

## On the linearity problem for mapping class groups

Tara E. Brendle  
Hessam Hamidi-Tehrani

### Abstract

Formanek and Procesi have demonstrated that  $\text{Aut}(F_n)$  is not linear for  $n \geq 3$ . Their technique is to construct nonlinear groups of a special form, which we call *FP-groups*, and then to embed a special type of automorphism group, which we call a *poison group*, in  $\text{Aut}(F_n)$ , from which they build an FP-group. We first prove that poison groups cannot be embedded in certain mapping class groups. We then show that no FP-groups of any form can be embedded in mapping class groups. Thus the methods of Formanek and Procesi fail in the case of mapping class groups, providing strong evidence that mapping class groups may in fact be linear.

**AMS Classification** 57M07,20F65; 57N05,20F34

**Keywords** Mapping class group, linearity, poison group

## 1 Introduction

The question of whether mapping class groups are linear has been around for some time. The recent work of Bigelow [2] and also Krammer [14] in determining that the braid group is linear has renewed interest in the subject, due to the close relationship between mapping class groups and braid groups. Let  $S_{g,b;n}$  denote a surface of genus  $g$  with  $b$  boundary components and  $n$  fixed points. Let  $\mathcal{M}_{g,b;n}$  denote the mapping class group of  $S_{g,b;n}$ . We assume throughout that maps fix boundary components pointwise. Bigelow and Budney [3] and independently Korkmaz [13] recently determined that  $\mathcal{M}_{2,0,0}$  is linear. Korkmaz also showed in [13] that mapping class groups contain very large linear subgroups, namely, the hyperelliptic subgroups. However, the question of linearity remains open for mapping class groups of surfaces of genus 3 or greater.

Let  $F_n$  denote the free group of rank  $n$ . It is well known that  $\text{Out}(F_2)$  and  $\text{Aut}(F_2)$  are linear. The former fact is due to Nielsen [17], and the latter follows

by [7] from the linearity of the 4-string braid group  $B_4$ , which is due to Krammer [14].

On the other hand, Formanek and Procesi demonstrated in [9] that  $\text{Aut}(F_n)$  is not a linear group for  $n \geq 3$ . A simple corollary of this result is that  $\text{Out}(F_n)$  is not linear for  $n \geq 4$ . The well-known fact due to Nielsen [15] that  $\mathcal{M}_{g,0,0}$  is isomorphic to  $\text{Out}(\pi_1(S_{g,0,0}))$  suggests that it may be possible to apply the methods of Formanek and Procesi to mapping class groups, though it may not be immediately clear how to do so.

Formanek and Procesi define a class of nonlinear groups, which we will generalize slightly and refer to as *Formanek and Procesi groups*, or *FP-groups* for short. We will show that the existence of FP-subgroups of  $\mathcal{M}_{g,0,1}$  would imply that  $\mathcal{M}_{g+k,0,0}$  is not linear for  $k \geq 1$ . We will also focus our attention on a special kind of automorphism group, which we call a *poison group*. We will describe the particular method of Formanek and Procesi for constructing FP-groups from poison subgroups.

This work originated in an attempt to use the methods of Formanek and Procesi to show that  $\mathcal{M}_{g,0,0}$  is not linear for  $g \geq 3$ . We prove instead that the essential building blocks of the Formanek and Procesi method do not exist in mapping class groups, first in a special case.

**Theorem A** *Poison subgroups cannot be embedded in  $\mathcal{M}_{g,0,1}$ .*

Thus the particular technique of Formanek and Procesi fails to show that certain mapping class groups are not linear. We then generalize this result as follows.

**Theorem B** *FP-groups do not embed in  $\mathcal{M}_{g,b,n}$  for any  $g$ ,  $b$ , and  $n$ .*

Our paper is organized as follows. In Section 2, we give an overview of the methods of Formanek and Procesi for constructing a nonlinear subgroup of  $\text{Aut}(F_n)$  from a poison subgroup. In Section 3, we establish connections between certain mapping class groups and the automorphism group of a closed surface. In Section 4 we prove Theorem A. In Section 5 we prove Theorem B using very different techniques from those used in Section 4. Though Theorem A is a special case of Theorem B, we include a separate proof of Theorem A both for the sake of highlighting the particular construction of Formanek and Procesi and also because the methods used are interesting in their own right. The reader should note, however, that Sections 3, 4, and 5 are completely independent of one another. For example, the reader interested only in Theorem B could read Sections 1, 2, and 5 without any loss of continuity.

**Acknowledgements** The authors would like to express their sincere gratitude to Joan Birman and Alex Lubotzky for suggesting the search for poison subgroups in mapping class groups, and also to Matthew Zinno for helping to point us in the other direction. We thank all three, as well as Walter Neumann, Brian Mangum, Gabriel Rosenberg, and Abhijit Champanerkar for many useful discussions. We are also grateful to the referee for many helpful questions and suggestions.

The first author was partially supported under NSF Grant DMS-9973232. The second author was partially supported by PSC-CUNY Research Grant 63463 00 32.

## 2 The method of Formanek and Procesi

Let  $G$  be any group, and let  $H(G)$  denote the following HNN-extension of  $G$ :

$$H(G) = \langle G, t \mid t(g)t^{-1} = (1;g); g \in G \rangle$$

In other words, conjugation by  $t$  in the HNN-extension carries the diagonal subgroup  $G$  onto its second factor. Formanek and Procesi show in the following theorem that such groups exhibit special behavior under a linear representation.

**Theorem 2.1** (Formanek and Procesi, [9]) *Let  $G$  be a group. Then the image of the subgroup  $G$  in  $H(G)$  under any linear representation of  $H(G)$  is nilpotent-by-abelian-by-finite.*

**Corollary 2.2** *Let  $G$  be a group, and  $K$  a normal subgroup of  $H(G)$  such that the image of  $G$  in  $H(G)/K$  is not nilpotent-by-abelian-by-finite. Then  $H(G)/K$  is not linear.*

**Proof** Let  $\rho : H(G)/K \rightarrow GL_N(k)$  be a linear representation where  $k$  is a field. Let  $\pi : H(G) \rightarrow H(G)/K$  be the natural projection map. Then  $\rho \circ \pi$  is a linear representation of  $H(G)$  and hence by Theorem 2.1,  $\rho(G)$  is nilpotent-by-abelian-by-finite. Thus  $\rho$  is not faithful.  $\square$

We will call a group of the type described in Corollary 2.2 a *Formanek and Procesi group*, or *FP-group* for short. We now describe the particular construction of Formanek and Procesi in demonstrating the nonlinearity of  $\text{Aut}(F_n)$  for  $n \geq 3$ .

Let  $G$  be any group. Let  $x_1, x_2, x_3$  be elements of  $G$  such that  $hx_1x_2x_3i = F_3$ . Let  $\alpha, \beta \in \text{Aut}(G)$  be two maps such that

- (1)  $\alpha(x_j) = x_j, \beta(x_j) = x_j, \forall j = 1, 2,$  and
- (2)  $\alpha(x_3) = x_3\alpha(x_1), \beta(x_3) = x_3\beta(x_2), \forall i = 1, 2.$

We will call the subgroup  $h\alpha, \beta i$  a *poison subgroup* of  $\text{Aut}(G)$ . We can define poison subgroups of the mapping class group  $\mathcal{M}_{g,0,1}$  analogously, since in this case the mapping class group acts on  $\pi_1(S_{g,0,1})$ . Notice that the second condition implies that  $h\alpha, \beta i = F_2$ . Thus poison groups, being isomorphic to the linear group  $F_2$ , are not themselves a kind of FP-group. However, as the following lemma shows, their existence in an automorphism group  $\text{Aut}(G)$  implies that  $\text{Aut}(G)$  is not linear (hence the name "poison groups", though it suggests a bias towards linearity).

**Lemma 2.3** *Let  $G$  be any group. If  $\text{Aut}(G)$  contains a poison subgroup, then it contains an FP-subgroup isomorphic to  $H(F_2)$ .*

**Proof** Let  $h\alpha, \beta i$  be a poison subgroup in  $\text{Aut}(G)$ . Following Formanek and Procesi's argument in [9], let  $\gamma_i \in \text{Aut}(G)$  denote conjugation by  $x_i$ . Consider the group

$$H = \langle h\alpha, \beta i, \gamma_1, \gamma_2, \gamma_3 \rangle$$

First, note that  $h\alpha, \beta i$  is a normal subgroup of  $H$  since both  $\gamma_1$  and  $\gamma_2$  preserve the subgroup  $hx_1x_2x_3i$ . Now let  $w(a; b)$  denote any non-trivial reduced word in the free group on the letters  $a$  and  $b$ . By definition of a poison subgroup, we know that  $w(\alpha, \beta)(x_i) = x_i$  for  $i = 1, 2$ . This tells us that if  $w(\alpha, \beta)$  is in  $h\alpha, \beta i$ , then  $w(\alpha, \beta)$  must induce conjugation by an element in  $hx_1x_2x_3i = F_3$ , which commutes with  $x_1$  and  $x_2$ . But the only such element is the identity. Hence  $w(\alpha, \beta)$  must be the identity map. But we know this is not the case since

$$w(\alpha, \beta)(x_3) = x_3w(\alpha, \beta)(x_1x_2) \tag{1}$$

This tells us that the images of  $\gamma_1$  and  $\gamma_2 \pmod{h\alpha, \beta i}$  will generate a free group. Clearly, the images of  $\gamma_1$  and  $\gamma_2$  also generate the quotient of  $H$  by  $h\alpha, \beta i$ , and so we have a split exact sequence

$$1 \rightarrow h\alpha, \beta i \rightarrow H \rightarrow \langle \gamma_1, \gamma_2 \rangle \rightarrow 1 \tag{2}$$

Thus the only relations we have in a presentation for  $H$  are given by conjugation, as follows:

$$H = \langle \alpha, \beta, \gamma_1, \gamma_2, \gamma_3 \mid \gamma_i \alpha \gamma_i^{-1} = \alpha, \gamma_i \beta \gamma_i^{-1} = \beta, \gamma_3 \alpha \gamma_3^{-1} = \alpha \gamma_1, \gamma_3 \beta \gamma_3^{-1} = \beta \gamma_2, i; j = 1, 2; \tag{3}$$

Rewriting the second set of relations, we obtain  $\varphi_3(\varphi_i \varphi_j) \varphi_3^{-1} = \varphi_i \varphi_j \quad i = 1, 2$ . Since  $h_{\varphi_1} \varphi_2 i = h_{\varphi_1} \varphi_2 i = F_2$ , we have that  $H = H(F_2)$ , with  $\varphi_3$  playing the role of the element  $t$ . Since  $F_2$  is not nilpotent-by-abelian-by-finite,  $H(F_2)$  is an FP-group.  $\square$

### 3 The connection with mapping class groups

Our motivation for the work in this paper is the following observation, the proof of which we defer to the end of the section.

**Claim 3.1** *If a poison subgroup exists in  $\mathcal{M}_{g,0,1}$  for  $g \geq 2$ , then the groups  $\mathcal{M}_{g+k,0,0}$  are not linear for  $k \geq 1$ .*

We have been abusing terminology a bit by talking about poison subgroups in  $\mathcal{M}_{g,0,1}$  and also in the context of automorphism groups. The distinction between the two contexts is unnecessary for our purposes, as the following lemma shows, since these mapping class groups are isomorphic to automorphism groups.

**Lemma 3.2**  $\mathcal{M}_{g,0,1} = \text{Aut}(\varphi_1(S_{g,0,0}))$ ; for  $g \geq 2$ :

**Proof** We begin with the exact sequence

$$1 \rightarrow \text{Inn}(\varphi_1(S_{g,0,0})) \rightarrow \text{Aut}(\varphi_1(S_{g,0,0})) \rightarrow \text{Out}(\varphi_1(S_{g,0,0})) \rightarrow 1.$$

By the well-known theorem of Nielsen [15], we have that  $\text{Out}(\varphi_1(S_{g,0,0})) = \mathcal{M}_{g,0,0}$ . In addition, since  $\varphi_1(S_{g,0,0})$  is centerless, we can replace  $\text{Inn}(\varphi_1(S_{g,0,0}))$  with  $\varphi_1(S_{g,0,0})$  (see, for example, [6]) to obtain

$$1 \rightarrow \varphi_1(S_{g,0,0}) \rightarrow \text{Aut}(\varphi_1(S_{g,0,0})) \rightarrow \mathcal{M}_{g,0,0} \rightarrow 1. \tag{4}$$

By [4], we also have the following exact sequence:

$$1 \rightarrow \varphi_1(S_{g,0,0}) \rightarrow \mathcal{M}_{g,0,1} \rightarrow \mathcal{M}_{g,0,0} \rightarrow 1. \tag{5}$$

Every short exact sequence  $1 \rightarrow N \rightarrow E \rightarrow G \rightarrow 1$  induces a homomorphism  $G \rightarrow \text{Out}(N)$ , defined as follows. Let  $g \in G$ , and let  $e_g$  be a lift of  $g \in E$ . Now,  $E$  acts on  $N$  by conjugation, hence we can think of  $e_g$  as an element of  $\text{Aut}(N)$ . However, since  $N$  is not necessarily abelian, this map is only well defined up to conjugation by an element of  $N$ . Thus we get a map  $G \rightarrow \text{Out}(N)$ . According to Corollary 6.8 of [6], given any short exact sequence as

above, with  $N$  centerless, there is a unique "middle group"  $E$  corresponding to any given homomorphism  $G \rightarrow \text{Out}(N)$ .

In Sequence 4 above, it is clear that the map induced is the Nielsen isomorphism between  $M_{g,0,0}$  and  $\text{Out}(\pi_1(S_{g,0,0}))$ . In Sequence 5, as discussed in [4], the image of a generator  $a$  of  $\pi_1(S_{g,0,0})$  is the so-called "spin map" associated to each curve, which induces conjugation by that curve, but can be more easily understood as a product of opposite Dehn twists about the boundary of an annular neighborhood of the curve  $a$ . In other words, if  $\gamma_+$  and  $\gamma_-$  are the two boundary curves, then the spin map associated to the curve  $a$  can be written as  $T_{\gamma_+} T_{\gamma_-}^{-1}$ , where  $T$  denotes the Dehn twist about the curve  $\gamma$ . Let  $\tilde{a} \in M_{g,0,0}$ , and let  $\tilde{a}$  denote a lift of  $a$  in  $M_{g,0,1}$ . Then  $\tilde{a} T_{\gamma_+}^{-1} T_{\gamma_-}^{-1} \tilde{a}^{-1} = T_{\tilde{\gamma}_+}^{-1} T_{\tilde{\gamma}_-}^{-1}$ , which is precisely the spin map associated to  $\tilde{a}$ . Thus, we are simply looking at the action of  $\tilde{a}$  on  $\pi_1(S_{g,0,0})$ , but since  $\tilde{a}$  does not necessarily fix the basepoint,  $\tilde{a}$  is getting mapped to the class of  $\tilde{a}$  in  $\text{Aut}$ , modulo inner automorphisms. In other words, the induced map from  $M_{g,0,0} \rightarrow \text{Out}(\pi_1(S_{g,0,0}))$  is also the Nielsen isomorphism. Now since  $\pi_1(S_{g,0,0})$  has a trivial center, we apply Corollary 6.8 of [6], and the lemma is proved.  $\square$

**Remark 3.3** The isomorphism given in Lemma 3.2 has received some attention in the literature, though perhaps not as much as it deserves. The map itself is the obvious one, namely, any homeomorphism of a surface with one fixed point induces a natural automorphism of the fundamental group of the closed surface with the fixed point taken as base point. From the geometric point of view, it is not immediately clear that this map from  $M_{g,0,1}$  to  $\text{Aut}(\pi_1(S_{g,0,0}))$  should be a surjection, i.e., it is not necessarily obvious that all elements of  $\text{Aut}(\pi_1(S_{g,0,0}))$  should be topologically induced.

**Lemma 3.4** *If  $\text{Aut}(\pi_1(S_{g,0,0}))$  is not linear, then  $M_{g,1,0}$  is not linear.*

Before proving the lemma, we make a few observations. From Chapter 4, Section 1 of [4] and Lemma 3.2 we have the short exact sequence

$$1 \rightarrow \mathbf{Z} \rightarrow M_{g,1,0} \rightarrow \text{Aut}(\pi_1(S_{g,0,0})) \rightarrow 1 \tag{6}$$

We note that  $\mathbf{Z}$  is actually the center of  $M_{g,1,0}$ , generated by a Dehn twist about the boundary curve. Now  $\text{Aut}(\pi_1(S_{g,0,0}))$  is the quotient of  $M_{g,1,0}$  by  $\mathbf{Z}$ . In general, the quotient of a linear group is not necessarily linear, but the extra information we have about the kernel in this case will allow us to draw the desired conclusion. The following two theorems are proved in [19]. Note that the term "closed" refers to the Zariski topology.

**Theorem 3.5** *Let  $G$  be a linear group and  $H$  a closed normal subgroup of  $G$ . Then  $G/H$  is also linear.*

**Theorem 3.6** *The centralizer of any subset of a linear group is closed.*

**Proof of Lemma 3.4** Since  $\mathbf{Z}$  is the center of  $\mathcal{M}_{g,1,0}$ , it is normal and also closed by the above. Thus we can apply Theorem 3.5 to the surjection given in Sequence 6, and Lemma 3.4 follows directly.  $\square$

We are now ready to prove the claim.

**Proof of Claim 3.1** Suppose that  $\mathcal{M}_{g,0,1}$  contains a poison subgroup. Then by the isomorphism of Lemma 3.2,  $\text{Aut}(\pi_1(S_{g,0,0}))$  also contains a poison subgroup. Then  $\text{Aut}(\pi_1(S_{g,0,0}))$  is not linear by Lemma 2.3. Now by Lemma 3.4,  $\mathcal{M}_{g,1,0}$  is also not linear. The claim follows from the fact that  $\mathcal{M}_{g,1,0}$  is a subgroup of  $\mathcal{M}_{g+k,0,0}$ , for  $k \geq 1$ . Although this fact is well-known, for the sake of completeness we include a proof as follows. Consider  $S_{g,1,0}$  as a subsurface of  $S_{g+k,0,0}$ . Let  $h$  be the homomorphism from  $\mathcal{M}_{g,1,0}$  to  $\mathcal{M}_{g+k,0,0}$  defined by extension to the identity on  $S_{g+k,0,0} \setminus S_{g,1,0}$ . Let  $f \in \ker(h)$  such that  $f \neq id$ . The mapping class  $h(f)$  of  $S_{g+k,0,0}$  keeps the subsurface  $S_{g,1,0}$  invariant up to isotopy. According to Section 7.5 in [11],  $h(f)$  induces a well defined mapping class in  $\pi_0(\text{Di}(S_{g,1,0}))$  (the group of homeomorphisms of  $S_{g,1,0}$  up to isotopy not necessarily fixing  $\partial S_{g,1,0}$ ). But since  $h(f) = id$  and by the definition of  $h$ , this implies that  $f$  induces the identity in  $\pi_0(\text{Di}(S_{g,1,0}))$ , which implies that  $f$  could only be a non-trivial power of a Dehn twist in the  $\text{Aut}(\pi_1(S_{g,1,0}))$ . Then by definition,  $h(f)$  will also be a non-trivial power of a Dehn twist, which is a contradiction.  $\square$

**Remark 3.7** We have defined poison subgroups in the context of  $\mathcal{M}_{g,0,1}$  and also in the context of automorphism groups, but the definition also makes sense in the context of any group action on another group. Thus one could use this as a general approach to the linearity question for any such group.

## 4 Poison subgroups cannot be embedded in $\mathcal{M}_{g,0,1}$

Our strategy for proving this result will be to decompose the surface  $S = S_{g,0,0}$  into subsurfaces in a particular way. We then use the machinery of graphs of groups (described in detail in [1]) to analyze the action of the generators of a

poison subgroup of  $\mathcal{M}_{g,0,1}$  on the elements  $x_1; x_2; x_3 \in \pi_1(S)$ . After completion of the proof of Theorem A, we discovered that similar methods involving graphs of groups and normal forms were used by Levitt and Vogtmann in [12] to give an algorithm for the Whitehead problem for surface groups. There is a major difference, however, in that we are not given the curves  $x_1; x_2;$  and  $x_3$ , and hence we cannot apply their algorithm directly, nor would our proof be significantly shortened by direct reference to their results. Thus we have kept the proof of Theorem A in its original form for the sake of self-containment. We have, however, found it useful to adopt their methods for the decomposition of the surface  $S$ .

Throughout this section assume that  $g \geq 2$ , since Theorem A is clear when  $g = 1$ . Fix a point  $p \in S$ , and identify  $S_{g,0,1}$  with  $(S; p)$ . We use the point  $p$  as the base point for the fundamental group of  $S$ . Let  $h = \langle x_1; x_2 \rangle$  be a poison subgroup in  $\mathcal{M}_{g,0,1}$ . Then there are elements  $x_1; x_2; x_3 \in \pi_1(S; p)$  such that  $hx_1; x_2; x_3 = F_3$  and

- (1)  $\sigma_i(x_j) = x_j \quad i; j = 1; 2$ , and
- (2)  $\sigma_i(x_3) = x_3 x_i \quad i = 1; 2$ .

In what follows, we will choose appropriate representatives for  $\sigma_i$  and  $x_j$  (denoted by the same names by abuse of notation) such that, among other things, a power of  $\sigma_i$  fixes a regular neighborhood of  $x_j$  pointwise. To this end our main tool will be the following result of Hass and Scott [10]. For  $y_1; y_2 \in \pi_1(S; p)$ , let

$$\text{Stab}(y_1; y_2) = \langle \sigma_i \in \mathcal{M}_{g,0,1} \mid \sigma_i(y_j) = y_j; \quad i = 1; 2; g \rangle$$

**Lemma 4.1** *Let  $y_1; y_2$  be distinct elements of  $\pi_1(S; p)$ , which are not proper powers. Then there exists a representative of  $y_i$  (denoted by  $y_i$ ) and a subsurface  $A$  formed by a regular neighborhood  $N$  of  $y_1 \cup y_2$  together with all disk components of  $S \setminus N$ , such that, for any  $\sigma \in \text{Stab}(y_1; y_2)$ ,  $\sigma$  has a representative homeomorphism  $\tilde{\sigma}$  such that  $\tilde{\sigma}(A) = A$ .*

This lemma follows from Theorem 2.1 in [10] together with the discussion in the beginning of page 32 in the same paper. For further details see Section 2.1 in [12].

**Remark 4.2** Notice that, in Lemma 4.1, if  $\sigma \in \text{Stab}(y_1; y_2)$ , the map  $\sigma$  induces a unique mapping class in  $\pi_0(\text{Diff}(A; \partial A))$  (see Section 7.5 in [11]).

Since it is possible that  $x_1$  and  $x_2$  are proper powers, we need the following well-known lemma, adapted from [12].



**Lemma 4.3** *Given a nontrivial element  $x \in \pi_1(S; \mathbb{R})$ , there exists a unique  $y \in \pi_1(S; \mathbb{R})$  and a unique  $t \geq 1$  such that  $y$  is not a proper power and  $x = y^t$ .*

**Proof** A proof is given in [12] (Lemma 2.3). Though we will not give details, we note that it is also possible to prove this lemma by elementary hyperbolic geometry, using the discrete action of  $\pi_1(S; \mathbb{R})$  on the upper half plane by hyperbolic isometries.  $\square$

**Corollary 4.4** *Let  $z_1, z_2 \in \pi_1(S; \mathbb{R})$  be such that  $z_1^N = z_2^N$  for some  $N \geq 1$ . Then  $z_1 = z_2$ .*

**Proof** Using Lemma 4.3 let  $y_i^{t_i} = z_i$  such that  $y_i$  is not a proper power and  $t_i \geq 1$ , for  $i = 1, 2$ . Let  $x = y_1^{t_1 N} = y_2^{t_2 N}$ . By the uniqueness guaranteed by Lemma 4.3, we have  $y_1 = y_2$  and  $t_1 N = t_2 N$ . Hence  $z_1 = z_2$ , as desired.  $\square$

Using Lemma 4.3, we can choose elements  $y_i$  which are not proper powers and  $t_i \geq 1$  such that  $x_i = y_i^{t_i}$  for  $i = 1, 2$ . Then we know that  $\pi_1(y_j^{t_j}) = y_j^{t_j}$ , which implies that  $\pi_1(y_j) = y_j$ , by Corollary 4.4. Notice that  $y_1$  and  $y_2$  are distinct since  $\langle x_1, x_2 \rangle = F_2$ . We choose  $y_i$  and  $A$  according to Lemma 4.1. Let  $\pi_0(\text{Di}(S; A))$  be the subgroup of  $M_{g,0,1}$  consisting of mapping classes which have a representative keeping  $A$  fixed pointwise. We now adapt Lemma 3.1 of [12] to our purposes, and repeat their argument nearly verbatim.

**Lemma 4.5** *The subgroup  $\pi_0(\text{Di}(S; A))$  has finite index in  $\text{Stab}(y_1, y_2)$ .*

**Proof** First note that  $A$  is not an annulus, since  $x_1$  and  $x_2$  generate a free group. Using Lemma 4.1 (and noting Remark 4.2), we can define a map from  $\text{Stab}(y_1, y_2)$  to  $\pi_0(\text{Di}(A; \mathbb{R}))$ . Now we claim that the image of  $\pi_0$  is finite. To see this, let  $k$  be any positive integer. Let  $T_k$  denote the set of homotopy classes of simple closed curves in  $A$  whose intersection number with  $y_1$  and  $y_2$  is at most  $k$ . Then  $T_k$  is finite, since  $\text{An}(y_1, y_2)$  is composed entirely of disks and annuli. Any map  $\phi \in \text{Stab}(y_1, y_2)$  will preserve the intersection number of a curve with  $y_1$  and  $y_2$ , and hence  $\text{Stab}(y_1, y_2)$  acts on the set  $T_k$ . Now choose a finite set  $W$  of simple closed curves in  $A$  whose image completely determines an element of  $\pi_0(\text{Di}(A; \mathbb{R}))$ . Let  $k$  be bigger than the intersection number of any element in  $W$  with  $y_1$  and  $y_2$ . Thus the class of  $\phi$  restricted to  $A$  in  $\pi_0(\text{Di}(A; \mathbb{R}))$  is completely determined by the action of  $\phi$  on  $T_k$ . But the set of permutations of  $T_k$  is finite, and hence the image of  $\text{Stab}(y_1, y_2)$  under  $\pi_0$  is finite.

Now let  $\rho : \pi_0(\text{Di}(A; \gamma)) \rightarrow \text{Out}(\pi_1(A; \gamma))$  be the natural homomorphism. The image of  $\rho$  is also finite by the above argument. Now any element  $z \in \ker(\rho)$  induces an inner automorphism on  $\pi_1(A; \gamma)$ , i.e.,  $\rho(z) = czc^{-1}$ . The element  $c$  has to commute with both  $y_1$  and  $y_2$ , which implies that  $c$  has to be a power of both  $y_1$  and  $y_2$  since the centralizer of an element in a surface group is cyclic (this is an exercise in elementary hyperbolic geometry), and  $y_1$  and  $y_2$  are not proper powers. But this implies that  $c = 1$  since  $x_1$  and  $x_2$  generate a free group. Hence  $\rho$  induces the identity on  $\pi_1(A; \gamma)$ . Picking a set of simple generators for  $\pi_1(A; \gamma)$ , one can use an isotopy of the surface to make sure that  $\rho$  keeps them fixed pointwise, by [8]. Then one can further isotope to make sure  $\rho$  keeps  $A$  invariant pointwise by Alexander's lemma [18]. Hence  $\ker(\rho)$  is contained in  $\pi_0(\text{Di}(S; A))$ , which proves the lemma.  $\square$

**Proposition 4.6** *There exists an integer  $M$  such that  $\rho_i^M$  fixes  $A$  pointwise (up to isotopy).*

**Proof** We know  $\rho_i \in \text{Stab}(y_1; y_2)$  for  $i = 1; 2$ . Hence by Lemma 4.5, there is an integer  $M_i \geq 0$  such that  $\rho_i^{M_i} \in \pi_0(\text{Di}(S; A))$ . Letting  $M = \text{LCM}(M_1; M_2)$ , we have  $\rho_i^M \in \pi_0(\text{Di}(S; A))$  for  $i = 1; 2$ .  $\square$

From this point on, we assume that we are working with the particular representative of  $\rho_i^M$  which fixes  $A$  pointwise.

Let  $B_1; \dots; B_r$  be the respective closures of each component of  $S \setminus A$ . Each component is  $B_j$  attached to  $A$  along one or more circles. Hence  $A \setminus B_j$  consists of  $n_j - 1$  circles, which we denote by  $\gamma_{j;1}; \dots; \gamma_{j;n_j}$ .

In what follows we will use this decomposition of  $S$  into the subsurfaces  $A; B_j$  to construct a graph of groups  $G$  whose fundamental group will give a decomposition of  $\pi_1(S; \gamma)$ . To that end, we introduce some notation.

For an oriented arc  $e$  let  $\text{start}(e)$  and  $\text{end}(e)$  be the starting and ending points of the arc  $e$ , respectively. Also, let  $e^{-1}$  be the same arc with the opposite orientation. In the following discussion, let the pair of indices  $j; k$  be such that  $1 \leq j \leq r$ , and  $1 \leq k \leq n_j$ .

Choose base points  $b_j \in B_j$ . Notice that  $\rho_i^M$  fixes each  $B_j$  setwise. Hence we further isotope  $\rho_i^M$  so that it fixes  $b_j$ , for  $i = 1; 2$ . See Figure 1.

Choose oriented arcs  $e_{j;k}$  connecting  $\gamma_{j;1}$  to  $b_j$  for  $1 \leq j \leq r$  and  $1 \leq k \leq n_j$ . Choose each arc  $e_{i;j}$  such that it intersects  $\gamma_{j;k}$  exactly once, and does not intersect any other  $\gamma$ 's. Moreover, we make the choices in such a way that

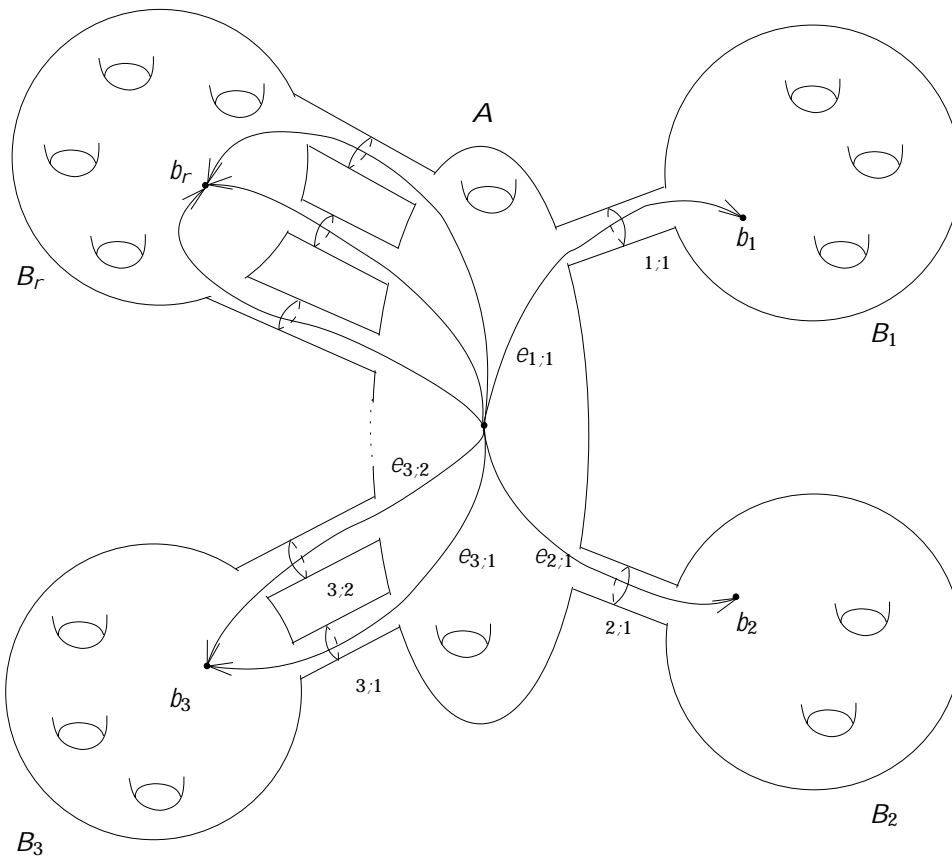


Figure 1: The decomposition of the surface  $S$

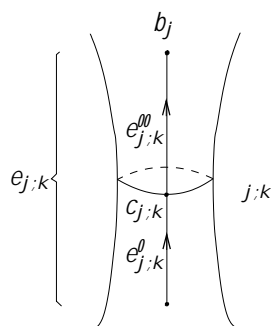


Figure 2: The subarcs of  $e_{j;k}$

if  $(j; k) \not\subseteq (j^0; k^0)$ , then  $e_{j;k}$  and  $e_{j^0;k^0}$  do not intersect except possibly at the endpoints. Let  $c_{j;k}$  be the point of intersection of  $e_{j;k}$  with  $e_{j;k}$ . Also, let  $e_{j;k}^\partial$  be the subarc of  $e_{j;k}$  connecting  $a_j$  to  $c_{j;k}$ , and let  $e_{j;k}^{\partial\partial}$  be the subarc from  $c_{j;k}$  to  $b_j$ . See Figure 2.

Let  $G$  be the graph embedded in  $S$  with vertices  $a_1; \dots; a_r$  and  $b_1; \dots; b_r$  and geometric edges  $e_{j;k}$  as above. As a technical point, the arcs with the opposite orientation  $e_{j;k}$  are also considered edges of the graph  $G$  but not drawn separately.

We use the graph  $G$  to construct a graph of groups. To each vertex of  $G$  we assign the fundamental group of the subsurface in which it is located, namely, to  $a_j$  we assign  $A = \pi_1(A; a_j)$ , to  $b_j$  we assign  $B_j = \pi_1(B_j; b_j)$ . To each edge  $e_{j;k}$  we assign  $\pi_1(j; k; c_{j;k}) = \mathbf{Z}$ . Also, let  $\pi_{e_{j;k}} = \pi_{e_{j;k}}$ . We also have natural injections of the edge groups into the adjoining vertex groups as follows: for any  $e_{j;k}$ , since  $start(e_{j;k}) = a_j$ , the vertex group for  $start(e_{j;k})$  is  $A$ . We have  $\pi_{e_{j;k}} : \pi_{e_{j;k}} \rightarrow A$  defined by  $\pi_{e_{j;k}}(x) = e_{j;k}^\partial x e_{j;k}^\partial$ . Corresponding to  $end(e_{j;k})$ , we have  $\pi_{e_{j;k}} : \pi_{e_{j;k}} \rightarrow B_j$  which is defined by  $\pi_{e_{j;k}}(x) = e_{j;k}^{\partial\partial} x e_{j;k}^{\partial\partial}$ . For the edges  $e_{j;k}$  set  $\pi_{e_{j;k}} = \pi_{e_{j;k}}$  and  $\pi_{e_{j;k}} = \pi_{e_{j;k}}$ .

Let  $G$  be the graph of groups constructed by the above data. By the generalized Van Kampen theorem,  $\pi_1(S; a_1)$  is isomorphic to the fundamental group of the graph of groups  $\pi_1(G; a_1)$ .

To understand the elements of  $\pi_1(G; a_1)$ , we quote some definitions from [1]. A loop based at  $a_1$  in  $G$  is a sequence

$$t = (g_0; i_1; g_1; \dots; i_n; g_n)$$

where  $i_i$  are edges of  $G$  and  $(i_1; \dots; i_n)$  is a loop in  $G$  with  $start(i_1) = a_1$  and  $end(i_n) = a_1$ . Also,  $g_0$  and  $g_n$  are in  $A$ , and for  $0 < i < n$ , each  $g_i$  is in the group assigned to  $end(i_i) = start(i_{i+1})$ . A loop  $t$  in  $G$  is reduced if either  $n = 0$  and  $g_0 \neq 1$ , or  $n > 0$  and whenever  $i_{i+1} = i_i$ , we have  $g_i \neq i_i(i_i)$ . Geometrically, one can think of  $t$  as a loop in  $S$ , with  $g_i$  being loops in respective subsurfaces, and  $i_i$  as arcs connecting these loops. From this point of view, a reduced loop on  $S$  does not "travel" to a component  $B_j$  unnecessarily.

By [1], any non-trivial element of  $\pi_1(G; a_1)$  can be written as  $jtj = g_0 i_1 g_1 \dots i_n g_n$ , where  $t$  is a reduced loop as above.

**Remark 4.7** The reduced loop representing 1 is the empty sequence.

**Remark 4.8** A non-reduced loop can be made into a reduced loop which represents the same element in  $\pi_1(G; a_1)$  by the process of *combing*. Namely, if a loop  $t$  of length  $n > 1$  is not reduced, it has a subsequence of the form

$(g_{i-1}; \dots; g_i; (h_i); \dots)$ . One can replace this subsequence with  $(g_{i-1}; \dots; (h_i))$ . This process reduces the length, so after finitely many steps one arrives at a reduced loop.

The following theorem is proved in [1].

**Theorem 4.9** *Let  $t = (g_0; \dots; g_1; \dots; g_n)$  and  $t^\theta = (g_0^\theta; \dots; g_1^\theta; \dots; g_m^\theta)$  be two reduced loops such that  $jtj = jt^\theta j$  in  $\pi_1(G)$ . Then  $n = m$ ,  $g_i = g_i^\theta$  for  $1 \leq i \leq n$ , and there exist  $h_i \in \pi_1$  such that*

- (1)  $g_0^\theta = g_0; \dots; (h_1)^{-1}$ ,
- (2)  $g_i^\theta = g_i; \dots; (h_i) g_{i+1}; \dots; (h_{i+1})^{-1}$ ,
- (3)  $g_n^\theta = g_n; \dots; (h_n) g_n$ .

Notice that in the above theorem the elements of the form  $(h)$  come from the circles  $e_{j;k}$ .

**Proof of Theorem A** Suppose  $h = \pi_1^{-1}(A)$  is a poison subgroup with respect to  $x_1; x_2; x_3 \in \pi_1(S_{g,0,0})$ . We construct the graph of groups  $G$  as above, with  $\pi_1(G) = \pi_1(S_{g,0,0})$ . In the following we will identify these two groups.

By Proposition 4.6, we can choose an integer  $M$  such that  $\pi_1^M$  fixes  $A$  pointwise. Since  $\pi_1^M$  also sends each  $B_j$  to itself fixing the base points, we can see that  $\pi_1^M(e_{j;k}) = e_{j;k} p_{j;k}$  where  $p_{j;k} \in B_j$ . Similarly  $\pi_1^M(e_{j;k}) = p_{j;k}^{-1} e_{j;k}$ .

We will now simplify notation a bit by letting  $\pi_1^M$  stand for  $\pi_1^M$ . Let  $x_3 = jtj$  where  $t$  is the reduced loop  $t = (g_0; \dots; g_1; \dots; g_{2n}; g_{2n})$ : Notice that since the graph  $G$  is "star-shaped", the length of the loop must be even. Therefore

$$\begin{aligned} (x_3) &= j(g_0; \dots; p_1 (g_1) p_2^{-1}; \dots; g_2; \dots; p_3 (g_3) p_4^{-1}; \dots; \dots; p_{2n-1} (g_{2n-1}) p_{2n}^{-1}; \dots; g_{2n})j \\ &\text{(each } p_i \text{ is in the group which makes this a well-defined path). Now by the} \\ &\text{condition } \pi_1(x_3) = x_3 x_1, \text{ which implies that } (x_3) = x_3 x_1^M, \text{ we get the equality} \\ j(g_0; \dots; p_1 (g_1) p_2^{-1}; \dots; g_2; \dots; p_3 (g_3) p_4^{-1}; \dots; \dots; p_{2n-1} (g_{2n-1}) p_{2n}^{-1}; \dots; g_{2n})j \\ &= j(g_0; \dots; g_1; \dots; g_{2n} x_1^M)j. \end{aligned}$$

Let  $t^\theta$  and  $t^{\theta\theta}$  be the paths appearing on the left and right hand sides of the above equation respectively. Since the path  $t$  is reduced, so is  $t^{\theta\theta}$ . If  $t^\theta$  is not reduced, by Remark 4.8 we can comb it to a reduced path  $t_{\text{red}}^\theta$ . By the equality

and Theorem 4.9,  $t_{\text{red}}^{\theta}$  must have the same length as  $t^{\theta}$ , which means  $t^{\theta}$  was reduced in the first place. Using Theorem 4.9 again, there is an  $h_1 \in \pi_1$  such that  $g_{2n} x_1^M = \pi_1(h_1) g_{2n}$ , i.e.,  $x_1^M = g_{2n}^{-1} \pi_1(h_1) g_{2n}$ . Similarly, using  $\pi_2$  in place of  $\pi_1$ , there exists an  $h_2 \in \pi_2$  such that  $x_2^M = g_{2n}^{-1} \pi_2(h_2) g_{2n}$ . But  $\pi_1 = \mathbf{Z}$ , therefore  $h_1, h_2$  commute, which implies  $x_1^M, x_2^M$  commute. This is a contradiction, since  $hx_1, x_2i = F_2$ .  $\square$

## 5 FP-groups do not embed in mapping class groups

We begin by showing how to narrow our search for an FP-subgroup in a mapping class group.

**Lemma 5.1** *Suppose that  $M_{g;b;n}$  contains an FP-subgroup. Then it contains an FP-subgroup  $H$  which is isomorphic to a quotient of  $H(F_2)$ . Moreover, the image of  $F_2 \curvearrowright \text{flg}$  in  $H$  is isomorphic to  $F_2$ .*

**Proof** Suppose  $M_{g;b;n}$  contains an FP-subgroup. Hence there is a group  $G$  and a homomorphism  $\pi : H(G) \rightarrow M_{g;b;n}$  such that  $(G \curvearrowright \text{flg})$  is not nilpotent-by-abelian-by-finite. Here  $H(G) = H(G) = \ker(\pi)$  is the FP-subgroup of  $M_{g;b;n}$ . By Tits' alternative for mapping class groups ([11] or [16]),  $(G \curvearrowright \text{flg})$  is either abelian-by-finite or contains a subgroup isomorphic to  $F_2$ . By assumption, the latter holds. Let  $x_1, x_2 \in G$  such that  $h(x_1, 1), (x_2, 1)i = F_2$ . Then it is easily seen that for  $G_1 = hx_1, x_2i$ ,  $H(G_1)$  is an FP-subgroup of  $M_{g;b;n}$  and  $G_1 = F_2$ .  $\square$

We now recall the following definition from [11]. A mapping class  $f$  is called *pure* if there exists a set (possibly empty)  $\mathcal{C} = \{c_1, \dots, c_k\}$  of non-parallel, non-trivial, non-intersecting simple closed curves on the surface such that:

- (1) The mapping class  $f$  fixes each curve in  $\mathcal{C}$  up to isotopy.
- (2) The mapping class  $f$  keeps each component of  $S \setminus \mathcal{C}$  invariant up to isotopy.
- (3) The restriction of  $f$  to each component of  $S \setminus \mathcal{C}$  is either the identity or pseudo-Anosov. (Recall that the restriction of  $f$  to a surface  $U$  is pseudo-Anosov if and only if for any non-trivial simple closed curve  $c$  in  $U$  not isotopic to  $\partial U$  and for any  $N > 0$ ,  $f^N(c)$  is not isotopic to  $c$ .)

For an integer  $m$ , let  $H_1(S; \mathbf{Z}=m\mathbf{Z})$  be the first homology group of  $S$  with coefficients in  $\mathbf{Z}=m\mathbf{Z}$ . We have an action of  $M_{g;b;n}$  on  $H_1(S; \mathbf{Z}=m\mathbf{Z})$ , which defines a natural homomorphism  $M_{g;b;n} \rightarrow \text{Aut}(H_1(S; \mathbf{Z}=m\mathbf{Z}))$ . The following theorem is due to Ivanov ([11], 1.8).

**Theorem 5.2** For any integer  $m \geq 3$ , the group

$$\Gamma_m = \ker(M_{g;b;n} \rightarrow \text{Aut}(H_1(S; \mathbf{Z}=m\mathbf{Z})))$$

is a normal subgroup of finite index in  $M_{g;b;n}$  consisting only of pure elements.

In the following discussion we will only need one such subgroup, so we set  $m = 3$  for simplicity. Any value  $m \geq 3$  would work as well.

The reader should note that in the following theorem, the generators  $\alpha_i, \beta_j$ , and  $t$  do not have precisely the same meaning as in Section 2.

**Theorem 5.3** Assume  $M_{g;b;n}$  contains an FP-subgroup. Then there exists an FP-subgroup of the form  $H = \langle \alpha_1, \alpha_2, \beta_1, \beta_2, t \rangle$  such that  $\alpha_1, \alpha_2, \beta_1$  and  $\beta_2$  are in  $\Gamma_3$  (in particular they are pure), and

- (1)  $\alpha_1 \alpha_2 = F_2$ ,
- (2)  $\beta_1$  commutes with  $\beta_2$ ,
- (3)  $t(\alpha_i \beta_j)t^{-1} = \alpha_i \beta_j$ .

**Proof** Let  $H$  be an FP-subgroup of the form  $\langle H(F_2) \rangle$  as in Lemma 5.1, where  $F_2 = h\alpha_1\alpha_2i$ . Let  $\alpha_i = (1; \alpha_i)$  and  $\beta_j = (\alpha_j; 1)$ . By abuse of notation, we denote  $\langle t \rangle$  by  $t$ . Then  $H = \langle \alpha_1, \alpha_2, \beta_1, \beta_2, t \rangle$  is an FP-subgroup satisfying (1) - (3) above, by definition of an FP-subgroup and Lemma 5.1. Using Theorem 5.2,  $\Gamma_3$  is a normal subgroup of  $M_{g;b;n}$  of finite index. Let  $N = [M_{g;b;n} : \Gamma_3]$ . Then  $\alpha_i, \beta_j \in \Gamma_3$  are pure, and  $\alpha_1 \alpha_2 = F_2$ . Replacing each of  $\alpha_i, \beta_j$  with their  $N$ th powers and keeping the same  $t$ , we get an FP-subgroup satisfying the conditions of the theorem. □

In the rest of this paper we assume that  $\alpha_i, \beta_j$  and  $t$  are maps as given in Theorem 5.3.

We can now exploit the machinery of pure mapping classes as developed in [11]. For a pure mapping class  $f$ , one can always find a representative homeomorphism (which we will also denote by  $f$ ) which fixes each curve in  $\mathcal{C}$  and each component setwise. Moreover, the mapping class  $f$  induces well-defined

mapping classes on components of  $S n C$  (see Section 7.5 in [11]). As an important technical point, for a component  $T$  of  $S n C$ , in order to get a well-defined mapping class  $f|_T$  in the mapping class group of  $T$ , one should allow the isotopies in  $T$  to move the points in the components of  $@T$  which are created as a result of cutting  $S$  open. Otherwise, an ambiguity results from combining  $f|_T$  with a Dehn twist in a component of  $@T$ . In other words, when the surface is cut open along  $C$ , all the new boundary components which appear will be dealt with essentially as punctures. The same remark holds when considering the mapping class group of a connected subsurface of  $S$ . In what follows, the phrase "up to isotopy" will usually be dropped, but should be understood in any discussion of topological equivalence.

In the above discussion, the collection  $C$  corresponding to a pure mapping class  $f$  may not be canonical, but in fact one can always choose a canonical collection of isotopy classes of disjoint simple closed curves, denoted by  $(f)$ , which we will define shortly. For two 1-submanifolds  $C_1$  and  $C_2$  of  $S$ , let

$$i(C_1; C_2) = \min \{ j \mid C_1 \setminus C_2 \text{ is isotopic to } C_j \}$$

In other words,  $i(C_1; C_2)$  is the geometric intersection number of  $C_1$  and  $C_2$ . We then define  $(f)$  by saying  $c \in (f)$  if the two following conditions hold:

- (1)  $f(c) = c$ ;
- (2) For any simple closed curve  $c$ , if  $i(c; c) \neq 0$ , then  $f(c) \neq c$ .

The collection  $(f)$  is called the *essential reduction system* for  $f$ . It is proved in [11] (see Chapter 7) that  $(f)$  is a finite collection of disjoint simple closed curves, and  $f$  restricted to each component of  $S n (f)$  is either the identity or pseudo-Anosov.

If  $f \in \mathcal{M}_{g;b;n}$  is not pure, then as discussed above there is some  $N > 0$  such that  $f^N$  is pure. Thus we can extend the definition of essential reduction systems by defining  $(f)$  to be equal to  $(f^N)$ . The notion of an essential reduction system was originally defined in [5] for a mapping class, and was generalized in [11] to an arbitrary subgroup of  $\mathcal{M}_{g;b;n}$ . Note that  $(f)$  is a topological invariant of the mapping class  $f$ . We use this notion to define an invariant for a pair of mapping classes in  $\mathcal{M}_{g;b;n}$ .

**Definition 5.4** For two mapping classes  $f, h \in \mathcal{M}_{g;b;n}$ , we let

$$i(f; h) = i((f); (h)):$$

Notice that this is invariant under simultaneous conjugacy:



**Proposition 5.5** For  $t, f, h \in \mathcal{M}_{g,b;n}$ ,  $i(tft^{-1}; tht^{-1}) = i(f; h)$ .

**Proof** First notice that  $(tft^{-1}) = t(\langle f \rangle)$ , for  $f, t \in \mathcal{M}_{g,b;n}$  (again see [11], Chapter 7). Then we have that

$$\begin{aligned} i(tft^{-1}; tht^{-1}) &= i(\langle tft^{-1} \rangle; \langle tht^{-1} \rangle) \\ &= i(\langle t \langle f \rangle \rangle; \langle t \langle h \rangle \rangle) \\ &= i(\langle f \rangle; \langle h \rangle) \\ &= i(f; h); \quad \square \end{aligned}$$

The invariant  $i(f; h)$  for  $f, h \in \mathcal{M}_{g,b;n}$  will be crucial in the proof of Theorem B. We recall the following lemma, proved in [11].

**Lemma 5.6** (Ivanov) *Let  $f$  be a pure mapping class. If  $X$  is a subsurface or a simple closed curve on the surface such that  $f^N(X) = X$  for some  $N \geq 1$ , then  $f(X) = X$ .*

The following definition is also inspired by [11].

**Definition 5.7** Let  $f \in \mathcal{M}_{g,b;n}$ , and let  $T$  be the isotopy class of a connected subsurface of  $S$ . We say  $f$  keeps  $T$  precisely invariant if  $f(T) = T$  and if  $f(c) \not\subset c$  for each curve  $c$  such that  $i(c; @T) \neq 0$ .

In particular we note that a pure mapping class  $f \in \mathcal{M}_{g,b;n}$  keeps all components of  $S \setminus \langle f \rangle$  precisely invariant, by the basic property of  $\langle f \rangle$ . Similarly,  $f$  keeps each regular neighborhood of  $c \in \langle f \rangle$  precisely invariant. We now develop a series of lemmas to prove Theorem B.

**Lemma 5.8** *Let  $f, g$  be pure mapping classes in  $\mathcal{M}_{g,b;n}$  such that  $f = g$ . Let  $T$  be a component of  $S \setminus \langle f \rangle$ . Then we have*

- (i)  $f(T) = T$ , up to isotopy.
- (ii)  $f(c) = c$  for each  $c \in \langle f \rangle$ .
- (iii)  $i(f; \langle g \rangle) = 0$ ; i.e.,  $\langle f \rangle$  and  $\langle g \rangle$  can be isotoped off each other.

**Proof** For any integer  $N$ ,  $f^N$  commutes with  $f$ . This implies that  $f(\langle f^N(T) \rangle) = \langle f^N(f(T)) \rangle = \langle f^N(T) \rangle$ . Suppose a simple closed curve  $c$  intersects  $@ \langle f^N(T) \rangle$  non-trivially. Then  $f^{-N}(c)$  intersects  $@T$  non-trivially, and so  $f(f^{-N}(c)) \not\subset f^{-N}(c)$ , by assumption. Applying  $f^N$  to both sides, we get  $f(c) \not\subset c$ . Hence  $f$

keeps  $N(T)$  precisely invariant. By the basic property of the essential reduction system, either  $fj_T = id$  or  $fj_T$  is pseudo-Anosov.

**Case 1** Assume  $fj_T = id$ . Since  $fj_{N(T)} = (Nj_T)fj_T(Nj_T)^{-1}$ ; we have  $fj_{N(T)} = id$  for all  $N$ . Notice that  $i(@N(T); @T) = 0$ , since  $f$  keeps  $N(T)$  precisely invariant for all  $N$ . Moreover, we claim that no component  $c$  of  $@N(T)$  can be isotopic to a simple closed curve in  $T$  which is not isotopic to a component of  $@T$ . Otherwise, one can find a simple closed curve in  $T$  such that  $i(c; ) \neq 0$ . But  $f( ) =$ , which contradicts the fact that  $f$  keeps  $N(T)$  precisely invariant. Similarly one can show that no component of  $@T$  can be isotopic to a simple closed curve in  $N(T)$  which is not isotopic to  $@N(T)$ . This shows that either  $N(T) = T$  or  $N(T)$  can be isotoped o  $T$ . This in turn implies that the collection of subsurfaces  $fN(T) j N \mathbb{Z}g$  is a collection of disjoint homeomorphic subsurfaces up to isotopy, and hence it is a finite collection. This shows that  $N(T) = T$  for some  $N$ , and since is pure,  $(T) = T$ ; up to isotopy, by Lemma 5.6.

**Case 2** Let  $fj_T$  be pseudo-Anosov. Again, since  $fj_{N(T)} = (Nj_T)fj_T(Nj_T)^{-1}$ ; we have  $fj_{N(T)}$  is pseudo-Anosov for all  $N$ . Also, notice that  $i(@N(T); @T) = 0$ , since  $f$  keeps  $N(T)$  precisely invariant for all  $N$ . Moreover, we claim that no component  $c$  of  $@N(T)$  can be isotopic to a simple closed curve in  $T$  which is not isotopic to a component of  $@T$ . Otherwise, since  $c \in @N(T)$  and  $f$  is pure and pseudo-Anosov on  $N(T)$ , we have  $f(c) = c$ . On the other hand,  $c$  is in the interior of  $T$  and  $f$  is pseudo-Anosov on  $T$ , hence  $f(c) \neq c$ , which is a contradiction. Similarly one can show that no component of  $@T$  can be isotopic to a simple closed curve in  $N(T)$  which is not isotopic to  $@N(T)$ . This shows that either  $N(T) = T$  or  $N(T)$  can be isotoped o  $T$ . The rest of the argument is exactly as in Case 1. This proves (i).

To prove (ii), let  $c \in (f)$ . Let  $T$  be component of  $Sn(f)$  such that  $c$  is a component of  $@T$ . Then  $(T) = T$ , by (i). This implies that permutes the components of  $@T$ , which by Lemma 5.6 implies that  $(c) = c$ , proving (ii).

To prove (iii), let  $c \in (f)$  and  $\in ( )$  such that  $i(c; ) > 0$ . Then by definition of an essential reduction system,  $(c) \neq c$ , which contradicts (ii).  $\square$

Let  $H = \langle h_1; z_1; z_2; t_i \rangle$  be an FP-subgroup of the type described in Theorem 5.3. Notice that by Lemma 5.8(iii),  $(i) [ (j)$  is collection of non-intersecting simple closed curves. For  $i = 1; 2$ , let  $C_i = (i) \setminus (i)$ ,  $A_i = (i) \cap C_i$  and  $D_i = (i) \cap C_i$ . Note that each of  $A_i; C_i$  or  $D_i$  could be empty.

**Lemma 5.9** For  $i = 1; 2$ ,  $A_i [ D_i \subset (i)$ .

**Proof** Without loss of generality, we prove  $A_i \in \langle \gamma_i \rangle$ . Let  $c \in A_i$ . Notice that by Lemma 5.8(ii),  $\gamma_i(c) = \gamma_i(c) = c$ . If  $c \notin \langle \gamma_i \rangle$ , by definition, there is a subsurface  $U$  containing  $c$  where  $U$  is a component of  $S \setminus \langle \gamma_i \rangle$ . Since  $\gamma_i|_U$  fixes  $c$ , it is not pseudo-Anosov and hence is the identity. Similarly since  $c \notin \langle \gamma_j \rangle$ , there is a subsurface  $V$  containing  $c$  where  $V$  is a component of  $S \setminus \langle \gamma_j \rangle$  such that  $\gamma_j|_V = id$ . Therefore  $\gamma_j|_{U \setminus V} = id$ . Since  $c$  is not isotopic to any component of  $\partial U$  or  $\partial V$ , and  $i(\partial U; \partial V) = 0$ ,  $c$  is not isotopic to any component of  $\partial(U \setminus V)$ . Then one can find a simple closed curve  $\gamma$  in  $U \setminus V$  such that  $i(c; \gamma) > 0$ . But  $\gamma_j|_{U \setminus V} = id$ , so  $\gamma_j(\gamma) = \gamma$ , which contradicts the fact that  $c \notin \langle \gamma_j \rangle$ .  $\square$

**Lemma 5.10**  $i(\gamma_1; \gamma_2) = 0$ .

**Proof** Recall that  $\langle \gamma_i \rangle = A_i [ C_i$  and  $\langle \gamma_j \rangle = C_j [ D_j$ . By definition of essential reduction system and Lemma 5.8(ii),  $i(\gamma_i; \gamma_j) = 0$  and so

$$i(A_i; C_j) = i(A_i; D_j) = i(C_1; C_2) = i(C_1; D_2) = 0;$$

for  $i, j = 1, 2$ . Therefore  $i(\gamma_1; \gamma_2) = i(A_1; A_2)$ . Now by Lemma 5.9,

$$\begin{aligned} i(\gamma_1 \gamma_2; \gamma_1 \gamma_2) &= i(A_1; A_2) + i(A_1; D_2) + i(D_1; A_2) + i(D_1; D_2) \\ &= i(A_1; A_2) + i(D_1; D_2); \end{aligned}$$

By part (3) of Theorem 5.3 and Proposition 5.5, we have that

$$\begin{aligned} i(A_1; A_2) &= i(\gamma_1; \gamma_2) \\ &= i(t(\gamma_1 \gamma_2)t^{-1}; t(\gamma_1 \gamma_2)t^{-1}) \\ &= i(\gamma_1 \gamma_2; \gamma_1 \gamma_2) \\ &= i(A_1; A_2) + i(D_1; D_2); \end{aligned}$$

Thus  $i(D_1; D_2) = 0$ . Hence

$$\begin{aligned} i(\gamma_1; \gamma_2) &= i(\langle \gamma_1 \rangle; \langle \gamma_2 \rangle) \\ &= i(C_1 [ D_1; C_2 [ D_2) \\ &= i(C_1; C_2) + i(C_1; D_2) + i(D_1; C_2) + i(D_1; D_2) \\ &= 0; \end{aligned}$$

which proves the lemma.  $\square$

For a connected subsurface  $U$  of  $S$ , we define a subgroup  $\mathcal{M}_3(U)$  of the mapping class group of  $U$  as follows:

$$\mathcal{M}_3(U) = \langle f f|_U j f \rangle \mathcal{M}_3 \text{ and } f(U) = Ug;$$

Notice that all elements of  $\pi_3(U)$  are pure. Also notice that if  $\phi_i$  (respectively  $\psi_i$ ) keeps  $U$  invariant, then by Theorem 5.3 we have  $\phi_{iU} \in \pi_3(U)$  (respectively  $\psi_{iU} \in \pi_3(U)$ ). The following lemma is proved in [11] (Lemma 8.13).

**Lemma 5.11** *Let  $\Gamma$  be a subgroup of the mapping class group of a connected surface  $U$  consisting of pure elements. If  $f \in \Gamma$  is a pseudo-Anosov element, then its centralizer in  $\Gamma$  is an infinite cyclic group generated by a pseudo-Anosov element.*

**Corollary 5.12** *Let  $\Gamma$  be a subgroup of the mapping class group of a connected surface  $U$  consisting of pure elements. If  $f, h \in \Gamma$  are pseudo-Anosov elements, then either  $f$  commutes with  $h$  or their respective centralizers in  $\Gamma$  intersect trivially.*

**Proof** Let  $C(f)$  denote the centralizer of  $f$  in  $\Gamma$ . Suppose there is an element  $1 \neq h \in C(f) \setminus C(h)$ . Then  $\langle f, h \rangle \in C(f)$ , which is cyclic by Lemma 5.11, so  $f$  commutes with  $h$ .  $\square$

We are going to encounter the following particular situation in different contexts, so we declare it a lemma:

**Lemma 5.13** *Let  $U$  be a component of  $S_n(\delta_i)$  for  $i = 1$  or  $i = 2$  such that  $\pi_3(U)$  is non-trivial. Assume that  $\phi_{iU} = id$  and  $\psi_i(U) = U$  for  $i = 1, 2$ . Then the respective centralizers of  $\phi_{1U}$  and  $\phi_{2U}$  in  $\pi_3(U)$  intersect non-trivially.*

**Proof** Without loss of generality, let  $U$  be a component of  $S_n(\delta_1)$ . Assume on the contrary that the centralizers of  $\phi_{1U}$  and  $\phi_{2U}$  in the mapping class group of  $U$  have only the identity map in common. This in particular implies that  $\phi_{iU} \neq id$  for  $i = 1, 2$ . The map  $\phi_{1U}$  is pseudo-Anosov, since  $U$  is a component of  $S_n(\delta_1)$ . Consider the subsurface  $t(U)$ . By part (3) of Theorem 5.3, we have

$$\phi_{t(U)} = (t_{1U})(\phi_{1U} \phi_{2U})(t_{1U})^{-1} = (t_{1U})(\phi_{1U})(t_{1U})^{-1}. \quad (7)$$

This implies that  $\phi_{t(U)} \neq id$  keeps  $t(U)$  invariant, since it is conjugate to  $\phi_{1U}$ , for  $i = 1, 2$ . Moreover,  $\phi_{1t(U)}$  is pseudo-Anosov. This in particular implies that  $t(U)$  is a component of  $S_n(\delta_1)$ , and  $t(U)$  can be isotoped to  $U$ , since  $\phi_{1U} = id$ . Moreover, by assumption and by (7), the centralizers of  $\phi_{1t(U)}$  and  $\phi_{2t(U)}$  intersect trivially in  $\pi_3(t(U))$ . By Lemma 5.8(i),  $\phi_i$  keeps  $t(U)$  invariant for  $i = 1, 2$ , since  $\phi_i$  commutes with  $\phi_1$ . Again, since  $\phi_{1t(U)}$  commutes with  $\phi_{2t(U)}$  and by the assumption about the centralizers, we have  $\phi_{it(U)} = id$ , for  $i, j = 1, 2$ . Now we can prove the following statements for  $N \geq 1$  simultaneously by induction on  $N$ :

- (1)  $j_{t^N(U)} \notin id$  keeps  $t^N(U)$  invariant, for  $i = 1;2$ .
- (2)  $j_{t^N(U)}$  is pseudo-Anosov (hence,  $j_i$  keeps  $t^N(U)$  invariant for  $i = 1;2$ ).
- (3) The respective centralizers of  $j_{t^N(U)}$  in  $\pi_3(t^N(U))$  intersect trivially, for  $i = 1;2$ .
- (4)  $j_{t^N(U)} = id$ , for  $i = 1;2$ .

We have already established all four statements for  $N = 1$ . The passage from  $N$  to  $N + 1$  follows similarly from the relation:

$$j_{t^{N+1}(U)} = (t_{j_{t^N(U)}})(j_{t^N(U)} j_{t^N(U)})(t_{j_{t^N(U)}})^{-1} = (t_{j_{t^N(U)}})(j_{t^N(U)})(t_{j_{t^N(U)}})^{-1}$$

The second statement above shows that  $t^N(U)$  can be isotoped to  $U$ , since  $j_U = id$ . Therefore,  $t^M(U)$  can be isotoped to  $t^N(U)$  for all  $M \notin N$ . This is clearly a contradiction, since the Euler characteristic of  $S$  is finite.  $\square$

**Lemma 5.14** For  $i = 1;2$ , let  $U_i$  be a component of  $S_n(i)$  such that  $j_{U_i}$  is pseudo-Anosov. Then either  $U_1$  and  $U_2$  are disjoint up to isotopy, or  $U_1$  is isotopic to  $U_2$ .

**Proof** First we show that if  $U_1$  and  $U_2$  are not disjoint, then either  $U_1 = U_2$  or  $U_2 = U_1$ . Suppose  $U_1 \not\subseteq U_2$  and  $U_2 \not\subseteq U_1$  but  $U_1$  cannot be isotoped to  $U_2$ . Throughout the proof, let  $j;k \in \{1;2\}$  be arbitrary such that  $j \notin k$ . Since  $i(\partial U_1; \partial U_2) = 0$ , there is some component  $c_j$  of  $\partial U_j$  such that  $c_j \subset U_k$  and  $c_j$  is not isotopic to any component of  $\partial U_k$ . By Lemma 5.8(i),  $j_i$  keeps  $U_1$  and  $U_2$  invariant for  $i = 1;2$ . Since  $j_i \in \pi_3$ , we have  $j_{U_j} \in \pi_3(U_j)$ . Since  $c_j$  is in the interior of  $U_k$  and  $j_i(c_j) = c_j$  by Lemma 5.8(ii), this implies that  $j_{U_k}$  is not pseudo-Anosov, hence by Lemma 5.11,  $j_{U_k} = id$  for  $i;k = 1;2$ .

Let  $U = U_1 \cup U_2$ . At this point we apply a similar argument as in the proof of Lemma 5.13, as follows. By the relation

$$j_{t(U_i)} = (t_{j_{U_i}})(j_{U_i} j_{U_i})(t_{j_{U_i}})^{-1} = (t_{j_{U_i}})(j_{U_i})(t_{j_{U_i}})^{-1}; \tag{8}$$

we see that  $j_{t(U_i)}$  is pseudo-Anosov. This in particular implies that  $t(U) = t(U_1) \cup t(U_2)$  can be isotoped to  $U$ , since  $j_U = id$ . Note that  $t(U_i)$  is a component of  $S_n(i)$ , so  $j_j$  keeps  $t(U_i)$  invariant for  $i;j = 1;2$ , by Lemma 5.8(i). Since  $j_i$  is pure, and  $t(c_j)$  is a boundary component of  $t(U_j)$ , we have  $j_i(t(c_j)) = t(c_j)$ . By the choice of  $c_j$  we know that  $t(c_j)$  is in the interior of  $t(U_k)$ . By Lemma 5.11 and the fact that  $j_{t(U_k)} \in \pi_3(U_k)$ , we have  $j_{t(U_k)} = id$ , for  $i;k = 1;2$ . Now by induction on  $N$  we can simultaneously prove the following statements for  $N \geq 1$ :

- (1) The map  $\iota_j^{t^N(U_i)}$  is pseudo-Anosov, for  $i = 1; 2$ .
- (2) We have  $\iota_j^{t^N(U_j)} = id$ , for  $i; j = 1; 2$ .

We have already established these two statements for  $N = 1$ . The passage from  $N$  to  $N + 1$  can be achieved by considering the conjugacy relation

$$\iota_j^{t^{N+1}(U_i)} = \iota_j^{t^N(U_i)} \iota_i^{t^N(U_i)} \iota_j^{t^N(U_i)} \iota_i^{-1} = \iota_j^{t^N(U_i)} \iota_i^{t^N(U_i)} \iota_i^{-1} \iota_j^{-1} \quad (9)$$

This proves statement (1) above. Now use Lemma 5.8(i) to see that  $\iota_i$  keeps  $t^{N+1}(U_j)$  invariant. This implies that  $\iota_j^{t^{N+1}(U_j)} = id$ , and by Lemma 5.11, we have statement (2).

In particular, statement (1) shows that  $t^N(U)$  can be isotoped to  $U$  for all  $N > 1$ , which is a contradiction as in Lemma 5.13. This proves that either  $U_1 = U_2$  or  $U_2 = U_1$ , or  $U_1$  and  $U_2$  can be isotoped to each other.

Now without loss of generality, suppose that  $U_1 = U_2$ , but  $U_1$  is not isotopic to  $U_2$ . Then there exists a component  $c_1$  of  $\partial U_1$  such that  $c_1$  is not isotopic to a component of  $\partial U_2$ . By Lemma 5.8(i),  $\iota_i$  keeps  $U_1$  and  $U_2$  invariant for  $i = 1; 2$ . Also, by Lemma 5.8(ii),  $\iota_i(c_1) = c_1$ , which implies  $\iota_j|_{U_2} = id$ , by Lemma 5.11. Again, using (8) we get statement (1) for  $N = 1$ . Hence  $t(U_i)$  is a component of  $S_n(\iota_i)$ . So  $\iota_i$  keeps  $U_j$  invariant. Thus  $\iota_i(t(c_1)) = t(c_1)$ , which gives  $\iota_j(U_2) = id$ , by Lemma 5.11. This proves statement (2) for  $N = 1$ . The passage from  $N$  to  $N + 1$  follows by using equation (9) above. Then again we have that  $U_2$  can be isotoped to  $t^N(U_2)$  for all  $N > 1$ , which is a contradiction. This proves that  $U_1$  is isotopic to  $U_2$ . □

**Lemma 5.15** *Let  $U$  be a component of both  $S_n(\iota_1)$  and  $S_n(\iota_2)$  such that  $\iota_j|_U$  is pseudo-Anosov for  $i = 1; 2$ . Then  $\iota_1|_U$  commutes with  $\iota_2|_U$ .*

**Proof** If  $\iota_1|_U$  and  $\iota_2|_U$  do not commute, then their centralizers in  $\pi_3(U)$  have trivial intersection by Corollary 5.12. This implies that  $\iota_j|_U = id$ , which contradicts Lemma 5.13. □

We are finally ready to prove Theorem B.

**Proof of Theorem B** Let  $U$  be a component of  $S_n(\iota_1)$  such that  $\iota_1|_U$  is pseudo-Anosov. We first prove that  $\iota_2|_U$  is either pseudo-Anosov or the identity. Suppose  $\iota_2|_U$  is neither pseudo-Anosov nor the identity (in particular,  $U$  is not a component of  $S_n(\iota_2)$ ). Let  $V_1; V_2; \dots; V_s$  be components of  $S_n(\iota_2)$ , which cover  $U$  up to isotopy. We can assume that the cover is minimal

in the sense that none of the  $V_k$  can be isotoped to  $U$ . By Lemma 5.14,  ${}_2j_{V_k} = id$  for all  $1 \leq k \leq s$ . (This does not mean that  ${}_2j_U = id$ , since  ${}_2$  may involve Dehn twists about boundary components of  $V_k$ .) By Lemma 5.10,  $i(@U; @V_k) = 0$  for all  $1 \leq k \leq s$ , which shows that  ${}_2$  keeps  $U$  invariant. Moreover,  ${}_2j_U$  is a non-trivial composition of Dehn twists about disjoint simple closed curves. Using Lemma 5.8(i),  ${}_1$  keeps  $U$  invariant. Since  ${}_1j_U; {}_2j_U \in \mathcal{C}_3(U)$  and  ${}_1j_U$  commutes with  ${}_2j_U$ , using Lemma 5.11 we see that  ${}_1j_U$  cannot be pseudo-Anosov. Moreover,  ${}_1j_U$  commutes with  ${}_1j_U$  so  ${}_1j_U = id$ . Now by Lemma 5.13, we get that the centralizers of  ${}_1j_U$  and  ${}_2j_U$  must intersect non-trivially. Lemma 5.11 then implies that  ${}_2j_U$  is either pseudo-Anosov or the identity, which is a contradiction.

We have proved that for a component  $U$  of  $S \setminus n$  (1) where  ${}_1j_U$  is pseudo-Anosov,  ${}_2j_U$  is either pseudo-Anosov or the identity. In the case that  ${}_2j_U$  is pseudo-Anosov,  ${}_1j_U$  and  ${}_2j_U$  commute by Lemma 5.15. Similarly, for a component  $V$  of  $S \setminus n$  (2) where  ${}_2j_V$  is pseudo-Anosov,  ${}_1j_V$  is either a commuting pseudo-Anosov or the identity.

Let  $S_1$  be the subsurface of  $S$  which is the union of subsurfaces  $T$  such that either  ${}_1j_T$  or  ${}_2j_T$  is pseudo-Anosov. We have proved that  ${}_1$  and  ${}_2$  both keep  $S_1$  invariant, and  ${}_1j_{S_1}$  commutes with  ${}_2j_{S_1}$ .

On  $S_2 = S \setminus S_1$  both  ${}_1$  and  ${}_2$  are compositions of Dehn twists about disjoint curves, by Lemma 5.10. Hence  ${}_1j_{S_2}$  and  ${}_2j_{S_2}$  commute. We conclude that  ${}_1$  and  ${}_2$  commute, contradicting part (1) of Theorem 5.3. This shows that FP-groups do not embed in  $M_{g,b,n}$ , as desired.  $\square$

## References

- [1] H. Bass. *Covering theory for graphs of groups*, J. Pure & App. Alg., **89** 1993, 3-47.
- [2] S. Bigelow. *Braid groups are linear*, J. AMS, to appear 2001.
- [3] S. Bigelow & R. Budney. *The mapping class group of a genus two surface is linear*, preprint.
- [4] J. Birman. *Braids, links and mapping class groups*, Annals of Mathematics Studies, **82**, Princeton University Press, 1974.
- [5] J. Birman, A. Lubotzky & J. McCarthy. *Abelian and solvable subgroups of mapping class group*, Duke Math J. **50** 1983, 1107-1120.
- [6] K. Brown. *Cohomology of groups*, Springer-Verlag, 1982.
- [7] J. L. Dyer, E. Formanek & E. K. Grossman. *On the linearity of automorphism groups of free groups*, Arch. Math (Basel) **38** 1982, 404-409.

- [8] D. B. A. Epstein. *Curves on 2-manifolds and isotopies*, Acta Math. **115** 1966, 83-107.
- [9] E. Formanek & C. Procesi. *The automorphism group of a free group is not linear*, J. Alg., **149** 1992, 494-499.
- [10] J. Hass & P. Scott, *Shortening curves on surfaces*, Topology, **33** (1) 1994, 25 - 43.
- [11] N. Ivanov. *Subgroups of Teichmüller modular groups*, Translations of Mathematical Monographs, Am. Math. Soc., 1992.
- [12] G. Levitt & K. Vogtmann. *A Whitehead algorithm for surface groups*, Topology, **39** 2000, 1239 - 1251.
- [13] M. Korkmaz. *On the linearity of certain mapping class groups*, preprint.
- [14] D. Krammer. *The braid group  $B_4$  is linear*, preprint.
- [15] W. Magnus. A. Karrass & D. Solitar. *Combinatorial group theory*, Dover Publications, Inc., 1976.
- [16] J. McCarthy. *A Tits alternative for subgroups of surface mapping class groups*, Trans. Amer. Math. Soc. **291** 1985, 583-612.
- [17] J. Nielsen. *Die Gruppe der dreidimensionalen Gittertransformationen*, Danske Vid. Selsk. Mat-Fys. 1924, 1 - 29.
- [18] D. Rolfsen. *Knots and links*, Publish or Perish, Inc., 1990.
- [19] B. Wehrfritz. *Infinite linear groups*, Ergebnisse der Mathematik und Ihrer Grenzgebiete, **76**, Springer-Verlag, 1973.

*Columbia University*

*Department of Mathematics*

*New York, NY 10027, USA*

and

*B.C.C. of the City University of New York*

*Department of Mathematics and Computer Science*

*Bronx, NY 10453, USA*

Email: [tbrendle@math.columbia.edu](mailto:tbrendle@math.columbia.edu), [hessam@math.columbia.edu](mailto:hessam@math.columbia.edu)