Normal generators for mapping class groups are abundant

Justin Lanier and Dan Margalit

Abstract. We provide a simple criterion for an element of the mapping class group of a closed surface to be a normal generator for the mapping class group. We apply this to show that every nontrivial periodic mapping class that is not a hyperelliptic involution is a normal generator for the mapping class group when the genus is at least 3. We also give many examples of pseudo-Anosov normal generators, answering a question of D. D. Long. In fact, we show that every pseudo-Anosov mapping class with stretch factor less than $\sqrt{2}$ is a normal generator. Even more, we give pseudo-Anosov normal generators with arbitrarily large stretch factors and arbitrarily large translation lengths on the curve graph, disproving a conjecture of Ivanov.

1. Introduction

Let S_g denote a connected, closed, orientable surface of genus g. The mapping class group $\mathrm{Mod}(S_g)$ is the group of homotopy classes of orientation-preserving homeomorphisms of S_g. The goal of this paper is to give new examples of elements of $\mathrm{Mod}(S_g)$ that have normal closure equal to the whole group; in this case we say that the element *normally generates* $\mathrm{Mod}(S_g)$.

In the 1960s, Lickorish [32] and Mumford [42] proved that $\mathrm{Mod}(S_g)$ is normally generated by a Dehn twist about a nonseparating curve in S_g. On the other hand for $k > 1$ the kth power of a Dehn twist is not a normal generator since it acts trivially on the mod k homology of S_g.

The Nielsen–Thurston classification theorem for mapping class groups categorizes elements of $\mathrm{Mod}(S_g)$ as either periodic, reducible, or pseudo-Anosov; see [16, Chapter 13]. Dehn twists and their powers are examples of reducible elements. Our primary focus in this paper is to find normal generators for $\mathrm{Mod}(S_g)$ among the periodic and pseudo-Anosov mapping classes. On the other hand, the criteria we develop to prove our results about periodic and pseudo-Anosov elements apply equally well to reducible elements.

Periodic elements. By the work of Harvey, Korkmaz, McCarthy, Papadopoulos, and Yoshihara [19, 25, 40, 46], there are several specific examples of periodic mapping classes that normally generate $\mathrm{Mod}(S_g)$. The first author recently showed [27] that for $k \geq 5$ and $g \geq (k-1)^2 + 1$, there is a mapping class of order k that normally generates $\mathrm{Mod}(S_g)$. We emphasize that the periodic elements produced in the proof of the latter result all have a special form, namely, they are realized by rotations of a surface embedded in $\mathbb{R}^3$; on the other hand, periodic mapping classes are not typically realized by such a rotation.

Our first theorem completely answers the question of which periodic elements normally generate. In the statement, the hyperelliptic involution is the element (or, conjugacy class) of $\mathrm{Mod}(S_g)$ depicted in Figure 1.
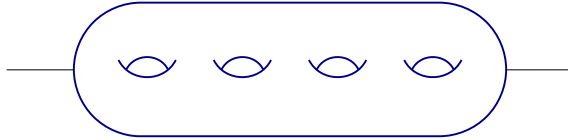

Figure 1. Rotation by π about the indicated axis is a hyperelliptic involution.

Theorem 1.1. *For $g \geq 3$, every nontrivial periodic mapping class that is not a hyperelliptic involution normally generates* $\mathrm{Mod}(S_g)$.

Additionally, we show in Proposition 3.3 that for $g \geq 3$ the normal closure of the hyperelliptic involution is the preimage of $\{\pm I\}$ under the standard symplectic representation of $\mathrm{Mod}(S_g)$. The Torelli group $\mathcal{I}(S_g)$ is the kernel of this representation, so the normal closure of the hyperelliptic involution contains $\mathcal{I}(S_g)$ as a subgroup of index 2. We also give in Section 3 a complete classification of the normal closures of periodic elements of $\mathrm{Mod}(S_g)$ when $g \leq 2$.

In Section 3 we give an extension of Theorem 1.1 to surfaces $S_{g,n}$ of genus $g \geq 3$ with n marked points (equivalently, punctures). Specifically, we show in Theorem 3.6 that the normal closure of every nontrivial periodic mapping class that is not a hyperelliptic involution contains the pure mapping class group $\mathrm{PMod}(S_{g,n})$. Here, a hyperelliptic involution is a periodic element that maps to a hyperelliptic involution in $\mathrm{Mod}(S_g)$ under the forgetful map $\mathrm{Mod}(S_{g,n}) \to \mathrm{Mod}(S_g)$. Note that $\mathrm{PMod}(S_{g,1})$ is equal to the full mapping class group $\mathrm{Mod}(S_{g,1})$.

Several well-known facts about mapping class groups follow from Theorem 1.1. For instance, the Torelli group $\mathcal{I}(S_g)$ is torsion free. Also, the level m congruence subgroup $\mathrm{Mod}(S_g)[m]$, the kernel of the action of $\mathrm{Mod}(S_g)$ on $H_1(S_g; \mathbb{Z}/m\mathbb{Z})$, is torsion free for $m \geq 3$. Additionally, it follows that there are no finite nontrivial normal subgroups of $\mathrm{Mod}(S_g)$ except for the cyclic subgroups of $\mathrm{Mod}(S_1)$ and $\mathrm{Mod}(S_2)$ generated by the hyperelliptic involution. One new consequence is that any normal subgroup of $\mathrm{Mod}(S_g)$ not containing $\mathcal{I}(S_g)$ is torsion free.

Another consequence of Theorem 1.1 involves the subgroup of $\text{Mod}(S_g)$ generated by nth powers of all elements. Let L be the least common multiple of the orders of the periodic elements of $\text{Mod}(S_g)$. For n not divisible by $L/2$ we show that the nth power subgroup of $\text{Mod}(S_g)$ is equal to the whole group $\text{Mod}(S_g)$; see Corollary 3.5. This improves on a result of Funar [18], who proved the analogous result with $L/2$ replaced by $4g + 2$. Funar's theorem answers in the negative a question of Ivanov, who asked in his problem list [22, Problem 13] if the nth power subgroup of $\text{Mod}(S_g)$ has infinite index for n sufficiently large.

It follows from the fact that $\text{Mod}(S_g)$ has a periodic normal generator that any homomorphism from $\text{Mod}(S_g)$ to a torsion-free group is trivial. Theorem 1.1 gives a strengthening: any homomorphism from $\text{Mod}(S_g)$ to a group without the "right" torsion must be trivial. Theorem 1.1, its extension Theorem 3.6 for punctured surfaces, and the surrounding ideas have already been leveraged in this way to prove several results about homomorphisms of mapping class groups:

(1) Mann and Wolf showed that for $g \geq 3$ any homomorphism $\text{Mod}(S_{g,1}) \to \text{Homeo}^+(S^1)$ is either trivial or equivalent to the standard Gromov boundary action [37].

(2) Chen and the first author proved that for $g \geq 3$ and $h < 2g - 1$ with $h \neq g$, any homomorphism $\text{Mod}(S_g) \to \text{Mod}(S_h)$ is trivial [8].

(3) Chen, Kordek, and the second author showed that any homomorphism from the braid group B_n to the braid group B_{2n} is either cyclic or is equivalent to one of the standard inclusions [7].

The first result answers a special case of a question of Farb [14, Questions 6.2]. The second result extends a special case of a result of Aramayona–Souto; they proved an analogous statement for surfaces of genus $g \geq 6$, possibly with punctures or boundary [2, Theorem 1.1].

The analogue of Theorem 1.1 for nonorientable surfaces was recently proved by Leśniak [31].

Pseudo-Anosov elements with small stretch factor. Having addressed the periodic elements, we turn to the case of pseudo-Anosov mapping classes. In a 1986 paper, Darren Long asked [33]:

   Can the normal closure of a pseudo-Anosov map ever be all of $\text{Mod}(S_g)$?

Long answered in the affirmative the analogous question for $g = 1$, where "pseudo-Anosov" is replaced with "Anosov." In Proposition 4.1 below we give a flexible construction that gives many pseudo-Anosov normal generators for each $g \geq 1$, thus answering Long's question.

Penner constructed a family of pseudo-Anosov mapping classes, one for each g, with the property that the stretch factors tend to 1 as g tends to infinity [43]. We show in Proposition 4.3 that each of these small stretch factor pseudo-Anosov mapping classes is a normal generator.

Our second main theorem shows that in fact every pseudo-Anosov mapping class with sufficiently small stretch factor is a normal generator.

Theorem 1.2. *If a pseudo-Anosov element of* $\mathrm{Mod}(S_g)$ *has stretch factor less than* $\sqrt{2}$ *then it normally generates* $\mathrm{Mod}(S_g)$.

For each $g \geq 3$ there are pseudo-Anosov elements of $\mathrm{Mod}(S_g)$ that satisfy the hypothesis of Theorem 1.2. Indeed, for $g \geq 4$ we may use the fact that there are pseudo-Anosov mapping classes with stretch factor less than $\phi^{2/(g-1)}$, where ϕ is the golden ratio (see [1, Proposition A.1]). For $g = 3$ we may appeal to the example given by Hironaka; see Table 1 in her paper [20, Theorem 1.5]. On the other hand, for $g \leq 2$ it is known that there are no pseudo-Anosov mapping classes that satisfy the hypothesis of Theorem 1.2. Indeed in these cases the smallest stretch factors are known and they are greater than $\sqrt{2}$; for $g = 1$ this is classical and for $g = 2$ this is due to Cho and Ham [9].

We can make precise one sense in which the mapping classes with stretch factor less than $\sqrt{2}$ are abundant. A theorem of Leininger and the second author of this paper [30, Theorem 1.3] has the following consequence: for any d there is a polynomial $Q(g)$ with degree d and with positive leading term so that the number of pseudo-Anosov elements of $\mathrm{Mod}(S_g)$ with stretch factor less than $\sqrt{2}$ is bounded below by $Q(g)$ for $g \gg 0$.

Another way to state Theorem 1.2 is:

> *If a pseudo-Anosov mapping class lies in any proper normal subgroup of* $\mathrm{Mod}(S_g)$*, its stretch factor is greater than* $\sqrt{2}$.

As such, Theorem 1.2 generalizes work of Agol, Farb, Leininger, and the second author of this paper. Farb, Leininger, and the second author proved that if a pseudo-Anosov mapping class is contained in $\mathrm{Mod}(S_g)[m]$ with $m \geq 3$ then the stretch factor is greater than 1.218 [15, Theorem 1.7]. Agol, Leininger, and the second author proved that if a pseudo-Anosov mapping class is contained in $\mathrm{Mod}(S_g)[2]$ then the stretch factor is greater than 1.00031 [1, Theorem 1.1]. Our Theorem 1.2 improves upon these results in two ways: from congruence subgroups to arbitrary normal subgroups and from 1.218 and 1.00031 to $\sqrt{2}$.

With regard to the first improvement, there do indeed exist examples of normal subgroups that are not contained in congruence subgroups (and hence not covered by the theorems of Agol, Farb, Leininger, and the second author). Such examples were recently constructed by Clay, Mangahas, and the second author [10], solving a problem posed in an earlier version of this paper [28, Problem 1.5].

With regard to the second improvement, we note that when using the bound $\phi^{2/(g-1)}$, we only obtain examples of pseudo-Anosov mapping classes with stretch factor less than 1.00031 when g is at least 3,106.

The bound $\sqrt{2}$ in the statement of Theorem 1.2 comes from the following considerations. We show in Section 6 that if the stretch factor of f is bounded above by $\sqrt{2}$ then there is a curve c so that the pairwise geometric intersection numbers

of c, $f(c)$, and $f^2(c)$ are bounded above by 2. In other words, the constraint on the stretch factor of f is converted into a constraint on the action of f on curves in S_g. The proof of Theorem 1.2 proceeds by classifying all possible triples of curves $(c, f(c), f^2(c))$ with pairwise geometric intersection number at most 2 and treating each possibility on a case-by-case basis. In order to relax the bound $\sqrt{2}$ to a slightly larger number, our approach would require a consideration of triples of curves $(c, f(c), f^2(c))$ with pairwise geometric intersection number at most n, where $n > 2$. Even for $n = 3$, a classification of such triples would be daunting. And even if this were accomplished, it is possible that there is a pseudo-Anosov f that is not a normal generator and a curve c so that c, $f(c)$, and $f^2(c)$ have pairwise intersection at most 3.

That said, it is natural to ask how sharp the bound $\sqrt{2}$ is in Theorem 1.2. In other words, what is the infimum of the stretch factors of all pseudo-Anosov mapping classes lying in any proper normal subgroup of any $\mathrm{Mod}(S_g)$? Farb, Leininger, and the second author [15] showed that for each $g \geq 2$ there are pseudo-Anosov elements of $\mathcal{I}(S_g)$ with stretch factor at most 62. Thus, the infimum lies between $\sqrt{2}$ and 62. On the other hand, Farb, Leininger, and the second author proved that for the kth term of the Johnson filtration of $\mathrm{Mod}(S_g)$ the smallest stretch factor tends to infinity as k does (independently of g), so this gives a sequence of normal subgroups for which the bound $\sqrt{2}$ of Theorem 1.2 becomes decreasingly sharp. It is an interesting problem, already raised by Farb, Leininger, and the second author [15], to understand the smallest stretch factors in various specific normal subgroups, such as the level 2 congruence subgroup $\mathrm{Mod}(S_g)[2]$.

Pseudo-Anosov elements with large stretch factor. Having given many examples of pseudo-Anosov normal generators with small stretch factor, we turn to the question of what other kinds of pseudo-Anosov normal generators may exist. In this vein, Ivanov [22, Problem 13] made the following conjecture in his 2006 problems paper:

Conjecture. *If f is a pseudo-Anosov element of a mapping class group $\mathrm{Mod}(S)$ with sufficiently large dilatation coefficient, then the subgroup of $\mathrm{Mod}(S)$ normally generated by f is a free group having as generators the conjugates of f. More cautiously, one may conjecture that the above holds for a sufficiently high power $g = f^N$ of a given pseudo-Anosov element f.*

(The term "dilatation coefficient" is interchangeable with the term "stretch factor.") In Proposition 4.2 below we give for each $g \geq 1$ a flexible construction of pseudo-Anosov normal generators for $\mathrm{Mod}(S_g)$ with arbitrarily large stretch factor. Since $\mathrm{Mod}(S_g)$ is not a free group, this in particular disproves the first part of Ivanov's conjecture.

Much more than this, we have the following theorem. In the statement, note that $\mathrm{Mod}(S_g)[m]$ is $\mathrm{Mod}(S_g)$ when $m = 1$ and it is $\mathcal{I}(S_g)$ when $m = 0$.

Theorem 1.3. *Let $g \geq 3$. For each $m \geq 0$ there are pseudo-Anosov mapping classes with arbitrarily large stretch factors whose normal closures in $\mathrm{Mod}(S_g)$ are equal to $\mathrm{Mod}(S_g)[m]$.*

It is natural to ask if other normal subgroups, such as the Johnson kernel and the other terms of the Johnson filtration, can be obtained as the normal closure in $\mathrm{Mod}(S_g)$ of a single pseudo-Anosov mapping class. In fact, it is an open question whether any of these groups can be obtained as the normal closure in $\mathrm{Mod}(S_g)$ of a single element.

The second part of Ivanov's conjecture is (perhaps intentionally) ambiguous: the word "a" can be interpreted as either "some" or "any." Dahmani, Guirardel, and Osin proved that every pseudo-Anosov mapping class has a large power (depending only on g) whose normal closure is an all pseudo-Anosov infinitely generated free group [11]. This confirms the second part of Ivanov's conjecture with the "some" interpretation. Their theorem also answers another question in Ivanov's problem list [22, Problem 3], which asks if there are any normal, all pseudo-Anosov subgroups of $\mathrm{Mod}(S_g)$.

Our next theorem disproves the second part of Ivanov's conjecture with the "any" interpretation. In the statement, the curve graph $\mathcal{C}(S_g)$ is the graph whose vertices are homotopy classes of simple closed curves in S_g and whose edges are pairs of vertices with disjoint representatives in S_g. Masur and Minsky proved that the asymptotic translation length for a pseudo-Anosov element of $\mathrm{Mod}(S_g)$ is a positive real number [39, Proposition 4.6].

Theorem 1.4. *For each $g \geq 3$ there are pseudo-Anosov mapping classes with the property that all of their odd powers normally generate $\mathrm{Mod}(S_g)$. Consequently, there are pseudo-Anosov mapping classes with arbitrarily large asymptotic translation lengths on $\mathcal{C}(S_g)$ that normally generate $\mathrm{Mod}(S_g)$.*

The pseudo-Anosov mapping classes used to prove Theorem 1.4 are not generic, as their invariant foliations have nontrivial symmetry groups (cf. [38]). Clay, Mangahas, and the second author proved that if a pseudo-Anosov mapping class has invariant foliations without symmetries then the normal closure of any sufficiently large power is a free group of infinite rank [10]. Their result confirms a suspicion raised in an earlier version of this paper [28]. In a similar direction, Maher–Tiozzo [36] further proved that the typical mapping class (in the sense of random walks) has normal closure isomorphic to an infinitely generated free group (Maher had previously proved that the typical mapping class in this sense in pseudo-Anosov [35]). We are also led to ask: if the asymptotic translation length of a pseudo-Anosov mapping class is large, can its normal closure be anything other than the mapping class group, a free group, or perhaps the extended Torelli group?

Moduli spaces. The group $\mathrm{Mod}(S_g)$ can be identified with the orbifold fundamental group of $\mathcal{M}_g$, the moduli space of Riemann surfaces of genus g. This means that

Theorems 1.1 and 1.2 can both be recast in terms of the geometry of normal covers of $\mathcal{M}_g$.

Periodic elements of $\mathrm{Mod}(S_g)$ correspond to orbifold points in $\mathcal{M}_g$ and so Theorem 1.1 can be interpreted as saying that the only orbifold points in a proper normal cover of $\mathcal{M}_g$ are those arising from the hyperelliptic involution. In particular, these orbifold points all have order 2 and lie along the hyperelliptic locus.

Torelli space is the normal cover of $\mathcal{M}_g$ corresponding to the Torelli group. This space can be described as the space of Riemann surfaces with homology framings. A further consequence of Theorem 1.1 is that every normal cover of $\mathcal{M}_g$ not covered by Torelli space is a manifold. In fact, a normal cover of $\mathcal{M}_g$ is a manifold if and only if it is not covered by the quotient of Torelli space given by the action of $-I \in \mathrm{Sp}_{2g}(\mathbb{Z})$.

If we endow $\mathcal{M}_g$ with the Teichmüller metric, then pseudo-Anosov elements of $\mathrm{Mod}(S_g)$ correspond exactly to geodesic loops in $\mathcal{M}_g$. The length of a geodesic loop is the logarithm of the corresponding stretch factor. Theorem 1.2 can thus be interpreted as saying that geodesic loops in $\mathcal{M}_g$ whose lengths are less than $\log \sqrt{2}$ do not lift to loops in any proper normal cover of $\mathcal{M}_g$. The existence of pseudo-Anosov normal generators for $\mathrm{Mod}(S_g)$ with arbitrarily large stretch factor, shown in Proposition 4.2, implies the existence of arbitrarily long geodesic loops in $\mathcal{M}_g$ that do not lift to loops in any proper normal cover.

The well-suited curve criterion. All of the results about normal generators in this paper are derived from a simple, general principle for determining when a mapping class f normally generates the mapping class group. We call this principle the well-suited curve criterion. The principle is that if we can find a curve c so that the configuration $c \cup f(c)$ is "simple" enough, then f normally generates the mapping class group. We give many concrete manifestations of this principle in the paper, namely, as Lemmas 2.2, 2.3, 2.4, 7.1, 7.2, 9.1, 9.3, and 9.4 and Proposition 7.3.

Our first example of the well-suited curve criterion, Lemma 2.2 below, takes the following form for $g \geq 3$:

> *If f lies in $\mathrm{Mod}(S_g)$ with $g \geq 3$ and if c is a nonseparating curve in S_g with $i(c, f(c)) = 1$ then f is a normal generator for $\mathrm{Mod}(S_g)$.*

The well-suited curve criterion is a very general and widely applicable principle, and there are many variations besides the ones introduced in this paper. We expect that this principle can be leveraged to address other problems about mapping class groups and related groups. For example, we use a variant to prove our result about normal generators for congruence subgroups, Theorem 1.3 below. We also use the principle to give normal generators for certain linear groups.

While having a simple configuration for c and $f(c)$ is a powerful sufficient criterion for a mapping class to be a normal generator, it is not a necessary condition. This point is underscored by Theorem 1.4, which gives examples of normal generators with large translation length on the curve graph $\mathcal{C}(S_g)$. For these examples, the

distance between every curve and its image is large, which means that each curve forms a complicated configuration with its image. Still there is a way to formulate the well-suited curve criterion as a necessary, as well as sufficient, condition for a mapping class to be a normal generator. We state in Proposition 7.3 below, the most general version of our well-suited curve criterion, which says that a mapping class is a normal generator if and only if a certain associated curve graph is connected.

Subsequent to our work, a version of the well-suited curve criterion for the braid group was given by Chen, Kordek, and the second author of this paper. They used the criterion to show that the normal closure of any nontrivial periodic element contains the commutator subgroup of the braid group [7, Lemma 4.2]. Also, Baik–Kin–Shin–Wu [3] applied the well-suited curve criterion to show that normal generators are abundant among the monodromies associated to a hyperbolic fibered 3-manifold with first Betti number equal to 2.

Overview of the paper.   The remainder of the paper is divided into two parts. In the first part, Sections 2–5, we give special cases of the well-suited curve criterion and then use them to prove all of our main results aside from Theorem 1.2. In the second part, Sections 6–10, we prove Theorem 1.2, which is by far our most technical result.

We begin the first part of the paper by proving in Section 2 several special cases of the well-suited curve criterion. In Section 3, we use the special cases of the criterion to prove our Theorem 1.1 about periodic elements and to give our extension of Funar's theorem. In Section 4 we again use the special cases of the criterion to answer Long's question in the affirmative, to disprove Ivanov's conjecture, and to prove Theorems 1.3 and 1.4. In Section 5 we apply our results about mapping class groups to give normal generators for certain linear groups.

The second part of the paper begins with Section 6, in which we relate small stretch factor to geometric intersection numbers for curves and lay out the plan for the proof of Theorem 1.2. In Section 7 we state and prove the general well-suited curve criterion, Proposition 7.3, and use it to prove one case of Theorem 1.2. Then in Section 8 we address the main technical obstacle of the paper, by giving a detailed analysis of all possible configurations of certain triples of mod 2 homologous curves that intersect in at most two points pairwise on any closed surface (there are 36 configurations up to stabilization). In Section 9 we give a variant of the well-suited curve criterion that applies to many of the configurations from Section 8. Finally, in Section 10 we use the results of Sections 6–9 in order to prove Theorem 1.2.

2. The well-suited curve criterion: special cases

As discussed in the introduction, the well-suited curve criterion is the principle that if f is a mapping class and c is a curve in S_g so that the configuration $c \cup f(c)$ is simple enough, then the normal closure of f is equal to the mapping class group. We will give several examples of this phenomenon in this section, in Lemmas 2.2–2.4. Besides serving as a warm-up for the full version of the well-suited curve criterion, these special cases also suffice to prove Theorem 1.1, to answer Long's question, and to resolve Ivanov's conjecture.

Curves and intersection number. In what follows we refer to a homotopy class of essential simple closed curves in S_g as a "curve" and we write $i(c, d)$ for the geometric intersection number between curves c and d.

 We will write $[c]$ for the element of $H_1(S_g; \mathbb{Z})/\{\pm 1\}$ represented by a curve c (the ambiguity comes from the two choices of orientation). We will write $[c] \mod 2$ for the corresponding element of $H_1(S_g; \mathbb{Z}/2)$. A useful fact is that whenever $i(c, d) = 0$ we have $[c] = [d]$ if and only if $[c] = [d] \mod 2$.

 Finally, we write $|\hat{\imath}|(c, d)$ for the absolute value of the algebraic intersection number between two elements of $H_1(S_g; \mathbb{Z})$ corresponding to c and d. We will refer this number as simply the algebraic intersection number, as we will have no need to discuss the signed algebraic intersection number.

Normal generators for the commutator subgroup. The following lemma, along with Lemmas 2.2–2.3, already appears in the paper by Harvey–Korkmaz [19, Lemma 3]. The ideas also appeared in the earlier works of McCarthy–Papadopoulos [40] and Luo [34]. All of our well-suited curve criteria will be derived from this lemma.

 The conclusions of Lemmas 2.2 and 2.3 only give that the normal closure of a given element f contains the commutator subgroup of $\mathrm{Mod}(S_g)$. For $g \geq 3$ it is well known that $\mathrm{Mod}(S_g)$ is perfect [16, Theorem 5.2] and so in these cases the lemmas imply that f is a normal generator.

Lemma 2.1. *Suppose c and d are nonseparating curves in S_g with $i(c, d) = 1$. Then the normal closure of $T_c T_d^{-1}$ is equal to the commutator subgroup of $\mathrm{Mod}(S_g)$.*

Proof. We will show the two inclusions in turn. For $g \geq 0$, it is known that $\mathrm{Mod}(S_g)$ is generated by Dehn twists about nonseparating curves. It is also known that the abelianization of $\mathrm{Mod}(S_g)$ is cyclic. Since the Dehn twists about any two nonseparating curves are conjugate it follows that $T_c T_d^{-1}$, and hence its normal closure, lies in the commutator subgroup of $\mathrm{Mod}(S_g)$.

 It remains to show that the commutator subgroup is contained in the normal closure of $T_c T_d^{-1}$. Lickorish proved that there is a generating set for $\mathrm{Mod}(S_g)$ where the generators are Dehn twists about nonseparating curves $\{c_1, \ldots, c_{3g-3}\}$

in S_g and where each $i(c_i, c_j)$ is at most 1 [16, Theorem 4.13]. The commutator subgroup of $\mathrm{Mod}(S_g)$ is thus normally generated by the various $[T_{c_i}, T_{c_j}]$. When $i(c_i, c_j) = 0$, the corresponding commutator $[T_{c_i}, T_{c_j}]$ is trivial. The nontrivial commutators $[T_{c_i}, T_{c_j}]$ are all conjugate in $\mathrm{Mod}(S_g)$ to $[T_c, T_d]$, where c and d are the curves in the statement of the lemma. Therefore it suffices to show that the single commutator $[T_c, T_d]$ is contained in the normal closure of $T_c T_d^{-1}$.

It is a general fact that if a and b are elements of a group G and N is a normal subgroup of G containing ab^{-1} then $[a, b]$ is contained in N. Indeed, if we consider the quotient homomorphism $G \to G/N$ then a and b map to the same element, and so $[a, b]$ maps to the identity. Applying this general fact to our situation, we have that $[T_c, T_d]$ is contained in the normal closure of $T_c T_d^{-1}$, as desired.   □

Two well-suited curve criteria for nonseparating curves. The next two lemmas are special cases of the well-suited curve criterion.

Lemma 2.2. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Suppose that there is a nonseparating curve c in S_g so that $i(c, f(c)) = 1$. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

Proof. Consider the commutator $[T_c, f]$. Since this commutator is equal to the product of $T_c f T_c^{-1}$ and f^{-1}, it lies in the normal closure of f. Since $f T_c f^{-1}$ is equal to $T_{f(c)}$ the commutator $[T_c, f]$ is also equal to $T_c T_{f(c)}^{-1}$. Since $i(c, f(c)) = 1$, the lemma now follows from Lemma 2.1.   □

Lemma 2.3. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Suppose that there is a nonseparating curve c in S_g so that $i(c, f(c)) = 0$ and $[c] \neq [f(c)]$. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*
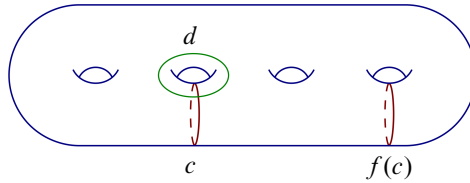

Figure 2. The curves c, $f(c)$, and d in the proof of Lemma 2.3.

Proof. It follows from the hypotheses on c and $f(c)$ that there is an additional curve d with $i(c, d) = 1$, $i(f(c), d) = 0$, and $[d] \neq [f(c)]$; see Figure 2. The commutator $[T_c, f] = T_c T_{f(c)}^{-1}$ again lies in the normal closure of f. Since $\mathrm{Mod}(S_g)$ acts transitively on pairs of disjoint, nonhomologous curves in S_g, there is an $h \in \mathrm{Mod}(S_g)$ taking the pair $(c, f(c))$ to the pair $(f(c), d)$. The conjugate $h[T_c, f]h^{-1}$,

which also lies in the normal closure of f, is equal to $T_{f(c)}T_d^{-1}$. Thus, the product

$$\left(T_c T_{f(c)}^{-1}\right)\left(T_{f(c)}T_d^{-1}\right) = T_c T_d^{-1}$$

lies in the normal closure of f. An application Lemma 2.2 completes the proof. $\square$

A well-suited curve criterion for separating curves. We will now use Lemma 2.3 to obtain an instance of the well-suited curve criterion for separating curves, as follows.

Lemma 2.4. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Suppose that there is a separating curve d in S_g with $i(d, f(d)) \leq 2$. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

Proof. Let d be a separating curve in S_g with $i(d, f(d)) \leq 2$. Since the intersection number between two separating curves is even, we have that $i(d, f(d))$ is either 0 or 2. In each case there is only one possible configuration up to homeomorphism and the genera of the complementary regions; see Figure 3.
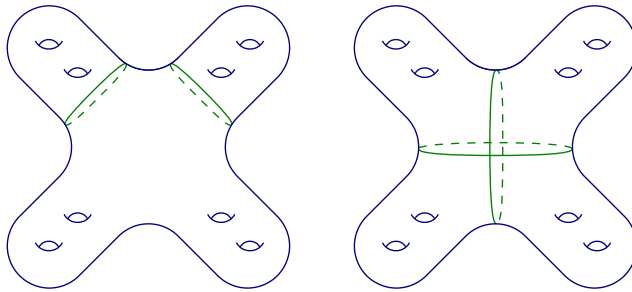

Figure 3. The configurations where c is separating and $i(c, f(c)) \leq 2$.

In each case we may find nonseparating curves a and b so that a and b lie on different sides of d and on the same side of $f(d)$. Since a and b lie on different sides of d, it follows that $f(a)$ and $f(b)$ lie on different sides of $f(d)$. Therefore, it must be that either a and $f(a)$ lie on different sides of $f(d)$ or b and $f(b)$ do (or both). Without loss of generality, suppose a and $f(a)$ lie on different sides of $f(d)$. Then clearly $i(a, f(a)) = 0$ and $[a] \neq [f(a)]$. An application of Lemma 2.3 completes the proof. $\square$

3. Application: periodic elements

In this section we apply the special cases of the well-suited curve criterion from Section 2 to determine the normal closure of each periodic element of each $\mathrm{Mod}(S_g)$. The main technical result is the following.

Proposition 3.1. *Let $g \geq 0$ and let f be a nontrivial periodic element of $\mathrm{Mod}(S_g)$ that is not a hyperelliptic involution. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

Since $\mathrm{Mod}(S_g)$ is perfect when $g \geq 3$, Proposition 3.1 immediately implies Theorem 1.1. Later in the section we will use Proposition 3.1 to determine the normal closures of all periodic elements of $\mathrm{Mod}(S_1)$ and $\mathrm{Mod}(S_2)$. At the end of the section, we prove Corollary 3.5, which gives a condition on n so that the nth power subgroup of $\mathrm{Mod}(S_g)$ is the whole group.

Our proof of Proposition 3.1 requires a lemma about roots of the hyperelliptic involution, Lemma 3.2 below. Before giving this lemma, we begin with some preliminaries.

Standard representatives. It is a classical theorem of Fenchel and Nielsen [16, Theorem 7.1] that a periodic mapping class $f \in \mathrm{Mod}(S_g)$ is represented by a homeomorphism ϕ whose order is equal to that of f. Moreover, ϕ is unique up to conjugacy in the group of homeomorphisms of S_g. We refer to any such ϕ as a *standard representative* of f.

The Birman–Hilden theorem. We now recall a theorem of Birman and Hilden. Let $i: S_g \to S_g$ be a hyperelliptic involution of S_g (a homeomorphism as in Figure 1) and let ι be the resulting hyperelliptic involution in $\mathrm{Mod}(S_g)$. Birman and Hilden proved [4, Theorem 1] that for $g \geq 2$ there is a short exact sequence

$$1 \to \langle \iota \rangle \to \mathrm{SMod}(S_g) \overset{\Theta}{\to} \mathrm{Mod}(S_{0,2g+2}) \to 1,$$

where $\mathrm{SMod}(S_g)$ is the centralizer in $\mathrm{Mod}(S_g)$ of ι and $S_{0,2g+2}$ is a sphere with $2g+2$ marked points. The map $\Theta: \mathrm{SMod}(S_g) \to \mathrm{Mod}(S_{0,2g+2})$ is defined as follows: it is proved [5, Theorem 4] that each element h of $\mathrm{SMod}(S_g)$ has a representative ψ that commutes with i, and so ψ can be pushed down to a homeomorphism $\overline{\psi}$ of the quotient $S_g/\langle i \rangle$, which is a sphere with $2g + 2$ marked points, namely, the images of the fixed points of ι. We note that the above exact sequence is not correct as stated for $g \leq 2$, as the given map $\mathrm{SMod}(S_g) \to \mathrm{Mod}(S_{0,2g+2})$ is not well-defined.

Roots of the hyperelliptic involution. The next lemma describes a property of roots of the hyperelliptic involution that will be used in the proof of Proposition 3.1.

Lemma 3.2. *Let $g \geq 1$. Suppose $f \in \mathrm{Mod}(S_g)$ is an nth root of a hyperelliptic involution (with $n > 1$). Then there is a power of f that is not the identity or the hyperelliptic involution and that has a standard representative with a fixed point.*

Proof. We first dispense with the case $g = 1$. In this case all periodic elements are given by rotations of either the hexagon or square. In particular, they all have fixed

points. Any nontrivial root f of the hyperelliptic involution is a periodic element, and so the first power of f satisfies the conclusion.

Now assume that $g \geq 2$. Fix a hyperelliptic involution $i: S_g \rightarrow S_g$ and corresponding mapping class ι as above. Suppose that f is the nth root of ι. It follows that f lies in $\mathrm{SMod}(S_g)$, the centralizer of ι. If Θ is the map from the aforementioned exact sequence of Birman and Hilden, then $\Theta(f)$ is a periodic element of $\mathrm{Mod}(S_{0,2g+2})$.

Let ψ be a standard representative of $\Theta(f)$. By ignoring the marked points, we may regard ψ as a finite-order homeomorphism of S^2. A theorem of Brouwer, Eilenberg, and de Kerékjártó [6, 13, 45] states that every finite-order orientation-preserving homeomorphism of S^2 is conjugate to a rotation. In particular, ψ has two fixed points in S^2.

There are two homeomorphisms of S_g that are lifts of ψ; they differ by the hyperelliptic involution i. One of these lifts is a standard representative ϕ for f and one is a standard representative for $f\iota$. If at least one of the fixed points of ψ is marked point, then both lifts of ψ to S_g have a fixed point. In particular, ϕ has a fixed point, as desired.

Now suppose that neither of the fixed points of ψ are marked points. In this case we will show that f^2 satisfies the conclusion of the lemma.

Let p be one of these fixed points and let $\widetilde{p}$ be one point of the preimage in S_g. One of the lifts of ψ fixes $\widetilde{p}$ and one interchanges it with $i(\widetilde{p})$. The one that fixes $\widetilde{p}$ cannot be a representative of f since no power of this homeomorphism is i (the only fixed points of i are the preimages of the marked points in $S_{0,2g+2}$). So the lift of ψ that interchanges $\widetilde{p}$ and $i(\widetilde{p})$ is a standard representative ϕ of f.

The homeomorphism ϕ^2 fixes $\widetilde{p}$. Clearly then ϕ^2 is also not equal to i or the identity, since i does not fix $\widetilde{p}$. Thus f^2 is not the identity or a hyperelliptic involution and its standard representative ϕ^2 fixes a point in S_g, as desired. $\square$

Proof of the theorem. We now prove Proposition 3.1, which, as described above, implies Theorem 1.1.

Proof of Proposition 3.1. Let f be a nontrivial periodic mapping class and assume that f is not a hyperelliptic involution. Let ϕ be a standard representative of f and let $\langle \phi \rangle$ denote the cyclic group of homeomorphisms of S_g generated by ϕ.

We treat separately three cases:

Case 1. the action of $\langle \phi \rangle$ is free;

Case 2. the action of $\langle \phi \rangle$ is not free and ϕ has order 2; and

Case 3. the action of $\langle \phi \rangle$ is not free and ϕ has order greater than 2.

We begin with the first case. If the action of $\langle \phi \rangle$ is free then this action is a covering space action. Every cyclic covering map $S_g \rightarrow S$ is equivalent to one of the covering maps indicated in Figure 4. In particular, we can find a curve c so that
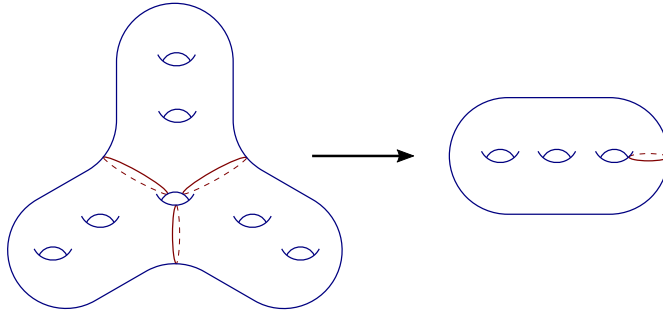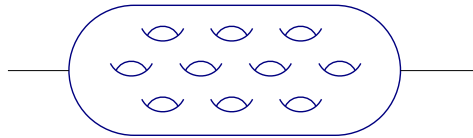
Figure 4. A cyclic cover of surfaces.

Figure 5. Rotation by π gives a mapping class of order 2.

f and c satisfy the hypotheses of Lemma 2.3. It follows from Lemma 2.3 that the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$, as desired.

We now treat the second case. There is a classification of homeomorphisms of S_g of order 2 that goes back to the work of Klein [24]; see Dugger's paper [12] for a modern treatment. In the cases where there is a fixed point in S_g, such a homeomorphism is conjugate to one of the ones indicated by Figure 5; there is a 1-parameter family, according to the number of handles above the axis of rotation. In particular, the conjugacy class of one of these homeomorphisms is completely determined by the genus of the quotient surface $S_g/\langle\phi\rangle$. When the genus of the quotient is 0, the homeomorphism is a hyperelliptic involution, which is ruled out by hypothesis. When the genus of the quotient is positive, we can again find a curve c as in Lemma 2.3. Again by Lemma 2.3 the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.

Finally, we treat the third case, where the order of ϕ is greater than 2 and the action of $\langle\phi\rangle$ on S_g is not free. Since the action of $\langle\phi\rangle$ on S_g is not free, some power of ϕ has a fixed point. We may choose this power so that ϕ has a fixed point and is not a hyperelliptic involution or the identity. This is obvious if f is not a root of a hyperelliptic involution and it follows from Lemma 3.2 otherwise. Without loss of generality we may replace ϕ by this power with a fixed point, since the normal closure of a power of f is contained in the normal closure of f; we may continue to assume that the order of the new ϕ is greater than two, for if not we may apply cases 1 and 2 above.
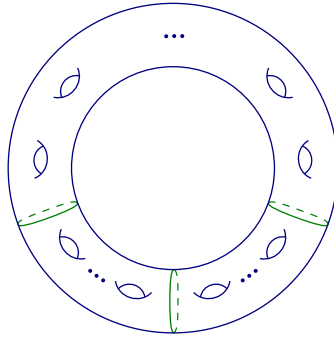
Figure 6. Three disjoint homologous curves in S_g.

A theorem of Kulkarni [26, Theorem 2] states that if ψ is a finite order homeomorphism of S_g that has a fixed point, then there is a way to represent S_g as a quotient space of some regular n-gon in such a way that ψ is given as rotation of the polygon by some multiple of $2\pi/n$. We apply this theorem to ϕ. Let P be the resulting n-gon (so S_g is obtained from P by identifying the sides of P in pairs in some way).

Let c be a line segment in P that connects the midpoints of two edges that are identified in S_g. Then c represents a curve in S_g. We may assume that c represents a nontrivial curve in S_g, for if all curves in S_g coming from these line segments were trivial, then we would have $g = 0$, in which case $\mathrm{Mod}(S_g)$ is trivial.

We may regard ϕ as a rotation of P by some multiple of $2\pi/n$. The image of the line segment c under ϕ is another line segment $\phi(c)$. Since ϕ has order greater than 2, it follows that $\phi(c)$ is not equal to c, and hence the number of intersections between these line segments is either 0 or 1. Moreover, this intersection number is equal to $i(c, f(c))$ (here we are regarding c as a curve in S_g).

If c is a separating curve in S_g, then the proposition follows from an application of the well-suited curve criterion for separating curves (Lemma 2.4). We may henceforth assume that c is nonseparating.

If $i(c, f(c)) = 1$, then it follows from Lemma 2.2 that the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$, as desired. Finally, suppose that $i(c, f(c)) = 0$. We would like to show that $[c] \neq [f(c)]$. To do this, it is enough to show that $[c]$, $[f(c)]$, and $[f^2(c)]$ are not all equal (if $[f(c)] = [c]$ then it would follow that $[f^2(c)] = [c]$).

Since the order of ϕ is greater than 2 the line segments c, $\phi(c)$, and $\phi^2(c)$ are all distinct. If $i(c, f^2(c)) = 1$ we may again use Lemma 2.2 to f^2 to conclude that the normal closure of f^2, hence f contains the commutator subgroup. So we may assume that c, $\phi(c)$, and $\phi^2(c)$ are all disjoint. If we cut S_g along the three curves corresponding to c, $\phi(c)$, and $\phi^2(c)$ then there is a region bordering all three curves, namely, the region of S_g containing the image of the center of P. On the

other hand, if we have three disjoint curves in S_g that are all homologous, then they must separate S_g into three components, and each region abuts two of the three curves (refer to Figure 6). Thus, $[c]$, $[f(c)]$, and $[f^2(c)]$ are not all equal and we are done. □

Normal closure of the hyperelliptic involution. To complete the classification of normal closures of periodic elements in $\mathrm{Mod}(S_g)$ it remains to deal with the case of the hyperelliptic involution and also the cases of $g = 1$ and $g = 2$. The following proposition already appears in the paper by Harvey and Korkmaz [19, Theorem 4(b)]. In the statement, let Ψ denote the standard symplectic representation $\mathrm{Mod}(S_g) \to \mathrm{Sp}_{2g}(\mathbb{Z})$.

Proposition 3.3. *Let $g \geq 3$ and let $\iota \in \mathrm{Mod}(S_g)$ be a hyperelliptic involution. Then the normal closure of ι in $\mathrm{Mod}(S_g)$ is the preimage of $\{\pm I\}$ under Ψ.*
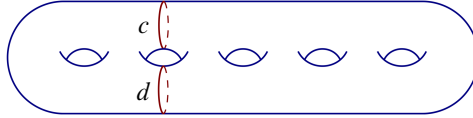

Figure 7. The normal closure of $T_c T_d^{-1}$ in $\mathrm{Mod}(S_g)$ is $\mathcal{I}(S_g)$.

Proof. The image of ι under Ψ is $-I$. Since the latter is central in $\mathrm{Sp}_{2g}(\mathbb{Z})$ it follows that the normal closure of ι is contained in the preimage of $\{\pm I\}$. It remains to show that the normal closure of ι contains the Torelli group $\mathcal{I}(S_g)$ (as in the introduction, $\mathcal{I}(S_g) = \ker \Psi$).

It is a theorem of D. Johnson [23, Theorem 2] that $\mathcal{I}(S_g)$ is equal to the normal closure in $\mathrm{Mod}(S_g)$ of the mapping class $T_c T_d^{-1}$, where c and d are the curves in S_g indicated in Figure 7 (this mapping class is called a bounding pair map of genus 1). Since $\iota(c) = d$ we have that

$$T_c T_d^{-1} = [T_c, \iota].$$

Thus, $T_c T_d^{-1}$ is equal to a product of two conjugates of ι. In particular, it is contained in the normal closure of ι. By Johnson's result, the normal closure of ι contains $\mathcal{I}(S_g)$ and we are done. □

Low genus cases. We now discuss the cases of $g = 1$ and $g = 2$. In both cases the hyperelliptic involution ι is unique and is central in $\mathrm{Mod}(S_g)$. Thus the normal closure of ι in both cases is the cyclic group of order 2 generated by ι. By Proposition 3.1 the normal closure of any other periodic element of $\mathrm{Mod}(S_g)$ contains the commutator subgroup, and hence the normal closure of an element f is completely determined by the image of f in the abelianization of $\mathrm{Mod}(S_g)$.

The abelianizations of $\mathrm{Mod}(S_1)$ and $\mathrm{Mod}(S_2)$ are isomorphic to $\mathbb{Z}/12\mathbb{Z}$ and $\mathbb{Z}/10\mathbb{Z}$, respectively; see [16, Section 5.1.3]. The image of any Dehn twist about a nonseparating curve in either case is equal to 1. Therefore, if f is a periodic element of $\mathrm{Mod}(S_g)$ that is not a hyperelliptic involution and f is equal to a product of k Dehn twists about nonseparating curves, then the normal closure of f in $\mathrm{Mod}(S_g)$ is equal to the preimage under the abelianization map $\mathrm{Mod}(S_g) \to \mathbb{Z}/n\mathbb{Z}$ of the group generated by k (here n is 12 when $g = 1$ and 10 when $g = 2$). In particular, if we consider any periodic element that is not a hyperelliptic involution then the normal closure has finite index in $\mathrm{Mod}(S_g)$. For a complete list of periodic elements in $\mathrm{Mod}(S_1)$ and $\mathrm{Mod}(S_2)$ and realizations of these elements as products of Dehn twists about nonseparating curves, see the paper of Hirose [21, Theorem 3.2] (we note that the second $f_{2,3}$ in Hirose's theorem should be $f_{2,4}$). In each case the image of the periodic element is not a generator for the abelianization. So for g equal to 1 or 2, the normal closure of any periodic element is a proper subgroup of $\mathrm{Mod}(S_g)$.

Finite generating sets. We already mentioned results of Harvey, Korkmaz, McCarthy, Papadopoulos, Yoshihara, and the first author of this paper that give periodic normal generators for the mapping class group. Much more than this, Korkmaz proved that only two conjugate elements of order $4g + 2$ are needed to generate $\mathrm{Mod}(S_g)$. Yoshihara proved that three conjugate elements of order 6 are required for $g \geq 10$. Finally, the first author proved that for $k \geq 6$ and $g \geq (k-1)^2 + 1$ only three conjugate elements of order k are needed, and for $k = 5$ and $g \geq 8$ only four conjugate elements are needed. Based on these results and Theorem 1.1 we are led to the following question.

Question 3.4. *Is there a number N, independent of g, so that if f is a periodic normal generator of $\mathrm{Mod}(S_g)$ then $\mathrm{Mod}(S_g)$ is generated by N conjugates of f?*

We emphasize that in the question N is independent of both g and f.

Power subgroups. As in the introduction, let $L = L(g)$ denote the least common multiple of the orders of the periodic elements of $\mathrm{Mod}(S_g)$. We have the following corollary of Theorem 1.1.

Corollary 3.5. *Let $g \geq 3$. Suppose that n is not divisible by $L/2$. Then the subgroup of $\mathrm{Mod}(S_g)$ generated by the nth powers of all elements is equal to the whole group $\mathrm{Mod}(S_g)$.*

This result improves on a theorem of Funar [18, Theorem 1.16 (2)] who showed the analogous result with $L/2$ replaced by $4g + 2$. Note that $4g + 2$ is a factor of $L/2$ since L has 4 and $2g + 1$ as factors. Thus our result indeed recovers the theorem of Funar. For comparison, when $g = 4$ the number $L/2$ is 360 and $4g + 2$ is 18.

Proof of Corollary 3.5. Since $L/2$ does not divide n, then for some prime p we have that $L/2$ has a factor p^j whereas n only has a factor p^k with $j > k \geq 0$.

If this p is an odd prime, then $\mathrm{Mod}(S_g)$ contains a periodic element f whose order has p^j as a factor. The element f^n is nontrivial and its order has p^{j-k} as a factor. Since p is odd it follows that f^n is not a hyperelliptic involution. By Theorem 1.1, f^n is a normal generator for $\mathrm{Mod}(S_g)$. Since f^n is contained in the nth power subgroup we are done in this case.

If instead p is 2, then $L/2$ has a factor 2^j whereas n only has a factor 2^k with $j > k \geq 0$. But then L has a factor 2^{j+1}. This implies that $\mathrm{Mod}(S_g)$ contains a periodic element f whose order has 2^{j+1} as a factor, and we have that f^n is nontrivial and its order has 2^{j+1-k} as a factor, which is at least 4. Hence f^n is nontrivial and not the hyperelliptic involution. Again an application of Theorem 1.1 completes the proof. $\square$

If we only consider periodic elements when analyzing power subgroups, we could at most hope to replace $L/2$ by L in the corollary (as the Lth power of every periodic element is trivial). However, for some g the analysis fails with this replacement. For instance, when g is a power of 2, the element of order $4g$ is the only element that is nontrivial when raised to the power $L/2$, and this power is the hyperelliptic involution.

Of course, there do exist values of n so that the nth power subgroup of $\mathrm{Mod}(S_g)$ is not $\mathrm{Mod}(S_g)$. Indeed for any group G with a proper subgroup H of finite index, there is an n so that the nth power group of G is a subgroup of H and hence is not equal to G.

Surfaces with marked points. We now give an extension of Theorem 1.1 to the case of surfaces with marked points. As in the introduction, by a hyperelliptic involution in $\mathrm{Mod}(S_{g,n})$ we mean a periodic element that maps to a hyperelliptic involution in $\mathrm{Mod}(S_g)$ under the forgetful map $\mathrm{Mod}(S_{g,n}) \to \mathrm{Mod}(S_g)$.

Theorem 3.6. *For $g \geq 3$ and $n \geq 1$, the normal closure of any nontrivial periodic mapping class that is not a hyperelliptic involution contains $\mathrm{PMod}(S_{g,n})$. In particular, when $n = 1$, the normal closure is equal to $\mathrm{Mod}(S_{g,n})$.*

Let Σ_n denote the symmetric group on n letters. There is a short exact sequence

$$1 \to \mathrm{PMod}(S_{g,n}) \to \mathrm{Mod}(S_{g,n}) \to \Sigma_n \to 1.$$

Let $\mathrm{AMod}(S_{g,n})$ denote the preimage in $\mathrm{Mod}(S_{g,n})$ of the alternating group A_n; this is a subgroup of $\mathrm{Mod}(S_{g,n})$ of index 2. Because the only normal subgroups of Σ_n are the trivial group, A_n, Σ_n, and (when $n = 4$) the Klein four group, it follows from Theorem 3.6 that the only possibilities for the normal closure of a nontrivial, nonhyperelliptic, periodic mapping class $f \in \mathrm{Mod}(S_{g,n})$ with $g \geq 3$ and $n \neq 4$ are $\mathrm{PMod}(S_{g,n})$, $\mathrm{AMod}(S_{g,n})$, and $\mathrm{Mod}(S_{g,n})$, according to whether the image of f in Σ_n is trivial, nontrivial even, or odd.

Proof of Theorem 3.6. Let f be a nontrivial periodic mapping class in $\mathrm{Mod}(S_{g,n})$ that is not a hyperelliptic involution. Since $\mathrm{PMod}(S_{g,n})$ is perfect, it suffices to show that the normal closure of f contains the commutator subgroup of $\mathrm{PMod}(S_{g,n})$. As in the proof of Proposition 3.1, we accomplish this by establishing well-suited curve criteria for $\mathrm{Mod}(S_{g,n})$ and then verifying that f satisfies one of these criteria.

The statements and proofs of the well-suited curve criteria established in Section 2 carry over to the case of surfaces with marked points, with S_g replaced by $S_{g,n}$, with $\mathrm{Mod}(S_g)$ replaced by $\mathrm{PMod}(S_{g,n})$, and with the following caveats. In the statement of Lemma 2.3, we should interpret $[c] \neq [f(c)]$ to mean that the images of c and $f(c)$ in S_g are not homologous mod 2. Similarly, in the statement of Lemma 2.4, we should interpret the assumption that d is separating to mean that the image of d in S_g is an essential separating curve. A key point is that, like in the case $n = 0$, the group $\mathrm{PMod}(S_{g,n})$ has a generating set consisting of Dehn twists about nonseparating curves, where the pairwise intersection numbers of the curves is at most 1; see [16, Section 4.4.4].

As in the proof of Theorem 1.1, we take ϕ to be a standard representative of f. As in the case $n = 0$, this is a homeomorphism of $S_{g,n}$ that represents f and has the same order. An essential point is that ϕ may be regarded as a homeomorphism of S_g. Because of this, we may apply the same classification results and theorems used in the proof of Proposition 3.1. As in the proof of Theorem 1.1 we consider three cases, according to whether the action of $\langle \phi \rangle$ is free, the action is not free and the order of ϕ is 2, or the action is not free and the order of ϕ is greater than 2.

In the first two cases, the proof is the same; we simply choose the curve c to avoid the marked points. In the third case, we may need to make a small perturbation of the line segment c in order to avoid the marked points; the proof then applies as stated. $\qquad\square$

4. Application: Long's question and Ivanov's conjecture

In this section we construct, for each $g \geq 1$, a pseudo-Anosov mapping class that normally generates $\mathrm{Mod}(S_g)$ (for $g = 1$ pseudo-Anosov mapping classes are usually called Anosov, but we will not make this distinction). As discussed in the introduction, it follows from our Theorem 1.2 that such mapping classes exist when $g \geq 3$. The examples in this section are simple and explicit and only use the special cases of the well-suited curve criterion given in Section 2, whereas the proof of Theorem 1.2 is more involved and requires the more general version of the well-suited curve criterion given in Section 7.

We begin with a very simple family of pseudo-Anosov mapping classes that normally generate, then we give examples that have large stretch factor, examples that have small stretch factor, and examples that have large translation length on the curve graph. Finally, we give examples with large stretch factor with normal

closure equal to any given congruence subgroup. As described in the introduction, the examples in this section answer Long's question in the affirmative and resolve Ivanov's conjecture in the negative.

First examples. For each $g \geq 1$, let $T_A = T_{a_1} \cdots T_{a_g}$ and $T_B = T_{b_1} \cdots T_{b_{g+1}}$, where the a_i and b_i are the curves in S_g indicated in Figure 8. The figure shows the case $g = 3$ but there is an obvious generalization for all other $g \geq 1$; when $g = 1$ the curves b_1 and b_2 are parallel.

Consider the mapping class $f = T_A^{-1} T_B$ (note f depends on g). The following proposition answers in the affirmative the question of Long from the introduction.
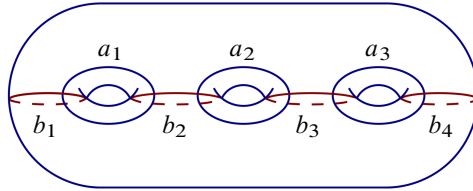

Figure 8. The a_i and b_i used in the definition of T_A and T_B.

Proposition 4.1. *For each $g \geq 1$ the mapping class f is pseudo-Anosov and it normally generates $\mathrm{Mod}(S_g)$.*

Proof. There are well-known constructions of pseudo-Anosov mapping classes due to Thurston and to Penner where certain products of Dehn twists are shown to be pseudo-Anosov; see [16, Theorems 14.1 and 14.4]. The product $f = T_A^{-1} T_B$ is both an example of the Thurston construction and an example of the Penner construction. In particular, f is pseudo-Anosov for all $g \geq 1$.

Consider the action of f on b_1. Since $i(b_1, b_i)$ is equal to 0 for all i and $i(b_1, a_i)$ is equal to 0 for $i > 1$ we have that $f(b_1) = T_{a_1}^{-1}(b_1)$. It follows that $i(f(b_1), b_1)$ is equal to 1. Thus by Lemma 2.2 the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$. For $g \geq 3$ it follows that f normally generates $\mathrm{Mod}(S_g)$. For g equal to 1 or 2, it is enough now to observe that the sum of the exponents of the Dehn twists in the definition of f is 1. As such the image of f in the abelianization of $\mathrm{Mod}(S_g)$ is 1, and hence a generator for the abelianization. Thus, in these cases, f is a normal generator as well. $\qquad\square$

Examples with large stretch factor. There is great flexibility in the construction of f. Indeed, we can alter any of the exponents, except on T_{a_1}, and we obtain another normal generator for $\mathrm{Mod}(S_g)$ when $g \geq 3$. If we also preserve the condition that the exponent sum is relatively prime to 12 or 10, we obtain a normal generator for $\mathrm{Mod}(S_1)$ or $\mathrm{Mod}(S_2)$, respectively. We take advantage of this flexibility to prove the following.

Proposition 4.2. *For each $g \geq 1$ there are pseudo-Anosov mapping classes with arbitrarily large stretch factors that normally generate* $\mathrm{Mod}(S_g)$.

Proof. Fix some $g \geq 1$ and consider again the mapping classes T_A and T_B as above. Much more than proving $T_A^{-1} T_B$ is pseudo-Anosov, Thurston proved that there is a positive number μ and a homomorphism

$$\langle T_A, T_B \rangle \to \mathrm{PSL}_2(\mathbb{R})$$

with

$$T_A \mapsto \begin{bmatrix} 1 & -\mu^{1/2} \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad T_B \mapsto \begin{bmatrix} 1 & 0 \\ \mu^{1/2} & 1 \end{bmatrix},$$

and so that $f \in \langle T_A, T_B \rangle$ is pseudo-Anosov if and only if its image in $\mathrm{PSL}_2(\mathbb{R})$ (or, rather, a representative in $\mathrm{SL}_2(\mathbb{R})$) has an eigenvalue $\lambda > 1$; moreover in this case λ is the stretch factor of f (the number μ is the Perron–Frobenius eigenvalue of NN^T, where N is the intersection matrix between the a_i and b_i; see [16]).

Consider then the mapping class $f_n = T_A^{-1} T_B^n$. Its image in $\mathrm{PSL}_2(\mathbb{R})$ is

$$\begin{bmatrix} 1 & \mu^{1/2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ n\mu^{1/2} & 1 \end{bmatrix} = \begin{bmatrix} 1 + n\mu & \mu^{1/2} \\ n\mu^{1/2} & 1 \end{bmatrix}.$$

The trace of this matrix is larger than 2, and so it has two real eigenvalues. The larger of these eigenvalues is strictly larger than the trace, which is $2 + n\mu$. In particular, each mapping class $T_A^{-1} T_B^n$ is pseudo-Anosov, and the stretch factors tend to infinity as n tends to infinity. As in the proof of Proposition 4.1, we have $i(f_n(b_1), b_1) = 1$ and so by Lemma 2.2 the normal closure of each of these mapping classes contains the commutator subgroup of $\mathrm{Mod}(S_g)$. This completes the proof for $g \geq 3$.

For $g = 2$ the image of f_n in the abelianization of $\mathrm{Mod}(S_2)$ is $3n - 2$. For each n equal to 1, 3, 5, or 7 mod 10 the image of f_n is therefore relatively prime to 10 and so the proposition is proven for $g = 2$. For $g = 1$ the image of f_n is $2n - 1$, which is relatively prime to 12 for n equal to 0, 1, 3, or 4 mod 6. $\qquad \square$

Examples with small stretch factor. There are many other explicit examples of pseudo-Anosov mapping classes that satisfy the special cases of the well-suited curve criterion given in Section 2 and hence normally generate $\mathrm{Mod}(S_g)$. Here we discuss one famous family of examples due to Penner.

Consider the curves a, b, and c in S_g indicated in Figure 9 and consider the order g rotation ρ of S_g indicated in the same figure. Penner proved that the product $\rho T_c T_b^{-1} T_a$ is pseudo-Anosov (to prove this he notes that the gth power is an example of his Dehn twist construction for pseudo-Anosov mapping classes). By inspection we observe that Penner's mapping class, together with the curve d in Figure 9, satisfies Lemma 2.3. We thus have the following proposition.
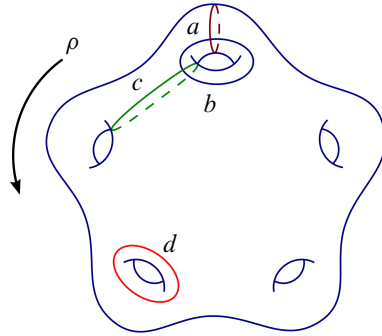
Figure 9. The curves and the rotation used in Penner's construction of pseudo-Anosov mapping classes with small stretch factor.

Proposition 4.3. *For each $g \geq 3$, Penner's mapping class $\rho T_c T_b^{-1} T_a$ is a normal generator for* $\mathrm{Mod}(S_g)$.

For $g = 2$ the mapping class ρ maps to 5 in $\mathbb{Z}/10\mathbb{Z}$ and so Penner's mapping class maps to 6. Thus the image of Penner's mapping class has index 2 in $\mathbb{Z}/10\mathbb{Z}$. On the other hand, $\rho T_c^2 T_b^{-2} T_a^2$ is pseudo-Anosov and normally generates.

Penner proved that the stretch factors of his mapping classes are bounded above by $11^{1/g}$ for all $g \geq 2$. Hence they form a sequence of pseudo-Anosov mapping classes whose stretch factors tend to 1 as g tends to infinity. As such, these examples served as the initial inspiration for our Theorem 1.2, which says that all pseudo-Anosov mapping classes with sufficiently small stretch factor are normal generators.

Smallest stretch factors. For $g \geq 3$ it is an open problem to determine which mapping classes realize the smallest stretch factor. Nevertheless, we do know in these cases that the smallest stretch factors are less than $\sqrt{2}$ and so by Theorem 1.2 the minimizers are normal generators.

For g equal to 1 or 2, we know precisely which mapping classes realize the smallest stretch factor in $\mathrm{Mod}(S_g)$. For $g = 1$ the minimum is given uniquely by the conjugacy class of $f_1 = T_{a_1} T_{b_1}^{-1}$, where a_1 and b_1 are as shown in Figure 8. The stretch factor of f_1 is $\phi^2 = (3 + \sqrt{5})/2$. The mapping class f_1 is not a normal generator since the image of f_1 in $\mathbb{Z}/12\mathbb{Z}$, the abelianization of $\mathrm{Mod}(S_1)$, is 0. On the other hand, the normal closure of f_1 does contain the commutator subgroup of $\mathrm{Mod}(S_1)$. This is because $i(f_1(b_1), b_1) = 1$ and so f_1 satisfies Lemma 2.2.

For $g = 2$, Cho and Ham [9] proved that the minimum stretch factor is the largest real root of $x^4 - x^3 - x^2 - x + 1$, which is approximately 1.72208. Lanneau and Thiffeault [29] gave an independent proof and also classified the conjugacy classes of all of the minimizers and gave explicit representatives of these conjugacy classes in terms of Dehn twists. Again using the labels in Figure 8, these are:

$$f_2 = T_{a_2} T_{b_3}^{-1} T_{a_1} T_{b_1}^2 T_{b_2} \quad \text{and} \quad f_2' = T_{a_2}^{-1} T_{b_2}^{-1} T_{b_3}^{-1} T_{a_1} T_{b_1}^2.$$

The images of these mapping classes in the abelianization $\mathbb{Z}/10\mathbb{Z}$ are 4 and 0, respectively, and so neither is a normal generator. As in the $g = 1$ case the normal closures of both f_2 and f_2' contain the commutator subgroup of $\text{Mod}(S_2)$ since $i(f_2(b_3), b_3) = 1$ and $i(f_2'(b_3), b_3) = 1$.

Examples with large translation length. In this section, we have already given examples of pseudo-Anosov normal generators for $\text{Mod}(S_g)$ with either small or large stretch factor. All of these examples have translation length in $\mathcal{C}(S_g)$ at most 2 since they each satisfy one of the well-suited curve criteria given in Lemmas 2.2 and 2.3. It follows that the asymptotic translation lengths are also bounded above by 2.

We will now prove Theorem 1.4, which states that for each $g \geq 3$ there are pseudo-Anosov mapping classes with the property that all of their odd powers are normal generators. Since the asymptotic translation length on the curve graph is multiplicative on cyclic subgroups, and since the asymptotic translation length of every pseudo-Anosov mapping class is positive, it follows that there are pseudo-Anosov normal generators with arbitrarily large asymptotic translation lengths. Because large translation length implies large stretch factor, Theorem 1.4 implies Proposition 4.2 in the cases when $g \geq 3$.

Proof of Theorem 1.4. Fix $g \geq 3$. Let D_{2g} denote the dihedral group of order $2g$. There is a standard action of D_{2g} on S_g by orientation-preserving homeomorphisms. In this way, we identify D_{2g} as a subgroup of $\text{Homeo}^+(S_g)$. We denote the quotient S_g/D_{2g} by X. As an orbifold, X is a sphere with five orbifold points.

Consider any pseudo-Anosov homeomorphism of X, thought of as a sphere with five marked points. Up to taking a power, we may assume that this homeomorphism lifts to a pseudo-Anosov homeomorphism ψ of S_g; see [16, Section 14.1.1].

The homeomorphism ψ lies in the normalizer of D_{2g}, and so we may identify ψ with an element of the automorphism group of D_{2g}; it acts by conjugation. Let r be the element of D_{2g} corresponding to rotation by $2\pi/g$ under the standard action of D_{2g} on the g-gon. Since r and r^{-1} are the only elements of D_{2g} that are conjugate in $\text{Homeo}^+(S_g)$, it must be that the automorphism of D_{2g} induced by ψ maps r to either r or r^{-1}. Let $k \in D_{2g}$ correspond to some reflection of the g-gon. Up to replacing ψ by $k\psi$ we may assume that the automorphism of D_{2g} induced by ψ maps r to r^{-1}. Since k fixes the invariant foliations for ψ and does not change the stretch factor, the new ψ is a pseudo-Anosov element f of $\text{Mod}(S_g)$. If we identify r as an element of $\text{Mod}(S_g)$, then $frf^{-1} = r^{-1}$. By construction $f^n r f^{-n} = r^{-1}$ for all odd n.

Consider the mapping class $[r, f^n]$ for n odd. We have

$$[r, f^n] = r(f^n r f^{-n})^{-1} = r^2.$$

Thus, r^2 lies in the normal closure of f^n in $\text{Mod}(S_g)$. By the well-suited curve criterion of Lemma 2.3 (and using the assumption $g \geq 3$) the mapping class r^2, hence f^n, is a normal generator for $\text{Mod}(S_g)$. $\qquad\square$

Torelli groups and congruence subgroups. Having found an abundance of pseudo-Anosov normal generators for the mapping class group, we now consider the question of which proper normal subgroups of $\mathrm{Mod}(S_g)$ arise as the normal closure of a single pseudo-Anosov mapping class.

Specifically, our next goal is to prove Theorem 1.3, which states that all level m congruence subgroups $\mathrm{Mod}(S_g)[m]$ arise as the normal closure of a pseudo-Anosov mapping class, and moreover that this mapping class can be chosen to have arbitrarily large stretch factor. The theorem covers the Torelli group as the case where $m = 0$. The construction in the proof is a variant of a construction of Leininger and the second author [30, Proof of Proposition 5.1].

The proof of Theorem 1.3 requires the following technical lemma, whose statement and proof are well known to experts, but are not easily found in the literature.

Lemma 4.4. *Let* $A = \{a_1, \ldots, a_m\}$ *and* $B = \{b_1, \ldots, b_n\}$ *be two multicurves that fill* S_g. *The stretch factors of the mapping classes*

$$f_k = \left(T_{a_1}^k \cdots T_{a_m}\right)\left(T_{b_1} \cdots T_{b_n}\right)^{-1}$$

tend to infinity as k does.

Since the T_{a_i} all commute with each other, since the T_{b_i} all commute with each other, and since the stretch factor of a mapping class is the same as that of its inverse, the lemma holds equally well if we attach the exponent k to any T_{a_i} or attach the exponent $-k$ to any T_{b_i}.

Proof of Lemma 4.4. Consider the intersection graph $\Gamma(A, B)$ with vertices $A \cup B$ and $i(a_i, b_j)$ edges for each pair $\{a_i, b_j\}$ (this is really a multigraph). Since A and B fill S_g, this graph is connected. Let D be half the maximum path-length distance between two A-vertices of $\Gamma(A, B)$.

Let A_k denote the multicurve obtained from A by replacing a_1 with k parallel copies of a_1, denoted $a_{1,1}, \ldots, a_{1,k}$. Let T_{A_k} and T_B be the products of the Dehn twists about the curves in A_k and B, respectively. We thus have

$$f_k = T_{A_k} T_B^{-1}.$$

Let $\Gamma(A_k, B)$ be the intersection graph for A_k and B. This graph has $m + n + k - 1$ vertices and is the blowup of $\Gamma(A, B)$ obtained by replacing the vertex a_1 with k vertices $a_{1,1}, \ldots, a_{1,k}$.

Let N be the $(n + k - 1) \times m$ matrix whose rows correspond to the curves of A_k, whose columns correspond to the curves of B, and whose entries are given by the corresponding intersection numbers. The matrix NN^T is a square matrix with rows and columns corresponding to the curves of A_k. Each entry is the number of paths of length 2 between the corresponding vertices of $\Gamma(A_k, B)$. Similarly, the entries of any power $(NN^T)^\ell$ are the numbers of paths of length 2ℓ between corresponding vertices of $\Gamma(A_k, B)$.

Similar to the proof of Proposition 4.2, the stretch factor of f_k is bounded below by $2 + \mu$, where μ is the Perron–Frobenius eigenvalue of NN^T. Thus it suffices to show that the latter tends to infinity with k.

Consider the matrix $(NN^T)^{D+1}$. Since D does not depend on k and since the Perron–Frobenius eigenvalue of $(NN^T)^{D+1}$ is the $(D+1)$st power of the Perron–Frobenius eigenvalue of NN^T it further suffices to show that the Perron–Frobenius eigenvalue of $(NN^T)^{D+1}$ tends to infinity with k.

The Perron–Frobenius eigenvalue of a Perron–Frobenius matrix is bounded below by the minimum row sum of the matrix. Fix a vertex v of $\Gamma(A_k, B)$ corresponding to a component of A_k. The corresponding row sum of NN^T is the number of paths of length $2D + 2$ in $\Gamma(A_k, B)$ starting at v.

By the definition of D and the definition of the blowup $\Gamma(A_k, B)$, there is a path from v to $a_{1,1}$ with length $L \leq 2D$. There are at least k ways to extend this path to a path of length $L + 2 \leq 2D + 2$, since $a_{1,1}$ is connected to some b_j and this b_j is connected to each of the blown-up vertices $a_{1,1}, \ldots, a_{1,k}$. Thus the minimum row sum of $(NN^T)^{D+1}$ is at least k, and the lemma follows. $\qquad\square$

The proof of Lemma 4.4 can be modified to prove a much more general result, namely, that for $A = \{a_1, \ldots, a_m\}$ and $B = \{b_1, \ldots, b_n\}$ as in the lemma, the stretch factors of the mapping classes

$$\left(T_{a_1}^{p_1} \cdots T_{a_m}^{p_m}\right)\left(T_{b_1}^{q_1} \cdots T_{b_n}^{q_n}\right)^{-1}$$

tend to infinity as the maximum of $\{p_1, \ldots, p_m, q_1, \ldots, q_n\}$ does. A special case of this fact was used (and proved) in the proof of Proposition 4.2 above. The proof of the more general fact is essentially the same, but is notationally unwieldy. The statement of Lemma 4.4 suffices for the proof of Theorem 1.3.

Proof of Theorem 1.3. As in the statement, let $g \geq 3$ be fixed. Let

$$a_0, \ldots, a_{\lceil g/2 \rceil}, b_0, \ldots, b_{\lfloor g/2 \rfloor}, d$$

be the curves in S_g as shown in Figure 10 for the case $g = 8$. The curve d is disjoint from each a_i when g is odd and it is disjoint from each b_i when g is even. Consider the mapping class

$$f_{m,k} = \left(T_{b_0}^{-1}(T_d^m)^{(-1)^{g+1}} T_{a_0}\right)\left(T_{a_1}^k \cdots T_{a_{\lceil g/2 \rceil}}\right)\left(T_{b_1} \cdots T_{b_{\lfloor g/2 \rfloor}}\right)^{-1}.$$

This mapping class is pseudo-Anosov. Indeed when g is odd it is conjugate by $T_{b_0}^{-1}$ to a pseudo-Anosov mapping class arising from the Thurston construction and when g is even it is conjugate by $T_{b_0}^{-1}T_d^{-m}$ to a pseudo-Anosov mapping class arising from the Thurston construction.

For each m and k, there is a curve c so that c and $f_{m,k}(c)$ form a bounding pair of genus 1, an example of which is shown in Figure 7. The curve c is shown
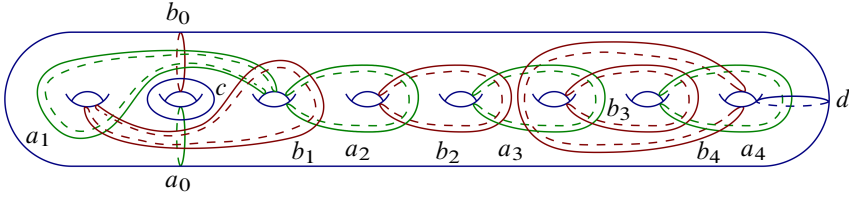
Figure 10. The curves used in the proof of Theorem 1.3 in the case $g = 8$.

in Figure 10. As in the proof of Proposition 3.3 the normal closure of $T_c T_{f_{m,k}(c)}^{-1}$ in $\mathrm{Mod}(S_g)$ is $\mathcal{I}(S_g)$. Since

$$T_c T_{f_{m,k}(c)}^{-1} = [T_c, f_{m,k}],$$

we see that the normal closure of $f_{m,k}$ in $\mathrm{Mod}(S_g)$ contains $\mathcal{I}(S_g)$.

The normal closure of $f_{m,k}$ is thus completely determined by its image under the symplectic representation $\Psi \colon \mathrm{Mod}(S_g) \to \mathrm{Sp}_{2g}(\mathbb{Z})$. Since each T_{a_i} and T_{b_i} lies in $\mathcal{I}(S_g)$ we have that $\Psi(f_{m,k})$ is equal to $\Psi(T_d^m)$. The latter is the mth power of a transvection about a primitive element of $\mathbb{Z}^{2g}$. Thus, by work of Mennicke [41, Section 10] its normal closure in $\mathrm{Sp}_{2g}(\mathbb{Z})$ is the level m congruence subgroup $\mathrm{Sp}_{2g}(\mathbb{Z})[m]$, which is the kernel of the reduction homomorphism

$$\mathrm{Sp}_{2g}(\mathbb{Z}) \to \mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$$

(it is essential here that $g > 1$). The preimage under Ψ of $\mathrm{Sp}_{2g}(\mathbb{Z})[m]$ is the level m congruence subgroup $\mathrm{Mod}(S_g)[m]$, as desired.

By Lemma 4.4 the stretch factor of $f_{m,k}$ tends to infinity as k does (in order to apply Lemma 4.4 directly we should replace T_d^m with $T_{d_1} \cdots T_{d_m}$ where each d_i is parallel to d). This completes the proof.   □

5. Application: linear groups

In this section we use the results we have proven so far about normal generators for mapping class groups to produce many examples of normal generators for certain linear groups. Using the standard surjective representation $\Psi \colon \mathrm{Mod}(S_g) \to \mathrm{Sp}_{2g}(\mathbb{Z})$, we can apply our results about mapping class groups to show that certain elements of $\mathrm{Sp}_{2g}(\mathbb{Z})$ are normal generators for $\mathrm{Sp}_{2g}(\mathbb{Z})$.

The normal closure of $\mathrm{Sp}_{2g}(\mathbb{Z})$ in $\mathrm{SL}_{2g}(\mathbb{Z})$ is $\mathrm{SL}_{2g}(\mathbb{Z})$. Thus, any normal generator for $\mathrm{Sp}_{2g}(\mathbb{Z})$ can also be interpreted as a normal generator for $\mathrm{SL}_{2g}(\mathbb{Z})$.

To begin, we obtain many examples of normal generators for $\mathrm{Sp}_{2g}(\mathbb{Z})$ by a direct application of Theorem 1.1. That is, if an element of $\mathrm{Sp}_{2g}(\mathbb{Z})$ is the image of a
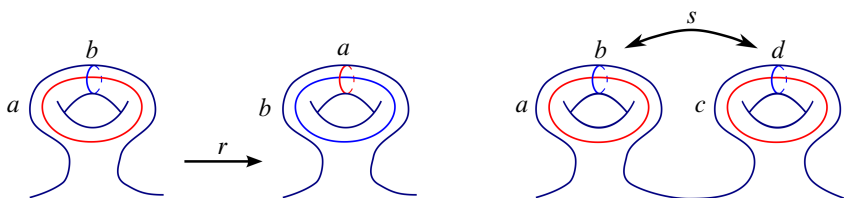
Figure 11. A handle rotation and a handle swap.

nontrivial periodic element that is not a hyperelliptic involution then it is a normal generator for $\mathrm{Sp}_{2g}(\mathbb{Z})$.

What is more, there are periodic elements of $\mathrm{Sp}_{2g}(\mathbb{Z})$ that are not images of periodic elements but are images of elements that satisfy the well-suited curve criterion. For instance, consider the block matrix

$$M = \begin{bmatrix} A & 0 \\ 0 & I_{2g-2} \end{bmatrix},$$

where A is the 2×2 matrix

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and I_n denotes the $n \times n$ identity matrix. Here and throughout all matrices in $\mathrm{Sp}_{2g}(\mathbb{Z})$ are written with respect to a standard symplectic basis $(x_1, y_1, \ldots, x_g, y_g)$ for $\mathbb{Z}^{2g}$.

The periodic matrix M is the image of the handle rotation r indicated in the left-hand side of Figure 11. The handle rotation r has infinite order, but it satisfies Lemma 2.2. Thus, r is a normal generator for $\mathrm{Mod}(S_g)$ and so M is a normal generator for $\mathrm{Sp}_{2g}(\mathbb{Z})$.

Similarly, consider the block matrix

$$N = \begin{bmatrix} B & 0 \\ 0 & I_{2g-4} \end{bmatrix},$$

where B is the 4×4 block matrix

$$\begin{bmatrix} 0 & I_2 \\ I_2 & 0 \end{bmatrix}.$$

The matrix N is the image of the handle swap s indicated in the right-hand side of Figure 11. Again s has infinite order, but since it satisfies Lemma 2.3 it is a normal generator for $\mathrm{Mod}(S_g)$, implying that N is a normal generator for $\mathrm{Sp}_{2g}(\mathbb{Z})$.

We can also use our pseudo-Anosov examples from Section 4 to find examples of irreducible (in fact, primitive) normal generators for $\mathrm{Sp}_{2g}(\mathbb{Z})$. Indeed, if we take the curves $(b_1, a_1, \ldots, b_g, a_g)$ in Figure 8 as a basis for $H_1(S_g; \mathbb{Z})$ (with appropriate orientations) then the mapping class

$$\left(T_{a_1} \cdots T_{a_g} \right)^{-1} \left(T_{b_1} \cdots T_{b_g} \right)$$

has primitive image in $\mathrm{Sp}_{2g}(\mathbb{Z})$. This image is a normal generator since the mapping class is.

One might hope to apply Theorem 1.1 and the surjectivity of the map $\mathrm{Mod}(S_g) \to \mathrm{Sp}_{2g}(\mathbb{Z})$ in order to show that every nontrivial, noncentral finite-order element of $\mathrm{Sp}_{2g}(\mathbb{Z})$ is a normal generator for $\mathrm{Sp}_{2g}(\mathbb{Z})$. This fails, however, because there are finite-order elements of $\mathrm{Sp}_{2g}(\mathbb{Z})$ so that each element of the preimage in $\mathrm{Mod}(S_g)$ has infinite order. For instance consider the elements of $\mathrm{Sp}_{2g}(\mathbb{Z})$ given by the block matrices

$$ M_k = \begin{bmatrix} -I_{2k} & 0 \\ 0 & I_{2g-2k} \end{bmatrix}, $$

where I_n is the $n \times n$ identity matrix. For $0 < k < g$ the matrix M_k does not have periodic preimage in $\mathrm{Mod}(S_g)$ (for $k = 0$ the identity lies in the preimage and for $k = g$ the hyperelliptic involution does). Moreover, the normal closure of M_k cannot be the entire group $\mathrm{Sp}_{2g}(\mathbb{Z})$ since M_k lies in the level 2 congruence subgroup, which is a proper normal subgroup of $\mathrm{Sp}_{2g}(\mathbb{Z})$. We suspect these are the only counterexamples.

Putman has pointed out to us another approach to showing that elements of $\mathrm{Sp}_{2g}(\mathbb{Z})$ are normal generators. By the Margulis normal subgroup theorem the normal closure of a nontrivial, noncentral element M of $\mathrm{Sp}_{2g}(\mathbb{Z})$ has finite index in $\mathrm{Sp}_{2g}(\mathbb{Z})$. Since $\mathrm{Sp}_{2g}(\mathbb{Z})$ enjoys the congruence subgroup property, this normal closure contains a level m congruence subgroup. Therefore, the problem of showing that M is a normal generator reduces to the problem of showing that the image of M in the finite group $\mathrm{Sp}_{2g}(\mathbb{Z}/m\mathbb{Z})$ is a normal generator.

6. Stretch factors and intersection numbers

The goal of the remaining five sections is to prove Theorem 1.2, which states that any pseudo-Anosov mapping class with stretch factor less than $\sqrt{2}$ is a normal generator. In this section we take the first step by relating the hypothesis about small stretch factor to the existence of a curve that has small intersection number with its image. After making this connection, we outline the rest of the proof of Theorem 1.2. We begin with some preliminaries about singular Euclidean metrics for pseudo-Anosov mapping classes.

Metrics for pseudo-Anosov mapping classes. For a pseudo-Anosov mapping class f with stretch factor λ there is a representative ϕ and two transverse measured foliations $\mathcal{F}_u$ and $\mathcal{F}_s$ on S_g with

$$ \phi(\mathcal{F}_u) = \lambda \mathcal{F}_u \quad \text{and} \quad \phi(\mathcal{F}_s) = \lambda^{-1} \mathcal{F}_s $$

(see [16, Section 13.2.3]). The measured foliations $\mathcal{F}_u$ and $\mathcal{F}_s$ determine a singular Euclidean metric on S_g. If we insist that this metric has unit area then there is a

one parameter family of choices for the pair $(\mathcal{F}_u, \mathcal{F}_s)$ and any choice suffices for our purposes; we refer to any one of these as an f-*metric* on S_g. An f-metric is also an f^k-metric for all nonzero k.

A curve in S_g has a well-defined length in a given f-metric (by a curve we again mean a homotopy class of curves) [17, Proposition 5.7]. The set of lengths is discrete in $\mathbb{R}$ and so there is a shortest such length, although the curve of shortest length may not be unique.

Stretch factors versus intersection numbers. The following fundamental tool was introduced by Farb, Leininger, and the second author of this paper [15, Proposition 2.6].

Proposition 6.1. *Let $g \geq 1$ and let $f \in \mathrm{Mod}(S_g)$ be pseudo-Anosov. Let c be a curve in S_g that is shortest with respect to some f-metric. If the stretch factor of f is less than or equal to $\sqrt[k]{n/2}$ then $i(c, f^k(c)) < n$.*

In the paper by Farb, Leininger, and the second author, the statement of Proposition 6.1 is less specific, saying only that if the stretch factor of f is less than $n/2$ then there exists a curve c with $i(c, f(c)) < n$. However, the proof there, plus the facts that the stretch factor of f^k is the kth power of the stretch factor for f and that an f-metric is also an f^k-metric, proves the statement given here, and we will need this more general statement. Specifically, we will apply the proposition in the cases $1 \leq k \leq 3$.

Outline of the proof of Theorem 1.2. We now explain how Proposition 6.1 will be used in the proof of Theorem 1.2, and we give an overview of the remaining four sections of the paper.

Let f be a pseudo-Anosov mapping class. For us, the first and most important consequence of Proposition 6.1 comes from the $k = 1$ case: if the stretch factor of f is less than $\sqrt{2}$ then there is a curve c with

$$i(c, f(c)) \leq 2$$

(here we are using the fact that $\sqrt{2} < 3/2$). If c is separating, then we may apply Lemma 2.4 to conclude that f is a normal generator. Thus, we may way assume that c is nonseparating. The proof of Theorem 1.2 then proceeds by considering three cases, as in Table 1. (Of course mod 2 homologous curve cannot have intersection number 1.) For Case 1, our first well-suited curve criteria, Lemmas 2.2 and 2.3, immediately apply. For Case 2, we will need an extension, Lemma 7.1 in Section 7. In that section we also give the full version of the well-suited curve criterion, a necessary and sufficient condition for a mapping class to be a normal generator. This full version, Proposition 7.3, implies Lemmas 2.2, 2.3, and 7.1.

Case 3 is by far the most difficult. Here we should not expect a simple argument to exist, since the equality $[c] = [f(c)] \mod 2$ holds, for example, for every element of the congruence subgroup $\mathrm{Mod}(S_g)[2]$ and every curve c.

Case	$i(c, f(c))$	$[c] = [f(c)]$ mod 2
1	0 or 1	✗
2	2	✗
3	0 or 2	✓

Table 1.

Unlike in Cases 1 and 2, it will not be enough to consider only c and $f(c)$; we will need to consider $f^2(c)$ in addition to c and $f(c)$. Thus, the first step in Case 3 is to observe the following further consequence of Proposition 6.1, applied in the case $k = 2$:

$$i(c, f^2(c)) \leq 2.$$

In Section 8, we will enumerate all possibilities for the configuration of the triple $(c, f(c), f^2(c))$, given that c is nonseparating, that f preserves the mod 2 homology class of c, that $i(c, f(c)) \leq 2$, and that $i(c, f^2(c)) \leq 2$ (actually we enumerate a larger class of triples).

In Section 9 we introduce an additional criterion for a mapping class to be a normal generator, namely, the existence of a good pair of curves. This new criterion, Lemma 9.1, is a generalization of our well-suited curve criterion for separating curves, Lemma 2.4.

We prove Theorem 1.2 in Section 10. In treating Case 3 we will show that many of the configurations from Section 8 admit a good pair of curves. We deal with the configurations that do not admit a good pair of curves on an ad hoc basis. In one subcase we apply the $k = 3$ case of Proposition 6.1 in order to constrain the possible configurations of the quadruple $(c, f(c), f^2(c), f^3(c))$.

7. The well-suited curve criterion: general version

In Section 2 we gave two special cases of the well-suited curve criterion for nonseparating curves, Lemmas 2.2 and 2.3. Using these special cases we were able to prove our main theorem about periodic normal generators and to construct pseudo-Anosov normal generators that answer Long's question and resolve Ivanov's conjecture.

In this section we give a more general version of the well-suited curve criterion for nonseparating curves, Chen's Lemma 7.1 below. Then, as discussed in Section 6, we use this lemma to give an application that will be used in Case 2 of the proof of Theorem 1.2, namely, Lemma 7.2. Specifically this lemma gives a well-suited curve criterion for the situation where $i(c, f(c)) = 2$ and $[c] \neq [f(c)]$ mod 2. At the end of the section we give the general version of the well-suited curve criterion, Proposition 7.3, and explain how to use it to recover Chen's lemma.
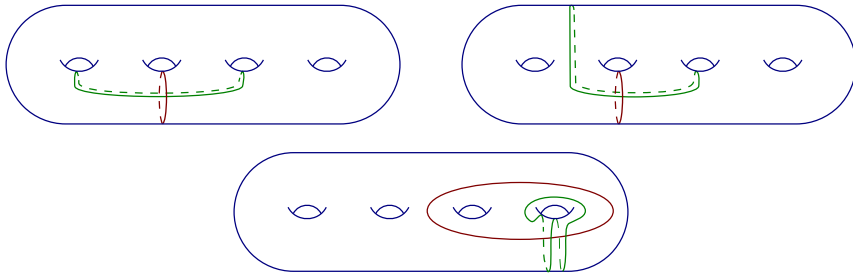
Figure 12. The three possible configurations of curves a and b with $i(a,b) = 2$ and $[a] \neq [b]$ mod 2

Chen's lemma. The next lemma is one generalization of our well-suited curve criteria for nonseparating curves, Lemmas 2.2 and 2.3. In those lemmas, we require that c and $f(c)$ have particularly simple arrangements, and in particular they have intersection number at most 1. In the next lemma the intersection number $i(c, f(c))$ can be arbitrarily large, as long as there is a curve d that forms a (specific) simple arrangement with both.

The formulation of the well-suited curve criterion in Lemma 7.1 was suggested to us by Lei Chen. Our original formulation, Proposition 7.3 below, is more general. The full generality of Proposition 7.3 is not needed for our application, Lemma 7.2 below. Lemma 7.1 is perhaps the simplest version of the well-suited curve criterion that suffices for the application.

Lemma 7.1. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Suppose that there are nonseparating curves c and d in S_g so that $i(c,d) = 1$ and $i(f(c), d) = 0$. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

If we instead have $i(c, d) = 0$ and $i(f(c), d) = 1$ then we may apply Lemma 7.1 to f^{-1} in order to deduce that the normal closure of f^{-1}, hence f, contains the commutator subgroup of $\mathrm{Mod}(S_g)$.

Proof of Lemma 7.1. The commutator $h = [T_c, f] = T_c T_{f(c)}^{-1}$ lies in the normal closure of f. By the hypotheses on intersection number, we have

$$i(d, h(d)) = i(d, T_c T_{f(c)}^{-1}(d)) = i(d, T_c(d)) = 1.$$

By Lemma 2.2, the normal closure of h, and so also the normal closure of f, contains the commutator subgroup of $\mathrm{Mod}(S_g)$. □

For the proof of Theorem 1.2 we will require the following specific application of Lemma 7.1.

Lemma 7.2. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Suppose that there is a nonseparating curve c in S_g so that $i(c, f(c)) = 2$ and so that $[c] \neq [f(c)]$ mod 2. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

Proof. Up to homeomorphism of S_g, there are only three configurations for c and $f(c)$, namely, the configurations shown in Figure 12. In each case there is a curve d with $i(c,d) = 1$ and $i(f(c),d) = 0$. The lemma thus follows from Lemma 7.1.   □

The general well-suited curve criterion. Our next proposition, Proposition 7.3, is the most general statement of the well-suited curve criterion that we will give. It implies Lemma 7.1, which in turn implies Lemmas 2.2 and 2.3. Not only is Proposition 7.3 the most general statement that we give, it is in a sense the most general statement possible, since it gives a necessary and sufficient condition for a mapping class to be a normal generator.

In order to state Proposition 7.3, we require a definition. Let $\mathcal{N}(S_g)$ denote the graph of nonseparating curves for S_g, that is, the subgraph of $\mathcal{C}(S_g)$ spanned by the vertices corresponding to nonseparating curves. Let f be an element of $\mathrm{Mod}(S_g)$. Let $\mathcal{N}_f(S_g)$ denote the abstract graph whose vertex set is the same as that of $\mathcal{N}(S_g)$ and whose edges correspond to pairs of vertices $\{c, h(c)\}$ where h is conjugate to f; we call $\mathcal{N}_f(S_g)$ the *graph of curves* for f.

Proposition 7.3. *Let $g \geq 0$. Then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$ if and only if $\mathcal{N}_f(S_g)$ is connected.*

Proof. First assume that $\mathcal{N}_f(S_g)$ is connected. We would like to show that the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$. Let c and d be nonseparating curves in S_g with $i(c,d) = 1$. By Lemma 2.1 it is enough to show that $T_c T_d^{-1}$ lies in the normal closure of f. As $\mathcal{N}_f(S_g)$ is connected there is a path $c = c_0, \ldots, c_n = d$ between c and d. For each i there is a conjugate f_i of f with $f_i(c_i) = c_{i+1}$. It follows that $T_{c_i} T_{c_{i+1}}^{-1} = [T_{c_i}, f_i]$ lies in the normal closure of f. But then the product

$$\left(T_{c_0} T_{c_1}^{-1}\right)\left(T_{c_1} T_{c_2}^{-1}\right) \cdots \left(T_{c_{n-1}} T_{c_n}^{-1}\right)$$

lies in the normal closure of f. This product is equal to $T_{c_0} T_{c_n}^{-1}$, or $T_c T_d^{-1}$, as desired.

For the other direction, assume that the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$. Let c and d be nonseparating curves in S_g. We would like to show that there is a path from c to d in $\mathcal{N}_f(S_g)$. There is an element h of the commutator subgroup of $\mathrm{Mod}(S_g)$ with $h(c) = d$. Indeed, for $g \geq 3$ this follows from the fact that the commutator subgroup of $\mathrm{Mod}(S_g)$ is equal to $\mathrm{Mod}(S_g)$ and the fact that $\mathrm{Mod}(S_g)$ acts transitively on the vertices of $\mathcal{N}_f(S_g)$. For g equal to 1 or 2 we can first choose an element h_0 of $\mathrm{Mod}(S_g)$ taking c to d, and take h to be any element of the form $h = T_d^k h_0$ that lies in the commutator subgroup.

As the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$ we have that $h = f_n \cdots f_1$, where each f_i is a conjugate of f. Let $c_0 = c$ and for $1 \leq i \leq n$ let

$$c_i = f_i(c_{i-1}) = f_i \cdots f_1(c_0).$$

Note $c_n = h(c_0) = d$. Since each f_i is a conjugate of f and $f_i(c_{i-1}) = c_i$ it follows that each pair of vertices $\{c_{i-1}, c_i\}$ spans an edge in $\mathcal{N}_f(S_g)$. So the sequence of vertices $c_0, \ldots, c_n$ gives a path in $\mathcal{N}_f(S_g)$ from c to d. Since c and d were arbitrary, $\mathcal{N}_f(S_g)$ is connected. $\qquad\square$

While the reverse implication of Proposition 7.3 is the one that is relevant to our applications in this paper, the forward implication also can be applied to prove the connectivity of curve graphs associated to mapping classes. For instance, in Theorem 1.4 we gave examples of normal generators for $\mathrm{Mod}(S_g)$ with arbitrarily large translation lengths on the curve graph $\mathcal{C}(S_g)$. The edges of the corresponding curve graph $\mathcal{N}_f(S_g)$ thus only connect vertices that are very far apart in $\mathcal{C}(S_g)$. Perhaps surprisingly, Proposition 7.3 implies that $\mathcal{N}_f(S_g)$ is connected.

Even for simple f, such as a Dehn twist, it is a challenge to describe $\mathcal{N}_f(S_g)$ explicitly. However, we can often show that $\mathcal{N}_f(S_g)$ is connected without having to give an explicit description. For example, if f is a Dehn twist about a nonseparating curve in S_g, then $\mathcal{N}_f(S_g)$ has edges for every pair of vertices with geometric intersection number 1, and is hence connected. In our proof of Chen's lemma at the end of the section we give an argument that a different $\mathcal{N}_f(S_g)$ is connected, again without needing to analyze all of the edges in the graph.

Chen's lemma via the Putman trick. The well-suited curve criterion of Proposition 7.3 is only as useful as our ability to show that a graph of curves $\mathcal{N}_f(S_g)$ is connected. Putman introduced an effective method for proving the connectedness of such graphs [44, Lemma 2.1]. We give here a specialized version which follows immediately from Putman's lemma and suffices for our purposes.

Lemma 7.4. *Suppose G is a group and X is a graph on which G acts. Suppose G acts transitively on the vertices and that G is generated by the set $\{g_i\}$. If there is an edge e of X with the property that each g_i fixes at least one vertex of e then X is connected.*

We can use this lemma and Proposition 7.3 to prove Lemma 7.1 as follows. Suppose f is an element of $\mathrm{Mod}(S_g)$ and c and d are nonseparating curves in S_g so that $i(c,d) = 1$ and $i(f(c), d) = 0$. We can choose a generating set for $\mathrm{Mod}(S_g)$ consisting of Dehn twists about curves $c_1, \ldots, c_{2g+1}$, where $c_1 = c$, $c_2 = d$, and all other c_i are disjoint from c. Let e be the edge of $\mathcal{N}_f(S_g)$ corresponding to $\{c, f(c)\}$. Since $T_{c_2}(f(c)) = f(c)$ and $T_{c_i}(c) = c$ for all $i \neq 2$, it follows from Lemma 7.4 that $\mathcal{N}_f(S_g)$ is connected, and so Chen's lemma follows from Proposition 7.3.

8. Curve configurations for mod 2 homologous curves

In this section we provide one of the main tools for dealing with Case 3 of the proof of Theorem 1.2, as outlined in Section 6. Specifically, we produce a list of configurations

of curves $(\gamma, \delta, \varepsilon)$. This list contains all triples of curves $(c, f(c), f^2(c))$ in S_g, where $i(c, f(c)) = i(f(c), f^2(c)) \leq 2$ and $i(c, f^2(c)) \leq 2$ and where all three curves are homologous mod 2.

In the remainder of this section (and this section only) we will use the word "curve" to mean a simple closed curve, as opposed to a homotopy class of curves. To emphasize this point, we will use Greek letters to denote curves in this section. For a surface S and a curve α we will write $S \setminus \alpha$ for the surface obtained from S by cutting along α.

The larger set of triples of curves that we will classify is the set of triples of curves $(\gamma, \delta, \varepsilon)$ in any S_g with the following properties:

- all three curves are pairwise homologous mod 2,
- all three curves are pairwise homotopically distinct and in minimal position,
- all three curves have pairwise intersection number at most 2,
- $\gamma \cup \delta$ and $\delta \cup \varepsilon$ have the same number of complementary regions,
- $i(\gamma, \delta) = i(\delta, \varepsilon)$, and
- $|\hat{\imath}|(\gamma, \delta) = |\hat{\imath}|(\delta, \varepsilon)$.

(We give a more concise description of these triples below.) The triples $(c, f(c), f^2(c))$ described at the start of the section satisfy the properties listed here. The set of triples $(\gamma, \delta, \varepsilon)$, however, is a priori a larger, since we do not make the assumption that the pair (γ, δ) is equivalent to the pair (δ, ε) under a homeomorphism of S_g.

The main goal of this section is to show that any triple $(\gamma, \delta, \varepsilon)$ that satisfies this list of properties is (up to stabilization, defined below) one of the triples shown in Figures 14, 17, 20, or 23. This classification is proved as Propositions 8.1, 8.4, 8.6, and 8.7 below.

The strategy for classifying such triples $(\gamma, \delta, \varepsilon)$ is as follows. First, there are only four configurations for the pair (γ, δ) in S_g, up to changing the genus of the complementary regions. The four corresponding "templates" are shown in Figure 13. The goal is then to find all candidate third curves ε, using the hypotheses on the curves. To do this, we first draw on each of the templates the finitely many candidates for ε up to homeomorphism. We then enumerate all the ways of adding handles to each such configuration so that the resulting configuration has the desired properties. Up to changing the genus of the complementary regions (i.e., stabilization), there are finitely many such ways to add the handles.

Four configurations for pairs of mod 2 homologous curves. We first give names to the various types of configurations of nonseparating curves γ and δ with γ and δ in minimal position, with $|\gamma \cap \delta| \leq 2$, with γ and δ homotopically distinct, and with $[\gamma] = [\delta]$ mod 2. There are four types, as in the following table (Table 2).

We emphasize that the type of a pair of curves is independent of the order of the two curves.

Type	$\|\gamma \cap \delta\|$	$\|\hat{\imath}\|(\gamma, \delta)$	$[\gamma] = [\delta]$	$[\gamma] = [\delta] \bmod 2$
I	0	0	✓	✓
II	2	0	✓	✓
III	2	0	✗	✓
IV	2	2	✗	✓

Table 2.

In what follows it will be convenient to depict the different types of configurations by drawing the surface obtained by cutting S_g along the curve δ. The cut surface $S_g \setminus \delta$ is connected and has two boundary components, and on this surface the curve γ either becomes a curve (type I) or a pair of arcs (types II, III, and IV). See Figure 13 for sample configurations of the four types. The picture for type II is drawn in such a way as to suggest an analogy with types III and IV.

In Figure 13 the identification of the two components of the boundary of $S_g \setminus \delta$ is indicated by the pair of black dots. That is, the two boundary components are glued so that the black dots are identified in the closed surface. In the pictures for types II, III, and IV the identification of the two boundary circles can be achieved by a vertical translation followed by a reflection about the plane passing through γ. In what follows we will draw configurations of pairs (γ, δ) and triples $(\gamma, \delta, \varepsilon)$ using similar pictures, and in those pictures we will always use this identification of the boundary components of $S_g \setminus \delta$. There are only two choices for the identification so that γ is a union of curves in the closed surface, and the identification we have specified is the only one that makes γ a connected curve.

Given a configuration of curves on a surface, we may perform a *stabilization* to the configuration, which means we add a handle to one of the configuration's complementary regions. In other words, performing a stabilization means choosing a component of the complement of the configuration, deleting the interiors of two disjoint closed disks that lie in that component, and then identifying the two new boundary components so that the result is an orientable surface. A sequence of such stabilizations is also called a stabilization. If a pair of curves is of type I, II, III, or IV, then the stabilized configuration is of the same type.

More generally, we may *add handles* to a configuration, which means that we delete the interiors of two disjoint disks – possibly in different complementary components – and identify the two new boundary components. Unlike stabilization, the process of adding handles can, for instance, destroy the property that a pair of curves is homologous mod 2.

The sample configurations in Figure 13 are *minimal* in the sense that any pair of curves of a given type can be obtained from the corresponding sample configuration by stabilization.
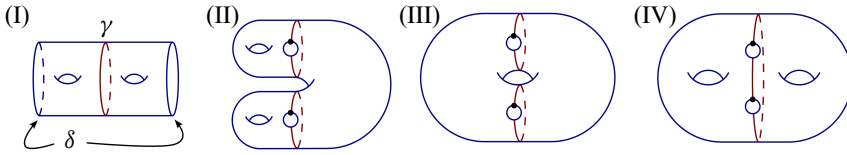
Figure 13. The minimal configurations of nonseparating curves γ and δ with $i(\gamma, \delta) \leq 2$ and $[\gamma] = [\delta]$ mod 2; in each picture the surface is cut along δ.

Configurations of triples of mod 2 homologous curves. Our main goal in this section is to classify certain types of configurations of ordered triples of curves $(\gamma, \delta, \varepsilon)$. Before we embark on this classification we will first restate the goal in terms of the four types of pairs of curves described above.

Suppose that $(\gamma, \delta, \varepsilon)$ is an ordered triple of curves where each pair of curves has type I, II, III, or IV (in particular each pair is in minimal position). We will say that the triple of curves is of type I, II, III, or IV if all three curves are pairwise nonhomotopic and both pairs $\{\gamma, \delta\}$ and $\{\delta, \varepsilon\}$ are of the given type. The motivation here is that in the proof of Theorem 1.2 we will be interested in triples of the form $(c, f(c), f^2(c))$ and in this case the pair $\{c, f(c)\}$ is of the same type as the pair $\{f(c), f^2(c)\}$ (the pair $\{c, f^2(c)\}$ might not be of the same type as the other two pairs).

It will continue to be convenient to draw our configurations on a cut surface $S_g \setminus \delta$. On $S_g \setminus \delta$ each of the curves γ and ε becomes a curve or a pair of arcs, just as in Figure 13 (again, up to stabilization). Moreover, since $\{\gamma, \delta\}$ has the same type as $\{\delta, \varepsilon\}$ (and since the type of a pair is independent of the order of the two curves) the γ-curve/arcs and the ε-curve/arcs on $S_g \setminus \delta$ are of the same type, in that they correspond to the same picture in Figure 13.

We may now restate the main result of this section as follows:

> *Every ordered triple of curves of type I, II, III, or IV is a stabilization of one of the minimal configurations given in Figures 14, 17, 20, or 23, respectively.*

The four different types of triples are addressed in Propositions 8.1, 8.4, 8.6, and 8.7, respectively. Again, these classifications will be used in the proof of Theorem 1.2 in Section 10.

Triples of type I. We can right away classify triples of type I. This case is much simpler than the cases of triples of types II, III, and IV.

Proposition 8.1. *Up to homeomorphism, every ordered triple of curves of type I in S_g is a stabilization of one of the minimal configurations in Figure 14.*

Proof. Let $(\gamma, \delta, \varepsilon)$ be an ordered triple of curves of type I in S_g. Consider the cut surface $S_g \setminus \delta$. The curves γ and ε correspond to curves in $S_g \setminus \delta$ that separate
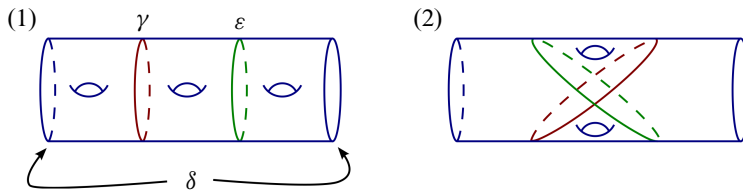
Figure 14. The two minimal configurations of ordered triples of curves of type I, cut along δ.

the two boundary components from each other. Since γ and ε are distinct and since they intersect in at most two points, it follows that the triple $(\gamma, \delta, \varepsilon)$ is given by a stabilization of one of the two configurations in Figure 14. □

Templates. In Propositions 8.4, 8.6, and 8.7 below we will classify ordered triples of curves of type II, III, and IV. We will do this by first listing simpler versions of each type of configuration, called templates.

Before we define templates, we first discuss the process of adding handles to a configuration of curves. Let $\gamma_1, \ldots, \gamma_k$ be a collection of curves in a surface. As above, adding handles to this configuration means deleting the interiors of an even number of disjoint disks in the complement of $\cup \gamma_i$ and gluing the resulting boundary components in pairs. The order in which handles are added does not affect the final configuration. In our arguments, we will only add handles to a triple of curves in such a way that the resulting configuration is a triple of the same type.

Let $(\gamma, \delta, \varepsilon)$ be an ordered triple of curves of type II, III, or IV in S_g. Set $S = S_1$ if the type is II or IV and set $S = S_2$ if the type is III (where S_1 and S_2 denote the closed surfaces of genus 1 and 2, respectively). A *template* for $(\gamma, \delta, \varepsilon)$ is an ordered triple of curves $(\gamma_0, \delta_0, \varepsilon_0)$ in S so that the following conditions hold:

(1) γ_0 and δ_0 are configured in S as in Figure 15, and

(2) the triple $(\gamma, \delta, \varepsilon)$ is obtained from $(\gamma_0, \delta_0, \varepsilon_0)$ by adding handles.

It follows from the second condition that the pairwise intersection numbers and algebraic intersection numbers of γ_0, δ_0, and ε_0 are inherited from those of γ, δ, and ε.

As one example, the configuration shown on the right-hand side of Figure 16 is a template in S_2 for the triple of curves of type III in S_9 shown on the left-hand side of the figure.

We note that the triples of curves in a template might fail to be a triple of type I, II, III, or IV in several ways: they might not be in minimal position, they might not be homotopically distinct, and the pair $\{\delta_0, \varepsilon_0\}$ might not have the same type as $\{\delta, \varepsilon\}$. In the example of Figure 16, the triple of curves in the template fails in all three ways.

Lemma 8.2. *Every ordered triple of curves of type II, III, or IV in S_g has a template in S_1, S_2, or S_1, respectively.*
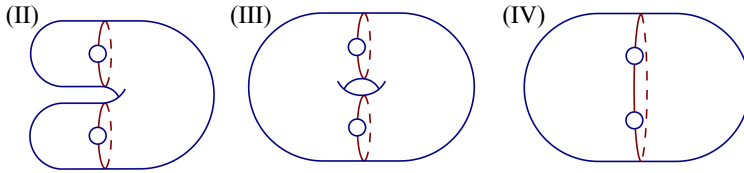
Figure 15. The configurations $\{\gamma_0, \delta_0\}$ in the templates of type II, III, and IV; the surfaces are cut along δ_0 and the resulting boundary components are identified as in Figure 13.

Proof. For concreteness we state the proof in the case of a triple of type III; the proofs for triples of type II and IV are similar. Let $(\gamma, \delta, \varepsilon)$ be an ordered triple of curves in S_g of type III. The configuration $\{\gamma, \delta\}$ in S_g is a stabilization of the standard pair of type III shown in Figure 13. The curves γ and δ divide S_g into two components, which we will refer to as R_1 and R_2. Each is a surface with two boundary components. Since $\{\delta, \varepsilon\}$ is of type III, the curve ε gives rise to collections of arcs in R_1 and R_2. Since $|\gamma \cap \varepsilon|$ and $|\delta \cap \varepsilon|$ are bounded above by 2, the intersection of ε with each R_i is at most two arcs.

Fix $i \in \{1, 2\}$. If there are two arcs of ε in R_i and both arcs have both their endpoints on the same component of the boundary of R_i, then these endpoints are unlinked on the boundary, in that they do not alternate. Indeed, if there were two arcs of ε in R_i that were linked on the boundary then we could find a curve in S_g that is disjoint from γ and δ, intersects one of these arcs of ε in one point, and is disjoint from ε otherwise. This contradicts the assumption that ε is homologous to γ and δ mod 2.

Consider a regular neighborhood of the union of the arcs of ε in R_1 with the boundary of R_1. By the previous paragraph, this neighborhood has genus 0. It follows that there is a collection of curves in R_1 so that when we cut R_1 along these curves we obtain a connected surface of genus 0. By the same argument there is a similar collection of curves in R_2. Cutting S_g along all of these curves and gluing disks to the resulting boundary components, we obtain a surface of genus 2. The curves arising from γ and δ are configured in S_2 like the curves γ_0 and δ_0 in Figure 15. By construction the configuration of curves in S_2 is a template for the original triple $(\gamma, \delta, \varepsilon)$ in S_g. $\qquad\square$

In the process of classifying triples of curves of types II, III, and IV, we will start in each case by classifying all templates of that type. The next lemma gives a condition that a configuration of curves must satisfy in order to be a template.

Lemma 8.3. *Suppose γ_0, δ_0, and ε_0 are curves giving a template for an ordered triple of curves $(\gamma, \delta, \varepsilon)$ in S_g of type II, III, or IV. Then ε_0 is a nonseparating curve in the underlying surface of the template.*

Proof. Suppose for contradiction that ε_0 is a separating curve in the template.
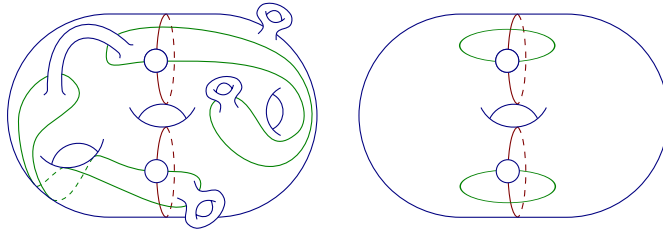
Figure 16. An example of an ordered triple of curves of type III (left) and its underlying template (right).

Because ε is nonseparating in S_g it must be that the ordered triple in S_g is obtained from the template by adding at least one handle to the template that connects the two sides of ε_0. The three curves of the template cut the underlying surface into regions. Say that the added handle connects regions R_1 and R_2. Since δ_0 is nonseparating, there exists a path α from R_1 to R_2 that avoids δ_0. The concatenation of α with a path along the added handle thus gives rise to a curve in S_g that intersects ε_0 in a single point and is disjoint from δ_0. This contradicts the assumption that ε and δ are homologous mod 2. $\square$

Linking. Let $(\gamma, \delta, \varepsilon)$ be an ordered triple of curves in S_g of type II, III, or IV. Since $|\gamma \cap \delta| = |\delta \cap \varepsilon| = 2$, each of $\gamma \cap \delta$ and $\varepsilon \cap \delta$ is a pair of points on δ. Because there are no triple intersections, these four points are distinct and so the two pairs can be either linked or unlinked. As such we may say that γ and ε are *linked* or *unlinked* along δ. In our classification of triples of type II, III, and IV we will use this notion to distinguish between various cases.

Triples of type II. Having established the preliminaries about templates, we now begin the process of classifying ordered triples of types II, III, and IV. We first classify ordered triples of type II.

Proposition 8.4. *Up to homeomorphism, every ordered triple of curves of type II in S_g is a stabilization of one of the minimal configurations in Figure 17.*

Proof. The proof has two main phases:

(1) classify all possible templates in S_1 for a triple of type II, and

(2) build all minimal configurations of type II by adding handles to the templates.

It follows from Lemma 8.2 that this two-phase process recovers all triples of type II.

Phase 1. Suppose that $(\gamma_0, \delta_0, \varepsilon_0)$ is a template in S_1 for a triple of curves $(\gamma, \delta, \varepsilon)$ in S_g. We will show that the template $(\gamma_0, \delta_0, \varepsilon_0)$ in S_1 is, up to homeomorphism, one of the four templates shown in Figure 19.
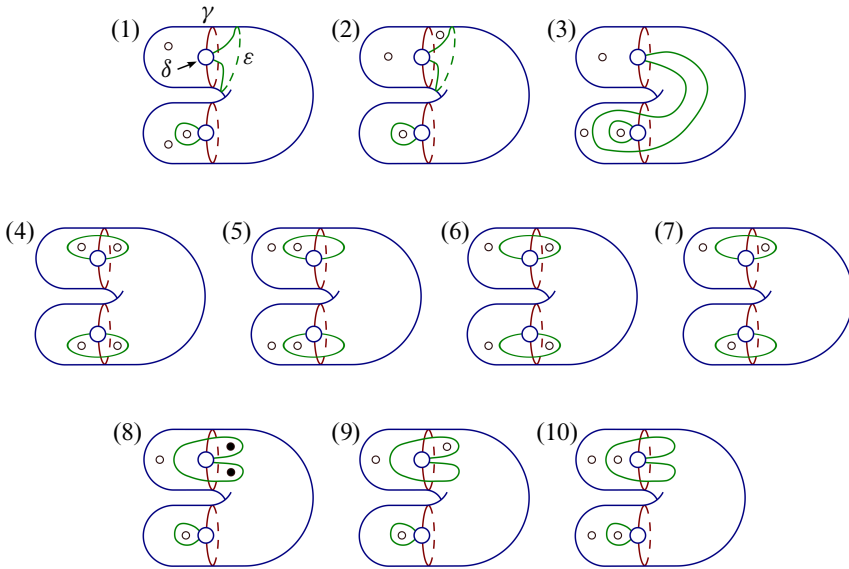
Figure 17. The ten minimal configurations of triples of type II; a white dot indicates a handle attached within a single region and a pair of black dots indicates a handle connecting two distinct regions.

Consider the cut surface $S_1 \smallsetminus \delta_0$. We call the two resulting boundary components δ_1 and δ_2. On the cut surface, γ_0 and ε_0 become pairs of arcs $\{\gamma_1, \gamma_2\}$ and $\{\varepsilon_1, \varepsilon_2\}$. The arcs γ_1 and γ_2 are as shown in configuration II in Figure 15. What remains is to determine the ways in which ε_1 and ε_2 may lie in the cut surface.

Since $(\gamma, \delta, \varepsilon)$ is of type II, each of the γ_i-arcs and ε_i-arcs connects some δ_i to itself. The labels of the γ_i-arcs and ε_i-arcs are inherited from the labels of the δ_i on which they have their endpoints.

The arcs γ_1 and γ_2 are both separating and they divide the cut surface $S_1 \smallsetminus \delta_0$ into one annulus R_0 and two disks, R_1 and R_2. We choose the names R_1 and R_2 according to which of δ_1 and δ_2 it intersects.

We may record the intersection numbers between the arcs γ_i and ε_i with a 2×2 matrix. We will arrange these matrices so that the rows correspond to the γ_i and the columns correspond to the ε_i. Since $|\gamma_0 \cap \varepsilon_0|$ is 0 or 2, the total sum of the entries in the matrix must be 0 or 2. Further, since each γ_i is separating, each off-diagonal entry of the matrix must be even.

Up to renumbering the two arcs in each pair, there are four such matrices:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}.$$

From these matrices, we can draw corresponding partial templates as in Figure 18, each unique up to homeomorphism of S_1. To do this, we must first observe that each
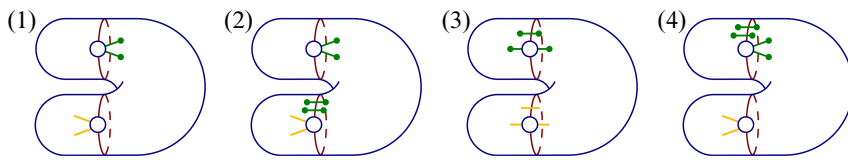
Figure 18. The partial templates for configuration II; the partial arcs with dots belong to ε_1 and the other arcs to ε_2.

of the four matrices determines whether or not γ_0 and ε_0 are linked along δ_0: if the sum of the entries in each column is odd then they are linked and otherwise they are unlinked. This is because the parity of the sum of the entries in the ith column is the number of times ε_i crosses γ_0. We treat the partial templates in turn.

Partial template 1. Since R_2 is a disk, there is a unique choice for the arc ε_2 up to homeomorphism. There are two choices for the arc ε_1 in R_0. One choice of ε_1 gives rise to a separating curve ε_0, which violates Lemma 8.3. The other choice of ε_1 gives template 1 in Figure 19.

Partial template 2. There is a unique choice for ε_2 and a unique choice for the intersection of ε_1 with R_2. Up to homeomorphism, there are two choices for the intersection of ε_1 with R_0, determined by the induced pairing of the two points of $\varepsilon_1 \cap \delta_1$ with the two points of $\varepsilon_1 \cap \gamma_2$. One of the possibilities leads to a separating curve ε_0, violating Lemma 8.3. The other leads to template 2 in Figure 19.

Partial template 3. The arcs ε_1 and ε_2 must bound disks with δ_1 and δ_2, respectively. The boundaries of these two disks specify two arcs of δ_1 and δ_2 (the arcs of δ_1 and δ_2 contained in the disks). Up to homeomorphism, there are then two possibilities: these two arcs are either equal in (the uncut) S_1 or not. If they are equal, then the resulting curve ε_0 is separating, violating Lemma 8.3. The other possibility leads to template 3 in Figure 19.

Partial template 4. There is a unique choice of ε_2 and a unique choice for the intersection of ε_1 with R_1. Up to homeomorphism, there are two possibilities for the intersection of ε_1 with R_0: the two sub-arcs are either nested or not. When they are nested the resulting curve ε is separating. When they are not nested, we obtain template 4 in Figure 19.

Since all of the templates found are shown in Figure 19, this completes the first phase of the proof.

Phase 2. As mentioned above, Lemma 8.2 gives that every minimal configuration of type II on a surface S_g arises from one of the templates by adding handles. Before addressing each template, we establish three constraints on the process of adding handles in order to obtain a minimal configuration of type II: the minimality condition, the checkerboard condition, and the type II condition.
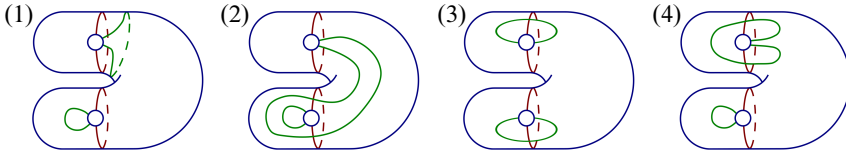
Figure 19. The four templates for triples of type II.

We can record the set of handles required to pass from a given template to a given type II configuration in S_g with a multiset (i.e., unordered list) of unordered pairs $\{R_i, R_j\}$ of not-necessarily-distinct regions in the complement of $\gamma_0 \cup \delta_0 \cup \varepsilon_0$.

There is a partial order on configurations of type II where one configuration is less than another if the latter is obtained from the former by stabilization. The desired configurations are the ones that are minimal in this partial order.

For a template and multiset of handles of pairs of regions as above to specify a minimal configuration, it must satisfy the following condition:

Minimality condition. Each pair $\{R_i, R_j\}$ appears at most once, and the pair $\{R_i, R_i\}$ may appear only if R_i is contained in a bigon or an annulus bounded by two of the curves.

The first part of the minimality condition forces our multiset to be a set; we will refer to the multiset as the handle set in what follows. Since the handle set must be finite (there are finitely many pairs of regions), we have thus reduced our problem of finding all minimal configurations to a finite check for each template.

There are certain pairs of regions that may not appear in the handle set. The curves γ_0, δ_0 and ε_0 are all homologous mod 2 in the templates. We may not add any handles destroying this property. For each pair of curves we may color the complementary regions by two colors so that two regions adjacent along an arc of one of the curves have different colors. The resulting restriction on the handle set is as follows.

Checkerboard condition. The handle set may not include any pair of regions that have different colors in any of the three colorings.

Alternatively, any path connecting a pair of regions in the handle set must cross all three curves γ_0, δ_0, and ε_0 the same number of times mod 2. The checkerboard condition eliminates many possibilities in the finite check.

One requirement for $(\gamma, \delta, \varepsilon)$ to be a triple of type II is that each of $\gamma \cup \delta$ and $\delta \cup \varepsilon$ must cut S_g into exactly three regions. It is already the case in each the templates that $\gamma_0 \cup \delta_0$ and $\delta_0 \cup \varepsilon_0$ cut S_1 into exactly three regions. We thus obtain one additional condition, as follows.

Type II condition. The handle set may not include any pair of regions (determined by the triple) that lie in two distinct regions determined by either of the pairs $\{\gamma_0, \delta_0\}$ or $\{\delta_0, \varepsilon_0\}$.

We are now ready to determine all minimal configurations of type II obtained from the four templates. The task for each template is to find all handle sets that respect the three constraints above and result in curves lying in minimal position. In particular, we must add handles to eliminate bigons and annuli formed by any pair of curves in the triple.

Templates 1, 2, and 3. In these cases all pairs of distinct regions fail either the checkerboard or the type II condition. Therefore, in adding handles to these templates we may only use stabilizations, subject to the minimality condition. It thus remains to determine which subsets of the set of regions give rise to configurations of curves that are in minimal position and are minimal in the partial order. The resulting minimal configurations are configurations 1–7 in Figure 17.

Template 4. In this case there is exactly one distinct pair of regions that satisfies both the checkerboard and type II conditions, namely, the two regions marked by black dots in configuration 8 in Figure 17. This pair of regions may be in the handle set or not. After making this choice, it remains to determine which stabilizations satisfy the minimality condition and result in configurations of curves that are in minimal position. The resulting minimal configurations are configurations 8–10 of Figure 17.

All of the minimal configurations found are shown in Figure 17 and so the proposition is proven. $\square$

Triples of type III. Our next goal is to classify ordered triples of curves of type III. We require one further lemma about templates of type III.

Lemma 8.5. *Suppose γ_0, δ_0, and ε_0 are curves in S_2 giving a template for an ordered triple of curves $(\gamma, \delta, \varepsilon)$ in S_g of type III. Then the two arcs of ε_0 in the cut surface $S_2 \setminus \delta_0$ are both separating or both nonseparating.*

Proof. Consider the cut surface $S_2 \setminus \delta_0$, and suppose that one of the ε_0-arcs in $S_2 \setminus \delta_0$ is separating and one is nonseparating. In S_g the configuration $\{\delta, \varepsilon\}$ is a stabilization of the configuration III in Figure 13. This means that after adding handles to the template on S_2, the two arcs of ε must be nonseparating in the cut surface $S_g \setminus \delta$, and their union must separate S_g.

The only way to make the separating ε_0-arc in S_2 into a nonseparating arc in $S_g \setminus \delta$ would be for one of the added handles to connect its complementary regions. In the surface with the added handles the union of the two ε-arcs is then nonseparating in $S_g \setminus \delta$. Thus $(\gamma_0, \delta_0, \varepsilon_0)$ is not a template for $(\gamma, \delta, \varepsilon)$. $\square$

Proposition 8.6. *Up to homeomorphism, every ordered triple of curves of type III in S_g is a stabilization of one of the minimal configurations in Figure 20.*

Proof. We complete the proof in the same two phases as in the proof of Proposition 8.4, first listing all possible templates in S_2 for a triple of type III and then building all minimal configurations of type III from these templates.
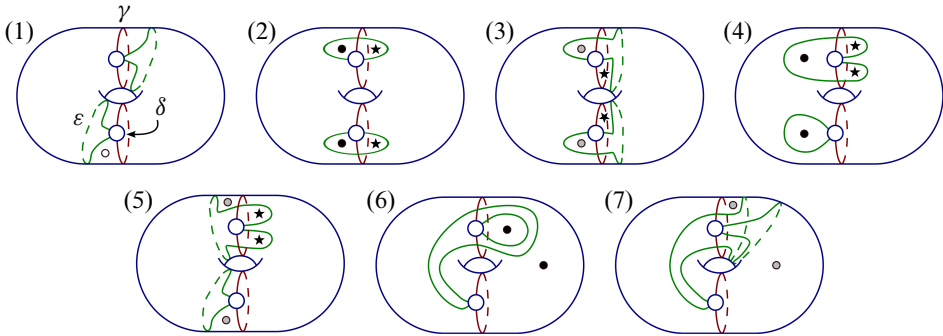
Figure 20. The sixteen minimal configurations of triples of type III; a white dot indicates a handle attached within a single region, a pair of black dots indicates a handle connecting two distinct regions, a pair of gray dots indicates two possibilities, namely, one handle attached within each of the two regions or a single handle connecting the two regions, and a pair of stars indicates two possibilities, namely, either no handle is added or a handle is added connecting the two regions.

Phase 1. Suppose that $(\gamma_0, \delta_0, \varepsilon_0)$ is a template in S_2 for a triple of curves $(\gamma, \delta, \varepsilon)$ in S_g. We will show that the template $(\gamma_0, \delta_0, \varepsilon_0)$ in S_2 is, up to homeomorphism, one of the seven templates shown in Figure 22.

The curves γ_0 and δ_0 are configured in S_2 as in the type III configuration in Figure 15. Consider the cut surface $S_2 \setminus \delta_0$, as in the figure. We call the two resulting boundary components δ_1 and δ_2.

In $S_2 \setminus \delta_0$ both γ_0 and ε_0 become pair of arcs $\{\gamma_1, \gamma_2\}$ and $\{\varepsilon_1, \varepsilon_2\}$. Each connects some δ_i to itself. The labels of the γ_i-arcs and the ε_i-arcs are inherited from the labels of the δ_i on which they have their endpoints. The arcs γ_1 and γ_2 divide $S_2 \setminus \delta_0$ into two annuli, R_1 and R_2.

We may record the intersection numbers between the γ_i-arcs and ε_i-arcs with a 2×2 matrix, where the rows and columns correspond to the γ_i and the ε_i, respectively. The total sum of the entries in the matrix must be 0 or 2. Up to renumbering the arcs in each pair, there are seven such matrices:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}.$$

The partial templates corresponding to these matrices are shown in Figure 21. As in the proof of Proposition 8.4, γ_0 and ε_0 are linked or unlinked along δ_0 according to the parities of the column sums. In the pictures, R_1 and R_2 are the left- and right-hand sides of each partial template, respectively. We treat the partial templates in turn.

Partial template 1. Given two points on a single boundary component of an annulus, there are two choices of arcs in the annulus connecting those points up to homeo-
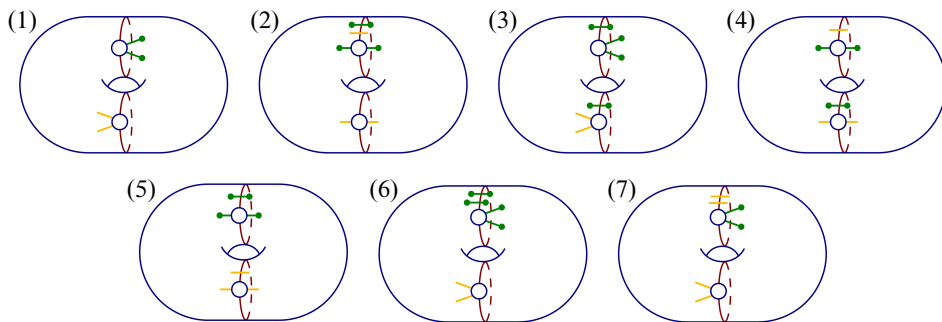
Figure 21. The partial templates for configuration III; the partial arcs with a dot belong to ε_1 and the other arcs to ε_2.

morphism fixing the boundary. Therefore, in each annulus R_i there are two choices of arcs that complete the partial template. This gives four possibilities.

In each R_i one of the choices of arcs is separating in the cut surface $S_2 \setminus \delta_0$ and one is nonseparating. By Lemma 8.5 we must either choose the two separating arcs or the two nonseparating arcs to form a template. If we choose the two separating arcs, then ε_0 is separating, violating Lemma 8.3. Thus, the only possibility is where ε_1 and ε_2 are nonseparating. This choice gives template 1 in Figure 22.

Partial templates 2 and 3. Neither of these partial templates can be completed to be a template. To see this we consider any of the four curves obtained as the boundary of a neighborhood of $\gamma_0 \cup \delta_0$. The algebraic intersection number of any of these curves with ε_0 is odd. On the other hand, the algebraic intersection number of any of these curves with γ_0 and δ_0 is 0. This contradicts the assumption that ε_0 is homologous to γ_0 and δ_0 mod 2.

Partial template 4. In this case there is a unique choice of ε_0 up to homeomorphism, and it is a separating curve. By Lemma 8.3 there is no template arising from this partial template.

Partial template 5. In order to complete ε_1 from this partial template, we need to choose two arcs, one in each annulus R_1 and R_2. This gives four possibilities. Up to homeomorphism there are only two possibilities for the resulting ε_1, one that is separating in the cut surface $S_2 \setminus \delta_0$ and one that is nonseparating. There are similarly four choices for ε_2, two separating and two nonseparating in $S_2 \setminus \delta_0$. By Lemma 8.5 we are left with four choices for the pair $\{\varepsilon_1, \varepsilon_2\}$. Two of these choices result in a separating ε_0, which violates Lemma 8.3. The other two choices are templates 2 and 3 in Figure 22.

Partial template 6. The intersection of ε_1 with R_2 is a pair of arcs. Up to homeomorphism, there are two possibilities: they can be nested or not nested. There are

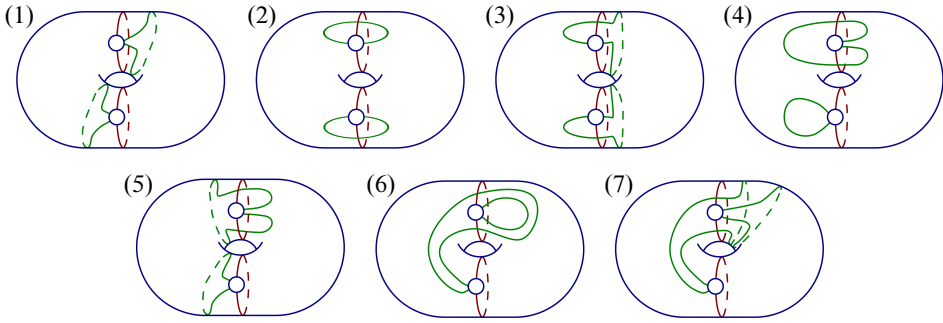Figure 22. The seven templates for triples of type III.

then two choices for the intersection of ε_1 with R_1. This gives four total choices for ε_1. There are two choices for ε_2, one separating and one not, but by Lemma 8.5 the choice of ε_2 is completely determined by the choice of ε_1. Thus there are four total candidates for ε_0. The two candidates corresponding to the nested arcs in R_2 are separating, and by Lemma 8.3 these are ruled out. The other two are templates 4 and 5 in Figure 22.

Partial template 7. Up to homeomorphism there are two choices for the intersection of ε_2 with R_1, corresponding to the two ways to pair the two points of $\varepsilon_2 \cap \delta_2$ with the two points of $\varepsilon_2 \cap \gamma_1$. As usual there are two choices for the intersection of ε_2 with R_2. This gives four choices for ε_2. For one of the four choices there is no choice of ε_1 that is disjoint from ε_2. For each of the other three choices for ε_2, the choice of ε_1 is determined by Lemma 8.5. One of the resulting candidates for ε_0 is a separating curve, violating Lemma 8.3. The other two are templates 6 and 7 in Figure 22.

Since all of the templates found are shown in Figure 22, this completes the first phase of the proof.

Phase 2. By Lemma 8.2 we may obtain the minimal configurations of triples of type III by adding handles to the seven templates. As in the proof of Proposition 8.4 the corresponding handle set is subject to the minimality and checkerboard conditions. In place of the type II condition, we will give an analogous type II condition, as follows.

Two curves $\{\delta, \varepsilon\}$ in S_g that intersect in two points and have algebraic intersection number 0 form a pair of type III if and only if $\delta \cup \varepsilon$ has two complementary regions, each with two boundary components. In some of the templates $\delta_0 \cup \varepsilon_0$ has three complementary regions. In these cases, we must add a handle connecting the two nonadjacent regions complementary to $\{\delta, \varepsilon\}$, so that the resulting $\{\delta, \varepsilon\}$ is of type III.

Type III condition. If $\delta_0 \cup \varepsilon_0$ has three complementary regions in S_2, the handle set must include a pair of regions (for the triple) that are contained in nonadjacent regions of S_2 determined by the pair $\{\delta_0, \varepsilon_0\}$.

For each of the seven templates we will follow three steps to produce all possible minimal configurations of triples of type III. Each step may produce multiple configurations.

Step 1. Check the number of components of the complement of $\delta_0 \cup \varepsilon_0$; if there are three components, add a single pair to the handle set as in the type III condition while respecting the checkerboard condition.

Step 2. Check in each of the resulting configurations if the ε_0-curve and the γ_0-curve are homotopically distinct in minimal position; if not add pairs to the handle set as above, respecting the checkerboard and minimality conditions.

Step 3. Check in each of the resulting configurations if there are any other pairs of regions that satisfy the checkerboard condition and the minimality condition. If so we may add these handles to create a new minimal configuration.

This first step is required in order to obtain a triple of type III. Note that this step is never a stabilization. The second step might be a stabilization, but it is also required in order to obtain a triple of type III. The third step is also never a stabilization, but it is required in order to obtain all minimal configurations. In practice, at most one handle is added in the third step.

We are now ready to determine all minimal configurations of type III obtained from the seven templates.

Template 1. In this case the complement of $\delta_0 \cup \varepsilon_0$ already has exactly two components, and so the type III condition is satisfied. However, there is an annulus bounded by γ_0 and ε_0. Every handle connecting distinct regions fails the checkerboard condition. Therefore, the only possibility is to add handles within regions.

The annulus between γ_0 and ε_0 is divided into two squares by δ_0. We need to add a handle to one of the two squares, but the two choices give homeomorphic configurations. The result is configuration 1 in Figure 20.

Templates 2, 4, and 6. In these cases the complement of $\delta_0 \cup \varepsilon_0$ has three components. Up to homeomorphism, there is a unique pair of regions that satisfies the checkerboard condition and the type III condition. These pairs of regions are marked by black dots in configurations 2, 4, and 6 in Figure 20. Adding each of these handles results in a minimal configuration. For templates 2 and 4, it is possible to further add a single handle between the two regions as in Step 3. The resulting minimal configurations are given by pictures 2, 4, and 6 in Figure 20. Each of the pictures 2 and 4 really represents two minimal configurations; the stars indicate the additional handle that can be added in Step 3.

Templates 3, 5, and 7. In these cases the complement of $\delta_0 \cup \varepsilon_0$ already has exactly two components, satisfying the type III condition. However the δ_0 and ε_0 curves are not in minimal position; in fact they form two bigons. Up to homeomorphism, there are two handle sets that we may use as in Step 2: in each case we may attach a
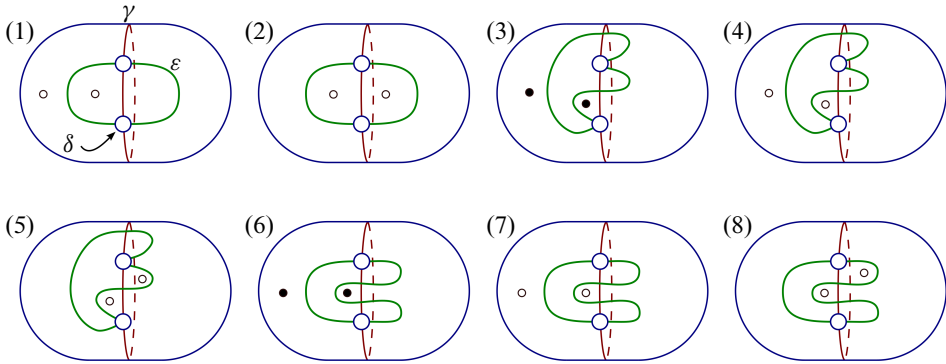
Figure 23. The eight minimal configurations of triples of type IV; a white dot indicates a handle attached within a single region and a pair of black dots indicates a handle connecting two distinct regions

handle that connects the pair of regions indicated by gray dots in pictures 3, 5, and 7 of Figure 20, or we may pair each region marked by a gray dot with itself. As such, we obtain two minimal configurations from each template. As in Step 3, the two minimal configurations produced from templates 3 and 5 so far allow one additional handle that still results in a minimal configuration. The corresponding regions are marked by stars in the figure. Thus, pictures 3 and 5 represent four different minimal configurations each.

Since all of the minimal configurations we have constructed are shown in Figure 20, this completes the proof of the proposition.   □

Triples of type IV. Finally, we classify ordered triples of type IV.

Proposition 8.7. *Up to homeomorphism, every ordered triple of curves of type IV in S_g is a stabilization of one of the minimal configurations in Figure 23.*

Proof. We complete the proof in the same two phases used in the proofs of Propositions 8.4 and 8.6, first listing all possible templates in S_1 for a triple of type IV, and then building all minimal configurations of type IV from these templates.

Phase 1. Suppose that $(\gamma_0, \delta_0, \varepsilon_0)$ is a template in S_1 for a triple of curves $(\gamma, \delta, \varepsilon)$ in S_g. We will show that up to homeomorphism this template is one of the three templates shown in Figure 25.

The curves γ_0 and δ_0 are configured in S_1 as in Figure 15. Consider the cut surface $S_1 \setminus \delta_0$ and call the resulting boundary components δ_1 and δ_2. In the cut surface, γ_0 and ε_0 each become a pair of arcs. By recording the pairwise intersections, we obtain a 2×2 matrix where the sum of the entries is 0 or 2. Up to symmetry there are five matrices:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}.$$
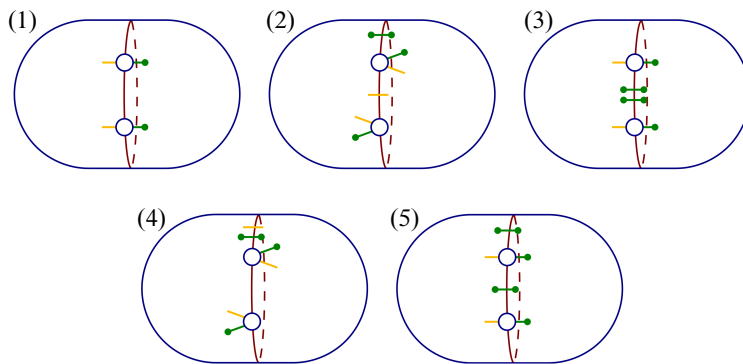
Figure 24. The partial templates for configuration IV; the partial arcs with a dot belong to ε_1 and the other arcs belong to ε_2.

Each matrix again determines whether γ_0 and ε_0 are linked or unlinked along δ_0. The corresponding partial templates are shown in Figure 24. To complete the first phase we must find all ways of completing the partial templates to templates. We treat the five cases in turn. Our work is simplified by the fact that the γ_0-arcs divide the $S_1 \setminus \delta_0$ into two disks R_1 and R_2 and the fact that there is a unique arc connecting two points in the boundary of a disk, up to homeomorphism.

Partial templates 1, 2, and 3. In each of these cases there is a unique way to complete the partial templates. The resulting templates are shown in Figure 25.

Partial templates 4 and 5. In these cases there is no way to complete the partial templates to templates. This is because the partial templates indicate that the two ε-arcs in each R_1 should be linked along the boundaries of the R_1, but this is impossible since R_1 is a disk and the arcs must be disjoint.

Since all of the templates found are shown in Figure 25, this completes the first phase of the proof.

Phase 2. By Lemma 8.2 we may obtain the minimal configurations of triples of type IV by adding handles to the three templates.

We follow the same steps as in the proof of Proposition 8.6 (the classification of type III triples) in order to build the minimal configurations of type IV from the templates. As in the proofs of Propositions 8.4 and 8.6, the added handles must satisfy the minimality condition and the checkerboard condition. Also, we must add handles so that in the resulting configuration the three curves are pairwise in minimal position and homotopically distinct.

One simplifying feature in the present situation is that the pair $\{\delta_0, \varepsilon_0\}$ in the three templates is automatically of type IV. Indeed, any two curves that intersect in two points and have nonzero algebraic intersection are in minimal position. Therefore we may skip Step 1 from the proof of Proposition 8.6 in each of the three cases. In other
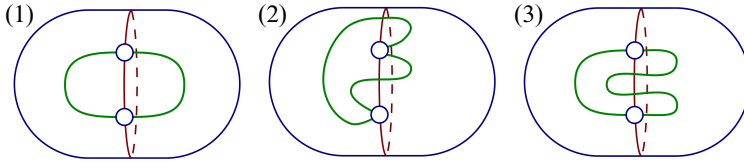
Figure 25. The three templates for triples of type IV.

words, there is no need for a type IV condition, analogous to the type II condition and the type II conditions used in the proof of Propositions 8.4 and 8.6.

Template 1. In this case γ_0 and ε_0 are disjoint and homotopic, so they bound two annuli in S_1. Each annulus is cut into two squares by δ_0. Up to homeomorphism, there are two ways to add handles to the template that satisfy the minimality and checkerboard conditions and that result in homotopically distinct γ_0- and ε_0-curves. The two resulting configurations are configurations 1 and 2 in Figure 23.

Template 2. In this case γ_0 and ε_0 are homotopic in S_1 and they intersect in two points. As such, they bound two bigons. Up to homeomorphism, there are three ways to add handles in order to obtain a minimal configuration, resulting in configurations 3, 4, and 5 in Figure 23.

Template 3. Again γ_0 and ε_0 are homotopic in S_1 and intersect in two points and so they bound two bigons. Up to homeomorphism, there are three ways to add handles to obtain a minimal configuration, resulting in configurations 6, 7, and 8 in Figure 23.

Since all of the minimal configurations we have constructed are shown in Figure 23, this completes the proof of the proposition. $\square$

9. A well-suited curve criterion for separating subsurfaces

In this section we give our main tool for addressing Case 3 of the proof of Theorem 1.2, as outlined in Section 6. One approach to addressing Case 3 would be to show that the curve graph $\mathcal{N}_f(S_g)$ from Section 7 is connected and to apply Proposition 7.3. However, the only edges of $\mathcal{N}_f(S_g)$ that we have in hand are the ones in the orbit of the edge $\{c, f(c)\}$. Since $[c] = [f(c)]$ mod 2 there is no path of such edges connecting c to any other curve in a different mod 2 homology class.

Our solution is to develop an analogue of our well-suited curve criterion for separating curves, Lemma 2.4. In that lemma we also have a curve d whose homology class (the trivial one) is preserved. However, we are able to use the fact that $i(d, f(d)) \leq 2$ in order to find a different curve a that satisfies one of our well-suited curve criteria for nonseparating curves, Lemma 2.3. We will follow a similar approach here.

Our first goal is to state our generalization of Lemma 2.4, namely, Lemma 9.1. Then we will state two specific consequences, Lemmas 9.3 and 9.4, that will be applied directly in the proof of Theorem 1.2.

The good pair lemma. Suppose that D is a subsurface of S_g, not necessarily connected. Let a be a curve in S_g, and say that a lies in the region R complementary to D. We say that the homology class $[a] \in H_1(S_g; \mathbb{R})$ is *local* with respect to D if it is not represented by any curve lying in a complementary region to D that is distinct from R. When there is no confusion, we will refer to a homology class as simply being "local," the subsurface D being understood.

Lemma 9.1. *Let $g \geq 0$ and let $f \in \mathrm{Mod}(S_g)$. Let D be a separating subsurface of S_g. Suppose that a and b are nonseparating curves in S_g whose homology classes are local with respect to D. If a and b lie in distinct complementary regions of D and lie in the same complementary region of $f(D)$ then the normal closure of f contains the commutator subgroup of $\mathrm{Mod}(S_g)$.*

When there exists a subsurface D and curves a and b as in the statement of Lemma 9.1 we say that a and b form a *good pair of curves* for the mapping class f (the subsurface D being understood).

Lemma 2.4 is in fact a special case of Lemma 9.1: on the subsurface D there is an annular neighborhood of the separating curve d and it is easy to find the appropriate curves a and b in this case. Our proof of Lemma 9.1 is a straightforward generalization of our proof of Lemma 2.4.

There is an asymmetry in the statement of Lemma 9.1. If instead a and b lie in the same complementary region of D but in different complementary regions of $f(D)$ then we may apply the lemma to f^{-1} instead of f; and of course f is a normal generator if and only if f^{-1} is. Because of this, we will also refer to such an a and b as a good pair for f.

Proof of Lemma 9.1. By hypothesis, a and b lie in distinct complementary regions of D. Therefore, the curves $f(a)$ and $f(b)$ lie in distinct complementary regions of $f(D)$. Also by hypothesis, a and b lie in the same complementary region of $f(D)$. Combining the last two sentences, it must be that either a and $f(a)$ lie in distinct complementary regions of $f(D)$ or b and $f(b)$ lie in distinct complementary regions of $f(D)$ (or both statements hold). Without loss of generality, suppose the former holds. Since $[a]$ is local with respect to D, it follows that $[f(a)]$ is local with respect to $f(D)$, and in particular that $[a] \neq [f(a)]$. Clearly, $i(a, f(a)) = 0$. An application of Lemma 2.3 completes the proof. $\square$

The next lemma gives two ways of finding curves whose homology classes are local in the sense of Lemma 9.1. In fact, such curves are easy to find.

Lemma 9.2. *Let $g \geq 0$, and let D be a separating subsurface of S_g.*

(1) *If D is connected, then every nonseparating curve in S_g in the complement of D represents a homology class that is local.*

(2) *Every curve in the complement of D that does not separate the complement of D represents a homology class that is local.*

Proof. Let a be a nonseparating curve in S_g that lies in the complementary region R of D. For both statements it suffices to find a curve b so that $i(a, b) = 1$ and so that b is contained in the smallest subsurface containing $D \cup R$. If a is nonseparating in R, then the curve b can be taken to lie in R. The second statement follows. If a is separating in R, then since a is nonseparating in S_g it must be that a induces a nontrivial partition of the components of the boundary of R. It then follows that we can find the desired curve b in $D \cup R$. The first statement follows. $\square$

Applications for iterates of curves. We now explain how Lemma 9.1 will be used in our proof of Theorem 1.2. As mentioned above, we will consider a mapping class f and a curve c. The role of D will be played by a neighborhood of $c \cup f(c)$, so that $f(D)$ corresponds to a neighborhood of $f(c) \cup f^2(c)$. If c and $f(c)$ form a pair of type I, then $c \cup f(c)$ is a bounding pair and D is not connected. On the other hand, if c and $f(c)$ form a pair of type II, III, or IV, then $i(c, f(c))$ is nonzero and so D is connected.

As in Section 8, we may cut S_g along $f(c)$ in order to obtain a surface with two boundary components. The curves c and $f^2(c)$ each correspond to a collection of arcs on the cut surface. Finding a good pair of curves a and b as in the statement of Lemma 9.1, with $D = c \cup f(c)$, then reduces to finding a pair of nonseparating curves a and b on the cut surface so that:

(1) a and b represent homology classes of S_g that are local with respect to D,

(2) a and b lie on different sides of the c-arcs, and

(3) a and b lie on the same side of the $f^2(c)$-arcs

(the term "sides" here is perhaps an abuse of terminology; what we precisely mean is that a and b lie in different components of the complement of c in the cut surface and in the same component of the complement of $f^2(c)$ in the cut surface). If $i(c, f(c)) = 0$ then by Lemma 9.2 the first condition is satisfied whenever a and b are not parallel to a component of the boundary of the cut surface. Otherwise by Lemma 9.2 the first condition is automatically satisfied.

We summarize the above discussion with the following two lemmas. The first addresses the case $i(c, f(c)) = 0$ and the second addresses the case $i(c, f(c)) \neq 0$. In both cases the hypotheses of the lemma force the genus g to be at least 3, which is why we can conclude that f is a normal generator and not just that its normal closure contains the commutator subgroup. At the same time, we could just as well assume here that g is at least 3, since this will be the case when we apply these lemmas in the proof of Theorem 1.2.
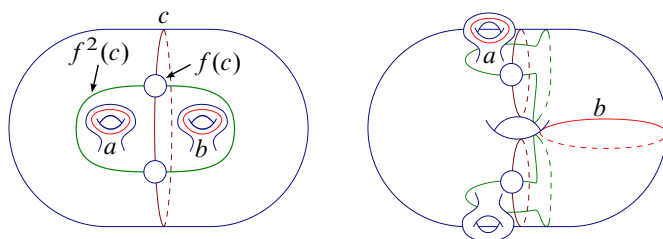
Figure 26. In each case the curves a and b form a pair of good curves for f, with the role of D being played by a neighborhood of $c \cup f(c)$.

Lemma 9.3. *Let $f \in \mathrm{Mod}(S_g)$. Suppose there is a curve c so that $i(c, f(c)) = 0$ and so that $[c] = [f(c)]$. Suppose further there are curves a and b in S_g so that on the surface obtained by cutting S_g along $f(c)$ we have that*

(1) *a and b lie on different sides of c,*

(2) *a and b lie on the same side of $f^2(c)$, and*

(3) *a and b are nonseparating curves in the surface obtained by further cutting along c.*

Then a and b form a good pair for f, and so f is a normal generator for $\mathrm{Mod}(S_g)$.

Lemma 9.4. *Let $f \in \mathrm{Mod}(S_g)$. Suppose there is a curve c and nonseparating curves a and b in S_g so that $i(c, f(c)) > 0$ and so that on the surface obtained by cutting S_g along $f(c)$ we have that*

(1) *a and b lie on different sides of c and*

(2) *a and b lie on the same side of $f^2(c)$.*

Then a and b form a good pair for f, and so f is a normal generator for $\mathrm{Mod}(S_g)$.

In Figure 26 we give two examples of good pairs as in Lemma 9.4. In the example on the left, the curves c, $f(c)$, and $f^2(c)$ lie in configuration (2) from Figure 23, and in the example on the right they lie in configuration (3) from Figure 20.

As with Lemma 9.1, there is an asymmetry in the statements of the two preceding lemmas. For example, in Lemma 9.4, if a and b lie on different sides of $f^2(c)$ but on the same side of c, then a and b form a good pair for f^{-1}. Again, f is a normal generator if and only if f^{-1} is, and so in this case we will also refer to a and b as a good pair for f.

10. Application: Pseudo-Anosov mapping classes with small stretch factor

In this section we use the tools developed in Sections 6–9 to prove Theorem 1.2, which states that a pseudo-Anosov element of $\mathrm{Mod}(S_g)$ with stretch factor less than $\sqrt{2}$ is a normal generator for $\mathrm{Mod}(S_g)$.

In Section 8 we introduced terminology for various types of pairs and ordered triples of curves, namely, types I, II, III, and IV. In this section we say that a pair or a triple of (homotopy classes of) curves is of type I, II, III, or IV if there are representatives of the curves that form a configuration of that type. The representatives chosen do not affect the designation into types.

Proof of Theorem 1.2. As discussed in the introduction, we may assume that $g \geq 3$ since there does not exist a pseudo-Anosov mapping class in $\mathrm{Mod}(S_g)$ with stretch factor less than $\sqrt{2}$ when $g < 3$. Let $f \in \mathrm{Mod}(S_g)$ be a pseudo-Anosov mapping class with stretch factor less than $\sqrt{2}$. We would like to show that f is a normal generator for $\mathrm{Mod}(S_g)$. We will apply the various well-suited curve criteria we have developed. Some of these lemmas only imply that the normal closure of f contains the commutator subgroup of f, but since $g \geq 3$ this implies that f is a normal generator.

We follow the plan we established in Section 6. Let c be a shortest curve in S_g with respect to some f-metric. Since the stretch factor of f is less than $\sqrt{2} < 3/2$ it follows from Proposition 6.1 that $i(c, f(c)) \leq 2$. If c is separating then f is a normal generator by our well-suited curve criterion for separating curves, Lemma 2.4. We may thus assume for the remainder of the proof that c is nonseparating.

Suppose first that $[c] \neq [f(c)]$ mod 2. There are three possibilities for $i(c, f(c))$: 0, 1, or 2. Applying Lemmas 2.3, 2.2, and 7.2 to the three cases, respectively, we conclude that f is a normal generator for $\mathrm{Mod}(S_g)$.

We may now assume for the remainder of the proof that $[c] = [f(c)]$ mod 2. This is Case 3 from Section 6. We have that $i(c, f(c)) = i(f(c), f^2(c))$ and $i(c, f^2(c))$ are both equal to 0 or 2. Again, this follows from the $(k, n) = (2, 4)$ case of Proposition 6.1 plus the fact that $i(c, f^2(c))$ is even (since f, and hence f^2, preserves $[c]$ mod 2).

It follows from the previous paragraph that $\{c, f(c)\}$, $\{f(c), f^2(c)\}$, and $\{c, f^2(c)\}$ are each of type I, II, III, or IV. Since $\{c, f(c)\}$ and $\{f(c), f^2(c)\}$ are of the same type (they differ by f), and since c, $f(c)$, and $f^2(c)$ are all distinct (f is pseudo-Anosov), the ordered triple $(c, f(c), f^2(c))$ is of type I, II, III, or IV. We treat the four possibilities in turn.

Type I. By Proposition 8.1, the configuration $(c, f(c), f^2(c))$ is a stabilization of one of the two configurations in Figure 14. In the first configuration there is a good pair of curves as in Lemma 9.3, and so f is a normal generator.

For the second configuration, there may not be a good pair of curves. Consider then the quadruple of curves $(c, f(c), f^2(c), f^3(c))$. Since $\sqrt{2} < \sqrt[3]{3}$, the stretch factor of f is less than $\sqrt[3]{3}$ and so by the $(k, n) = (3, 6)$ case Proposition 6.1 and the fact that $[c] = [f^3(c)]$ mod 2 we have $i(c, f^3(c)) \leq 4$.

The curves c and $f^2(c)$ form a pair of type II. As such, there are two separating curves d_1 and d_2 obtained as a boundary component of a neighborhood of $c \cup f^2(c)$.

We will show that at least one of $i(d_1, f(d_1))$ or $i(d_2, f(d_2))$ is at most 2. It will then follow from Lemma 2.4 that f is a normal generator.

The curves c and $f^2(c)$ cut each other into two arcs each, and we may think of d_1 and d_2 as being formed from these four arcs. Two of the arcs make up d_1 and the other two make up d_2. Similarly, $f(d_1)$ and $f(d_2)$ are made of the four corresponding arcs of $f(c)$ and $f^3(c)$. We may therefore understand how $f(d_1)$ and $f(d_2)$ intersect d_1 and d_2 by understanding how $f(c)$ and $f^3(c)$ intersect c and $f^2(c)$. Specifically,

$$i(f(d_1), d_1) + i(f(d_1), d_2) + i(f(d_2), d_1) + i(f(d_2), d_2)$$
$$\leq i(c, f(c)) + i(f^2(c), f(c)) + i(c, f^3(c)) + i(f^2(c), f^3(c)).$$

Let us examine the right-hand side. By assumption,

$$i(c, f(c)) = i(f(c), f^2(c)) = i(f^2(c), f^3(c)) = 0.$$

We already said that $i(c, f^3(c)) \leq 4$ and so the left-hand side is at most 4. Thus, $i(f(d_i), d_i) < 4$ for some i. Since d_i is separating it must be that $i(f(d_i), d_i) \leq 2$ as desired.

Type II. By Proposition 8.1, the configuration $(c, f(c), f^2(c))$ is a stabilization of one of the ten configurations in Figure 17. For configurations 1–5 there is a good pair of curves as in Lemma 9.4, and so f is a normal generator. For configurations 6–10, we consider either of the two separating curves obtained as a component of the boundary of a neighborhood of $c \cup f(c)$ (these are $\gamma \cup \delta$ in the pictures); call it d. Its image is one of the two separating curves e_1 and e_2 obtained as a component of the boundary of $f(c) \cup f^2(c)$ (these are $\delta \cup \varepsilon$ in the pictures). Each of $i(d, e_1)$ and $i(d, e_2)$ is either 0 or 2. It thus follows from Lemma 2.4 that f is a normal generator. (The argument for configurations 6–10 also applies to configurations 1–5.)

Type III. By Proposition 8.6, the configuration $(c, f(c), f^2(c))$ is a stabilization of one of the sixteen configurations in Figure 20. There are seven pictures in the figure. Some of the pictures correspond to more than one minimal configuration, but our arguments will apply uniformly to all the different minimal configurations represented by the same picture.

In pictures 1–5 we find a good pair of curves and so by Lemma 9.4 we conclude that f is a normal generator. For the configurations depicted in pictures 6 and 7, consider the four curves lying on the boundary of a neighborhood of (representatives of) c and $f(c)$ (these are γ and δ in the pictures). These four curves must map under f to the four curves lying on the boundary of a neighborhood of $f(c)$ and $f^2(c)$ (these are δ and ε in the pictures). In each of the pictures, one of the latter four curves is the curve d surrounding one of the black or gray dots. (Note that since d is the image of a nonseparating curve it must be itself nonseparating; hence the configuration where the gray dot represents a handle connecting the region to itself does not arise here.) The curve d is disjoint from the first four curves and not
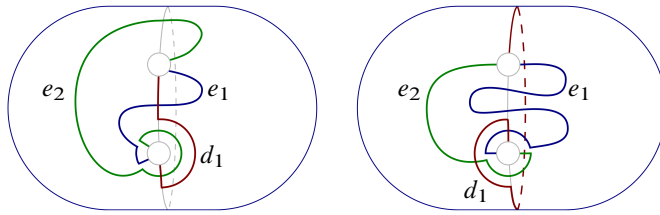
Figure 27. The curves d_1, e_1 and e_2 for configurations 3–5 (left) and 6–8 (right).

homologous to any of them (we can find a curve that intersects each of the first four curves in a single point and does not intersect d). By Lemma 2.3, we have that f is a normal generator in each case.

Type IV. By Proposition 8.7, the configuration $(c, f(c), f^2(c))$ is a stabilization of one of the eight configurations in Figure 23. For configurations 1 and 2 there is a good pair of curves and so by Lemma 9.4 we have that f is a normal generator.

Next consider configurations 3–5. The curves c and $f(c)$ (γ and δ in the picture) intersect in two points, and hence each cuts the other into two arcs. There are four curves in the surface comprised of one of the two arcs of c and one of the two arcs of $f(c)$. Two of these four curves make a left turn onto $f(c)$ from c, call them d_1 and d_2; see the left-hand side of Figure 27. We can similarly make two curves e_1 and e_2 in $f(c) \cup f^2(c)$ that turn left onto $f^2(c)$. The former pair of curves must map to the latter. Since $i(d_1, e_1) = i(d_1, e_2) = 1$ and since $f(d_1)$ is either e_1 or e_2, it follows from Lemma 2.2 that f is a normal generator.

For configurations 6–8 the argument works in the same way. The analogous curves d_1, e_1, and e_2 are shown in the right-hand side of Figure 27. Thus, f is a normal generator in this case as well. This completes the proof of the theorem. □

References

[1] I. Agol, C. J. Leininger, and D. Margalit, Pseudo-Anosov stretch factors and homology of mapping tori. *J. Lond. Math. Soc. (2)* **93** (2016), no. 3, 664–682 Zbl 1388.37033 MR 3509958

[2] J. Aramayona and J. Souto, Homomorphisms between mapping class groups. *Geom. Topol.* **16** (2012), no. 4, 2285–2341 Zbl 1262.57003 MR 3033518

[3] H. Baik, E. Kin, H. Shin, and C. Wu, Asymptotic translation lengths and normal generations of pseudo-Anosov monodromies for fibered 3-manifolds. 2019, arXiv:1909.00974

[4] J. S. Birman and H. M. Hilden, On the mapping class groups of closed surfaces as covering spaces. In *Advances in the theory of Riemann surfaces (Proc. Conf., Stony Brook, N.Y., 1969)*, pp. 81–115, Ann. of Math. Studies 66, Princeton Univ. Press, Princeton, NJ, 1971 Zbl 0217.48602 MR 0292082

[5] J. S. Birman and H. M. Hilden, Isotopies of homeomorphisms of Riemann surfaces and a theorem about Artin's braid group. *Bull. Amer. Math. Soc.* **78** (1972), 1002–1004 Zbl 0255.57002 MR 307217

[6] L. E. J. Brouwer, Über die periodischen Transformationen der Kugel. *Math. Ann.* **80** (1919), no. 1, 39–41 Zbl 47.0527.01 MR 1511946

[7] L. Chen, K. Kordek, and D. Margalit, Homomorphisms between braid groups. 2019, arXiv:1910.00712

[8] L. Chen and J. Lanier, Constraining mapping class group homomorphisms using finite subgroups. 2021, arXiv:2112.07843

[9] J.-H. Cho and J.-Y. Ham, The minimal dilatation of a genus-two surface. *Experiment. Math.* **17** (2008), no. 3, 257–267 Zbl 1153.37375 MR 2455699

[10] M. Clay, J. Mangahas, and D. Margalit, Right-angled Artin groups as normal subgroups of mapping class groups. *Compos. Math.* **157** (2021), no. 8, 1807–1852 Zbl 07384181 MR 4292178

[11] F. Dahmani, V. Guirardel, and D. Osin, Hyperbolically embedded subgroups and rotating families in groups acting on hyperbolic spaces. *Mem. Amer. Math. Soc.* **245** (2017), no. 1156 Zbl 1396.20041 MR 3589159

[12] D. Dugger, Involutions on surfaces. *J. Homotopy Relat. Struct.* **14** (2019), no. 4, 919–992 Zbl 1431.57026 MR 4025595

[13] S. Eilenberg, Sur les transformations périodiques de la surface de sphère. *Fund. Math.* **22** (1934), 28–41 Zbl 0008.37109

[14] B. Farb, Some problems on mapping class groups and moduli space. In *Problems on mapping class groups and related topics*, pp. 11–55, Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI, 2006 Zbl 1191.57015 MR 2264130

[15] B. Farb, C. J. Leininger, and D. Margalit, The lower central series and pseudo-Anosov dilatations. *Amer. J. Math.* **130** (2008), no. 3, 799–827 Zbl 1187.37060 MR 2418928

[16] B. Farb and D. Margalit, *A primer on mapping class groups*. Princeton Math. Ser. 49, Princeton Univ. Press, Princeton, NJ, 2012 Zbl 1245.57002 MR 2850125

[17] A. Fathi, F. Laudenbach, F. Laudenbach, and V. Poénaru, *Thurston's work on surfaces*. Math. Notes 48, Princeton Univ. Press, Princeton, NJ, 2012 Zbl 1244.57005 MR 3053012

[18] L. Funar, On power subgroups of mapping class groups. *J. Gökova Geom. Topol. GGT* **8** (2014), 14–34 Zbl 1395.57002 MR 3310570

[19] W. J. Harvey and M. Korkmaz, Homomorphisms from mapping class groups. *Bull. London Math. Soc.* **37** (2005), no. 2, 275–284 Zbl 1066.57020 MR 2119027

[20] E. Hironaka, Small dilatation mapping classes coming from the simplest hyperbolic braid. *Algebr. Geom. Topol.* **10** (2010), no. 4, 2041–2060 Zbl 1221.57028 MR 2728483

[21] S. Hirose, Presentations of periodic maps on oriented closed surfaces of genera up to 4. *Osaka J. Math.* **47** (2010), no. 2, 385–421 Zbl 1202.57022 MR 2722366

[22] N. V. Ivanov, Fifteen problems about the mapping class groups. In *Problems on mapping class groups and related topics*, pp. 71–80, Proc. Sympos. Pure Math. 74, Amer. Math. Soc., Providence, RI, 2006 Zbl 1281.57011 MR 2264532

[23] D. L. Johnson, Homeomorphisms of a surface which act trivially on homology. *Proc. Amer. Math. Soc.* **75** (1979), no. 1, 119–125 Zbl 0407.57003 MR 529227

[24] F. Klein, Ueber Realitätsverhältnisse bei der einem beliebigen Geschlechte zugehörigen Normalcurve der φ. *Math. Ann.* **42** (1893), no. 1, 1–29 Zbl 25.0689.03 MR 1510765

[25] M. Korkmaz, Generating the surface mapping class group by two elements. *Trans. Amer. Math. Soc.* **357** (2005), no. 8, 3299–3310 Zbl 1079.57018 MR 2135748

[26] R. S. Kulkarni, Riemann surfaces admitting large automorphism groups. In *Extremal Riemann surfaces (San Francisco, CA, 1995)*, pp. 63–79, Contemp. Math. 201, Amer. Math. Soc., Providence, RI, 1997 Zbl 0863.30050 MR 1429195

[27] J. Lanier, Generating mapping class groups with elements of fixed finite order. *J. Algebra* **511** (2018), 455–470 Zbl 1432.57046 MR 3834783

[28] J. Lanier and D. Margalit. Normal generators for mapping class groups are abundant. 2018, arXiv:1805.03666v1

[29] E. Lanneau and J.-L. Thiffeault, On the minimum dilatation of pseudo-Anosov homeomorphisms on surfaces of small genus. *Ann. Inst. Fourier (Grenoble)* **61** (2011), no. 1, 105–144 Zbl 1237.37027 MR 2828128

[30] C. J. Leininger and D. Margalit, On the number and location of short geodesics in moduli space. *J. Topol.* **6** (2013), no. 1, 30–48 Zbl 1293.57011 MR 3029420

[31] M. Leśniak, Torsion normal generators of the mapping class group of a nonorientable surface. *J. Pure Appl. Algebra* **226** (2022), no. 6, Paper No. 106964 Zbl 07461220

[32] W. B. R. Lickorish, A finite set of generators for the homeotopy group of a 2-manifold. *Proc. Cambridge Philos. Soc.* **60** (1964), 769–778 Zbl 0131.20801 MR 171269

[33] D. D. Long, A note on the normal subgroups of mapping class groups. *Math. Proc. Cambridge Philos. Soc.* **99** (1986), no. 1, 79–87 Zbl 0584.57008 MR 809501

[34] F. Luo, Torsion elements in the mapping class group of a surface. 2000, arXiv:math.GT/0004048v1

[35] J. Maher, Random walks on the mapping class group. *Duke Math. J.* **156** (2011), no. 3, 429–468 Zbl 1213.37072 MR 2772067

[36] J. Maher and G. Tiozzo, Random walks, WPD actions, and the Cremona group. *Proc. Lond. Math. Soc. (3)* **123** (2021), no. 2, 153–202 Zbl 07429310 MR 4307033

[37] K. Mann and M. Wolff, Rigidity of mapping class group actions on S^1. *Geom. Topol.* **24** (2020), no. 3, 1211–1223 Zbl 1454.57014 MR 4157553

[38] H. Masai, Fibered commensurability and arithmeticity of random mapping tori. *Groups Geom. Dyn.* **11** (2017), no. 4, 1253–1279 Zbl 1423.57037 MR 3737282

[39] H. A. Masur and Y. N. Minsky, Geometry of the complex of curves. I. Hyperbolicity. *Invent. Math.* **138** (1999), no. 1, 103–149 Zbl 0941.32012 MR 1714338

[40] J. McCarthy and A. Papadopoulos, Involutions in surface mapping class groups. *Enseign. Math. (2)* **33** (1987), no. 3-4, 275–290 Zbl 0655.57005 MR 925990

[41] J. Mennicke, Zur Theorie der Siegelschen Modulgruppe. *Math. Ann.* **159** (1965), 115–129 Zbl 0134.26502 MR 181676

[42] D. Mumford, Abelian quotients of the Teichmüller modular group. *J. Analyse Math.* **18** (1967), 227–244 Zbl 0173.22903 MR 219543

[43] R. C. Penner, Bounds on least dilatations. *Proc. Amer. Math. Soc.* **113** (1991), no. 2, 443–450 Zbl 0726.57013 MR 1068128

[44] A. Putman, A note on the connectivity of certain complexes associated to surfaces. *Enseign. Math. (2)* **54** (2008), no. 3-4, 287–301 Zbl 1182.57004 MR 2478089

[45] B. von Kerékjártó, Über die periodischen Transformationen der Kreisscheibe und der Kugelfläche. *Math. Ann.* **80** (1919), no. 1, 36–38 Zbl 47.0526.05 MR 1511945

[46] K. Yoshihara, Generating the mapping class group of a surface by torsion. Lecture given at *Knots 2016*, Nihon University, December 2016, `https://www.math.chs.nihon-u.ac.jp/~ichihara/Knots2016/Proceedings/yoshihara.pdf`

J. Lanier, Department of Mathematics, University of Chicago,
5734 S University Avenue, Chicago, IL 60637, USA
E-mail: jlanier8@uchicago.edu

D. Margalit, School of Mathematics, Georgia Institute of Technology,
686 Cherry Street, Atlanta, GA 30332, USA
E-mail: margalit@math.gatech.edu