien Produkte von beliebigen Gruppen,* Math. Ann. 109 (1934), 647–660.

[Se] E. Seneta, *Non-negative matrices and Markov chains,* Springer, New York, 1981.

[Sp] E. Spanier, *Algebraic topology,* McGraw-Hill, New York, 1966.

[St] J. Stallings, *Topology of finite graphs,* Invent. Math. 71 (1983), 551–565.

D. J. Collins
School of Mathematical Sciences
Queen Mary and Westfield College
London E1 4NS
United Kingdom

E. C. Turner
Department of Mathematics
State University of
New York at Albany
Albany, NY 12222