Pergamon

Topology Vol. 36, No. 2, pp. 381–410, 1997

0040-9383/96/$15.00 + 0.00

S0040-9383(96)00012-2

ON NON-SEPARATING SIMPLE CLOSED CURVES IN A COMPACT SURFACE

FENG LUO

(*Received for publication* 19 *February* 1996)

We introduce a semi-algebraic structure on the set $\mathscr{S}$ of all isotopy classes of non-separating simple closed curves in any compact oriented surface and show that the structure is finitely generated. As a consequence, we produce a natural finite dimensional linear representation of the mapping class group of the surface. Applications to the Teichmüller space, Thurston's measured lamination space, the harmonic Beltrami differentials, and the first cohomology groups of the surface are discussed. Copyright © 1996 Elsevier Science Ltd

1. INTRODUCTION

1.1. The main results

Given a compact oriented surface of positive genus Σ, let $\mathscr{S} = \mathscr{S}(\Sigma)$ be the set of all isotopy classes of non-separating simple closed unoriented curves in Σ. We introduce two relations, *orthogonal* and *disjoint* in $\mathscr{S}$ as follows. Two classes α and β in $\mathscr{S}$ are said to be *orthogonal*, denoted by $\alpha \perp \beta$, if they represent simple closed curves a and b intersecting transversely at one point (in this case, we also say that a is *orthogonal* to b and denote it by $a \perp b$). Two classes α and β in $\mathscr{S}$ are said to be *disjoint*, denoted by $\alpha \cap \beta = \phi$, if they represent simple closed curves a and b so that $a \cap b = \phi$. Our goal is to study $\mathscr{S}$ under these two relations.

Given two orthogonal simple closed curves p and q, define the product pq of p and q to be $D_p(q)$ where D_c is the positive Dehn twist (a right twist) about the simple closed curve c. Geometrically, pq and qp are obtained from $p \cup q$ by breaking the intersection into two embedded arcs as in Fig. 1.

Clearly the isotopy classes of pq and qp depend only on the isotopy classes of p and q. We use $[p]$ to denote the isotopy class of a simple curve p and define the product $[p][q]$ to be $[pq]$ when $p \perp q$. Our first result states that $\mathscr{S}$ is finitely generated in the product. Actually, a stronger form of the finiteness result holds. It is on the stronger form that we will focus. To this end, we introduce the following definition.

Definition. Given a subset χ of $\mathscr{S}$, the *derived set* χ' is $\chi \cup \{\alpha\beta \mid \alpha, \beta, \text{ and } \beta\alpha \text{ are in } \chi\}$. Inductively define χ^n to be the derived set of χ^{n-1} for $n > 1$. We define $\bigcup_{n=1}^{\infty} \chi^n$ to be the set *generated by* χ, and denote it by χ^∞.

Our first theorem is the following.

THEOREM I. *If Σ is a compact orientable surface of positive genus, then there is a finite subset $\mathscr{F}$ of isotopy classes of non-separating simple closed curves so that $\mathscr{F}^\infty = \mathscr{S}(\Sigma)$.*

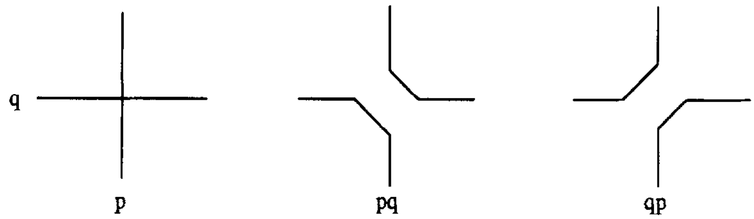

Fig. 1. In the figure, the orientation is the right-handed orientation in the plane.

The next theorem (in [1]) is an analogue of Dehn–Nielsen's theorem that the mapping class group of a closed surface is the outer automorphism group of the fundamental group of the surface. To be more precise, given any orientation preserving self-homeomorphism ϕ of Σ, ϕ induces a bijective map of $\mathcal{S}$ by the formula $\phi_*([a]) = [\phi(a)]$. Clearly, ϕ_* preserves disjointness, orthogonality and the product.

THEOREM II. *If h is a bijective map of the set of all isotopy classes of non-separating simple closed curves in a closed orientable surface so that h preserves disjointness, orthogonality and the product, then h is induced by an orientation preserving self-homeomorphism of the surface.*

In [12], N. Ivanov has sketched a proof of a stronger version of Theorem II for surfaces with genus $g \geq 2$.

1.2. The motivations

Theorems I and II are well known for the torus T^2. As usual, choose an oriented meridian m and an oriented longitude l for T^2. Each element in $\mathcal{S}$ is uniquely represented by $\pm(p[m] + q[l])$ in the first homology group where p and q are two relatively prime integers. By assigning the rational number p/q to the class, we identify $\mathcal{S}$ with $\mathbf{Q}\cup\{\infty\} \subset \mathbf{R}\cup\{\infty\}$ which is considered to be the natural boundary of the hyperbolic upper half-plane. Two classes p/q and r/s are orthogonal if and only if $ps - qr = \pm 1$. Furthermore, if $\beta\perp\gamma$, then $\beta\gamma$ and $\gamma\beta$ are symmetric with respect to the hyperbolic reflection about the geodesic ending at β and γ. Thus by the well known modular picture, one sees that $\mathcal{S}$ is generated by 0, 1, and ∞ as in Fig. 2.

To see Theorem I for the torus, let us take a bijective map h of $\mathcal{S} = \mathbf{Q}\cup\{\infty\}$ preserving the orthogonality. Thus $h(0)$, $h(1)$ and $h(\infty)$ are three pairwise orthogonal rational numbers. Therefore, there is an element $\psi \in \mathrm{GL}(2, \mathbf{Z})$ acting on $\mathbf{Q}\cup\{\infty\}$ as Möbius transformations so that $\psi(0) = h(0)$, $\psi(1) = h(1)$ and $\psi(\infty) = h(\infty)$. Since $\mathrm{GL}(2, \mathbf{Z})$ is the mapping class group of the torus, we may assume at that h leaves the three curves 0, 1 and ∞ fixed. Thus by the modular pictures above, the bijection h leaves each rational number fixed.

The other motivation for Theorem I comes from Thurston's compactification of the Teichmüller space [2, 3], Bonahon's interpretation of Thurston's compactification [4] and the trace formula $\mathrm{tr}(XY) + \mathrm{tr}(X^{-1}Y) = \mathrm{tr}(X)\mathrm{tr}(Y)$ for X and Y in $\mathrm{SL}(2, \mathbf{C})$. If two elements A and B in $\pi_1(\Sigma)$ have representatives in the free homotopy class $[a]$, $[b]$ in $\mathcal{S}$ so that $a\perp b$, then the products AB and $A^{-1}B$ are represented by the classes $[ab]$ and $[ba]$. As a consequence of the trace formula, the hyperbolic lengths of the classes A, B, AB and $A^{-1}B$ with the above property satisfy the following nonlinear relation: $\cosh(l_{AB}/2) + \cosh(l_{A^{-1}B}/2) = 2\cosh(l_A/2)\cosh(l_B/2)$, i.e., it satisfies the following,

$$f(\alpha\beta) + f(\beta\alpha) = 2f(\alpha)f(\beta), \quad \alpha\perp\beta. \tag{1}$$

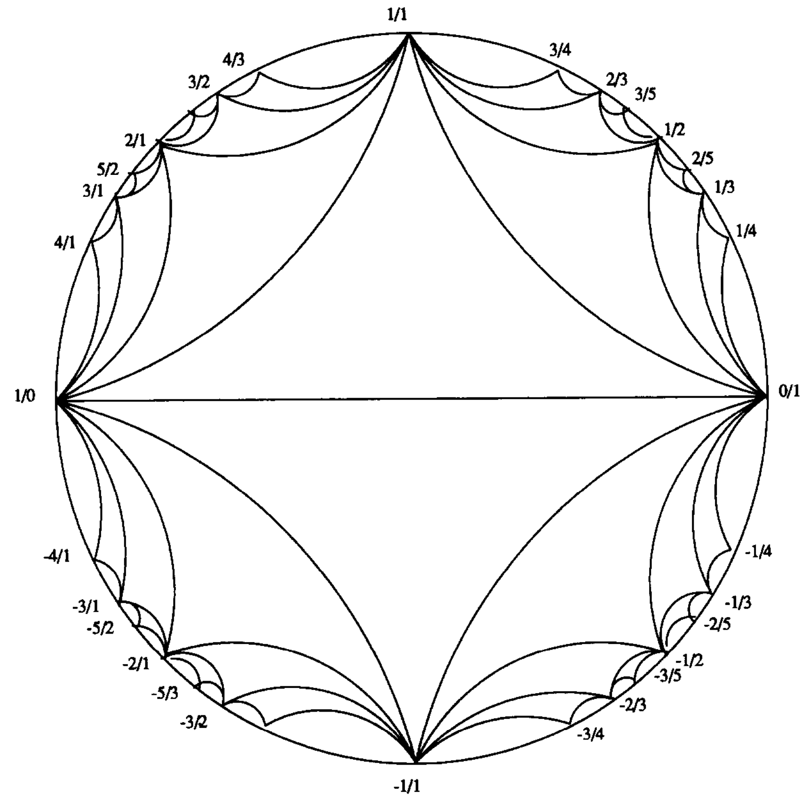

Fig. 2.

Now if we consider a degenerate family of hyperbolic metrics which tends projectively to a simple closed curve c (or more generally a measured lamination) in Thurston's compactification of the Teichmüller space, then the relation (1) degenerates to $\mathbf{I}([a], [c]) + \mathbf{I}([b], [c]) = \max(\mathbf{I}([ab], [c])), \mathbf{I}([ba], [c])$, i.e. it satisfies the following,

$$f(\alpha) + f(\beta) = \max(f(\alpha\beta), f(\beta\alpha)), \quad \alpha \perp \beta \tag{2}$$

where $\mathbf{I}(\cdot, \cdot)$ is the geometric intersection number. Since eq. (2) is piecewise linear ($\max(x, y) = \frac{1}{2}(x + y + |x - y|)$), which may be one of the reasons why the action of the mapping class group on the measured lamination space is piecewise linear.

Since both the Teichmüller space and the measured lamination space are finite dimensional, this prompts us to ask about the existence of finite generators for $\mathscr{S}$.

1.3. Some consequences

As a consequence of Theorem I, we produce a finite dimensional linear representation of the mapping class group $\mathscr{G} = \mathscr{G}(\Sigma)$ of the surface Σ. Recall that the mapping class group $\mathscr{G}(\Sigma)$ is defined to be $\mathrm{Homeo}^+(\Sigma)$/isotopies where homeomorphisms are orientation preserving and $\mathscr{G}$ acts naturally on $\mathscr{S}$ by permuting the isotopy classes. The finite dimensional linear representations of $\mathscr{G}$ are constructed as follows. Take V to be the vector space $\mathbf{C}^{\mathscr{S}}$ of all complex valued functions on $\mathscr{S}$, i.e. $V = \{f: \mathscr{S} \to \mathbf{C}\}$. The mapping class group $\mathscr{G}$ acts naturally linearly on V by permuting the $\mathscr{S}$ coordinates. Let a, b, c and d be four specified complex numbers with $cd \neq 0$. Define a linear subspace $V_{a,b,c,d}$ of V by the equations

$$af(\alpha) + bf(\beta) = cf(\alpha\beta) + df(\beta\alpha), \quad \alpha \perp \beta. \tag{3}$$

By Theorem I, $V_{a,b,c,d}$ is finite dimensional and is clearly invariant under the linear action of the mapping class group.

PROPOSITION. *If the genus of the surface is bigger than one, then the only space $V_{a,b,c,d}$ which contains a non-constant function is $V_{2a,2a,a,a} = V_{2,2,1,1}$. Furthermore, the dimension of $V_{2,2,1,1}$ is at least $(2g+n)(2g+n-1)/2$ where g is the genus and n is the number of boundary components of the surface.*

The basic idea of the proof (in [1]) is the following. If a simple closed curve s is orthogonal to another simple closed curve t, then there is a universal relation that $[s(ts)] = [t]$ (and also $[(st)s] = [t]$). This relation $[s(ts)] = [t]$ implies the well known braid relation $D_sD_tD_s = D_tD_sD_t$ up to isotopy in the mapping class group. By iterating the relation $[s(ts)] = [t]$ several times for three pairwise orthogonal simple closed curves which are not in a torus with a hole, we obtain that $2a = 2b = c = d$. To show that $V_{2,2,1,1}$ contains non-constant solutions, we take the square of the algebraic intersection number $f(\alpha) = i^2(\alpha, \gamma)$ for a fixed class γ in $\mathscr{S}$. One sees easily that f satisfies the skein relation $f(\alpha\beta) + f(\beta\alpha) = 2f(\alpha) + 2f(\beta)$ whenever $\alpha \perp \beta$. Since for all choices of γ in $\mathscr{S}$, these functions span a linear subspace of dimension at least $(2g+n)(2g+n-1)/2$ in $V_{2,2,1,1}$ (the actual dimension of the subspace is $(2g+n)(2g+n-1)/2$ if $n \neq 0$ and is $g(2g+1)$ if $n = 0$), the estimate on the dimension follows.

Theorem I may give rise to a characterization of the length spectrum of a hyperbolic metric in a closed surface. Given a closed orientable surface Σ, let Teich (Σ) be the Teichmüler space of Σ. For each equivalence class $[d]$ of hyperbolic metric in Teich(Σ), we produce a function $f_{[d]}$ from $\mathscr{S}$ to $\{t \in \mathbb{R} \mid t > 1\}$ by setting $f_{[d]}(\alpha) = \cosh(l_d(\alpha)/2)$ where $l_d(\alpha)$ is the length of the geodesic in the class α in the metric d. By the remark above, we have

$$f_{[d]}(\alpha\beta) + f_{[d]}(\beta\alpha) = 2f_{[d]}(\alpha)f_{[d]}(\beta), \quad \alpha \perp \beta. \tag{4}$$

The set $\mathscr{T}$ of all functions from $\mathscr{S}$ to $\{t \in \mathbb{R} \mid t > 1\}$ satisfying (1) is finite dimensional by Theorem I. It is well known that the map from Teich(Σ) to $\mathscr{T}$ sending $[d]$ to $f_{[d]}$ is injective and continuous. It is natural to ask whether the map is an onto map.

One possible approach to the above problem is to find the dimension of $\mathscr{T}$. Since harmonic Beltrami differentials are deformations of the hyperbolic metrics in the Teichmüller space, given a function f satisfying the relation (1) above, it is tempting to call a function g satisfying

$$g(\alpha\beta) + g(\beta\alpha) = 2f(\alpha)g(\beta) + 2f(\beta)g(\alpha), \quad \alpha \perp \beta \tag{5}$$

a *harmonic Beltrami differential* in the *conformal structure* given by f. Denote $T_f(\mathscr{T})$ the linear space (the tangent space of $\mathscr{T}$ at f) of all functions g satisfying (5). One would expect to have a Riemann–Roch theorem which calculates the dimension of $T_f(\mathscr{T})$.

1.4. Sketch of the proofs

Since the proofs of Theorems I and II are quite long, we sketch the main idea of the proof below. The main step in the proof of Theorem I is an induction on a norm $\|\Sigma\| = 3g + n$ where Σ is a compact orientable surface of genus g with n boundary components. We show that there are finitely many subsurfaces Σ_i in Σ of smaller norms so that non-separating simple closed curves in Σ_i considered as non-separating simple closed curves in $\mathscr{S}(\Sigma)$ generate $\mathscr{S}(\Sigma)$. The main technical difficulty is due to the fact that separating simple closed curves in Σ_i may become non-separating in Σ.

We use several induction steps to achieve the above result. In each step of the induction process, we construct a finite collection G of isotopy classes of simple closed curves $[a_1], \dots, [a_m]$ and induct on a norm $\|\alpha\|_G = \sum_{i=1}^m I(\alpha, [a_i])$. Given a class $\alpha \in \mathscr{S}$, we would like to find two classes β and γ in $\mathscr{S}$ so that $\beta \perp \gamma$, $\alpha = \beta\gamma$ and the norms of β, γ and $\gamma\beta$ are smaller than $\|\alpha\|_G$. To this end, let us take $p \in \alpha$ so that $\|\alpha\|_G = \sum_{i=1}^m |p \cap a_i|$ and find an arc c in Σ so that $c \cap (p \cup \bigcup_{i=1}^m a_i) = \partial c$ and c approaches both end points of ∂c from different sides of p (we call it a *cutting arc*; see Definition 2.4). Indeed, assuming c has been constructed, we obtain three new non-separating simple closed curves representing β, γ and $\gamma\beta$ as in Fig. 4. Furthermore, $\alpha = \beta\gamma$. We call this an H-reduction on the curve p. Now if the arc c satisfies $\partial c = c \cap (\bigcup_{i=1}^m a_i)$, then clearly $\|\beta\|_G$, $\|\gamma\|_G$ are less than $\|\alpha\|_G$ (Lemma 2.5) and $\|\gamma\beta\|_G \leqslant \|\alpha_G\|$.

Let us illustrate this by considering a torus, i.e. the modular picture is generated by hyperbolic reflections on the three sides of the ideal triangle with vertices $0 = [l]$, $1 = [ml]$ and $\infty = [m]$. Given a class $\alpha = \pm(p[m] + q[l])$ where p and q are relatively prime integers, define a norm $\|\alpha\| = |p| + |q| = \inf\{|a \cap m| + |a \cap l| \,|\, a \in \alpha\}$. Let $a \in \alpha$ be a representative so that $\|\alpha\| = |a \cap m| + |a \cap l|$ where $|a \cap l| = |p|$ and $|a \cap m| = |q|$. Then all intersection points in $a \cap m$ (and in $a \cap l$, respectively) have the same intersection signs. Thus, if one of the numbers $|a \cap m|$ or $|a \cap l|$ is bigger than 1, say $|a \cap l| \geqslant 2$, then there are two adjacent intersection points x and y in the curves l so that there is an arc c in l joining x and y with $c \cap a = \partial c$ and $c \cap m = \phi$. Then, the H-reduction on the curve a at c produces two classes β and γ so that $\beta \perp \gamma$, $\alpha = \beta\gamma$ and the norms of β, γ and $\gamma\beta$ are smaller than $\|\alpha\|$. Finally, if both p and q are at most 1, then α is one of the four classes $[m]$, $[l]$, $[ml]$ and $[lm]$. Thus the result follows.

There are two steps in the proof of Theorem I. In the first step, we show that there are finitely many non-separating simple closed curves $\{c_1, \dots, c_k\}$ in $\mathscr{S}$ so that $\mathscr{X} = \{\alpha \in \mathscr{S} \,|\, \alpha$ is disjoint from one of $[c_i]\}$ satisfies $(\mathscr{X})^\infty = \mathscr{S}$ (Lemma 2.10). This step is relatively easy to achieve. The next step is the major step in which we replace these curves c_i by non-boundary parallel essential *separating* simple closed curves in Σ. This is achieved by combining Propositions 2.8, 2.15, and 2.18. Knowing this, we apply the induction on the norm $\|\Sigma\|$ of the subsurfaces obtained as the closure of the components of the complement of c_i in Σ and end of proof. The proof of Theorem I is given in Section 2.

To prove Theorem II (in [1]), we show that the Lickorish–Humphries basis $F_0 = \{[a_1], \dots, [a_{2g+1}]\}$ (as a subset of $\mathscr{S}$; see Fig. 3) for the mapping class group of a closed surface of genus g satisfies the following property: Let F_n be the set $\{\alpha \,|\, \alpha = \beta\gamma$, where β and γ are in $F_{n-1}\} \cup F_{n-1}$. Then $\mathscr{S} = \bigcup_{n=1}^\infty F_n$. This also shows that the subgroup generated by the Dehn twists on the curves $[a_i]$'s is the subgroup generated by the Dehn twists on the set of all non-separating simple closed curves. Indeed $D_{pq} = D_p D_q D_p^{-1}$ up to isotopy when $p \perp q$ (see [5–7] for details on generating the mapping class group by Dehn twists). Now any two Lickorish–Humphries bases are related by a self homeomorphism of the surface and a bijective map of $\mathscr{S}$ preserving disjointness and orthogonality sends a Lickorish–Humphries basis to a Lickorish–Humphries basis. Thus the result follows. This is an analogue of the modular

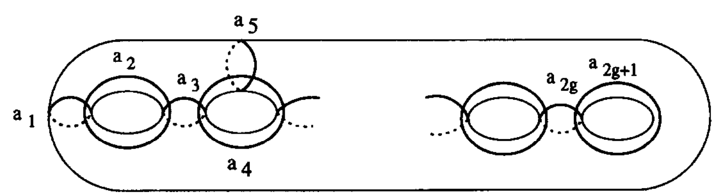

Fig. 3. A Lickorish–Humphries basis.

group case where we use two curves 0 and ∞ as bases. The basic technique of the proof is the same as that used in Theorem I. We do not have to consider the third curve $\gamma\beta$ here but we do need to produce the minimal set of generators in the new sense.

1.5. Some observations and questions

There is an interesting similarity between the relations (1)–(3) above and the functional equations for the well known elementary functions from the set of integers $\mathbf{Z}$ to $\mathbf{R}$. Namely, the trigonometric functions $\cosh(\lambda t)$ satisfy $f(x+y)+f(x-y)=2f(x)f(y)$ with $f(x)>1$, the absolute value functions $|\lambda t|$ satisfy $f(x)+f(y)=\max(f(x+y),f(x-y))$, and the square functions λt^2 satisfy $f(x+y)+f(x-y)=2f(x)+2f(y)$. Here λ is a parameter. It is easy to show that these are all non-constant solutions to the functional equations above. Note that $\cos(\lambda t)$ also satisfies $f(x+y)+f(x-y)=2f(x)f(y)$.

This leads one to ask several questions.

Question 1. What is the dimension of the vector space $V_{2,2,1,1}$?

Question 2. Given a closed orientable surface Σ and a function $f:\mathscr{S}\to\{t\in\mathbf{R}\,|\,t>1\}$ satisfies the relation $f(\alpha\beta)+f(\beta\alpha)=2f(\alpha)f(\beta)$ whenever $\alpha\perp\beta$, is it true that $f=f_{[d]}$ for some hyperbolic metric d in the surface?

It can be shown easily that there are non-constant solutions to the relation $f(\alpha\beta)+f(\beta\alpha)=2f(\alpha)f(\beta)$ which take some values equal to one (see [11] for a solution to the length spectrum problem). One would also ask whether there is a similar notion of the cosine function for curves in surfaces.

In view of the linear skein relations for Jones type knot invariants, one may call a function in $V_{2,2,1,1}$ a two-dimensional "Jones invariant". More generally, one may define two-dimensional "Jones invariants" as follows. Take a finite collection of classes $\alpha_1,\ldots,\alpha_k$ in $\mathscr{S}$ (or more generally the set of all isotopy classes of simple closed curves in the surface) and finite collection of non-zero numbers $d_1,\ldots,d_k$. A two-dimensional "Jones invariant" is a function $f:\mathscr{S}\to\mathbf{C}$ satisfying a linear skein relation:

$$\sum_{i=1}^{k} d_i f(\phi_*(\alpha_i)) = 0 \tag{6}$$

for all ϕ in the mapping class group of the surface.

A natural candidate for the collection $\alpha_1,\ldots,\alpha_k$ seems to be the one satisfying that α_i is either disjoint from or orthogonal to α_j for $i\neq j$. The goal of this approach is to construct finite dimensional linear representations of the mapping class group.

As an example, take three pairwise orthogonal simple closed curves a, b and c so that $a\cap b\cap c\neq\phi$ and $a\cup b\cup c$ is not in a torus with a hole. Let $\alpha=[a]$, $\beta=[b]$ and $\gamma=[c]$. Clearly each function f in $V_{2,2,1,1}$ satisfies the following relation:

$$f(\alpha\beta)+f(\beta\alpha)+f(\beta\gamma)+f(\gamma\beta)+f(\gamma\alpha)+f(\alpha\gamma)=4f(\alpha)+4f(\beta)+4f(\gamma). \tag{7}$$

Is it true that the set of all solutions to relation (7) forms a finite dimensional vector space?

F. Bonahon asked whether one could estimate the number of generators in Theorem I. It follows from the proof that the number of generators for $\Sigma_{g,n}$ is at most c^{3g+n} for some universal constant c. But the smallest number might be quadratic in $3g+n$.

2. PROOF OF THEOREM

2.1. *Notation.* We will use Σ to denote a compact oriented connected surface of genus g with n boundary components. A 1-submanifold in Σ is either a simple closed curve or an embedded arc (possibly non-compact). A submanifold p is *proper* if p is closed in Σ and $\partial p = p \cap \partial\Sigma$. All intersections of two submanifolds in Σ are assumed to be transverse If p and q are two oriented 1-submanifolds, the *intersection sign* at a point $P \in p \cap q$ is defined in the usual way. If P and Q are in $p \cap q$, then the intersection signs at P and Q being the same is independent of the orientations on p, q and Σ.

Notation:

$N(G)$	a small regular neighborhood of a graph (a one-dimensional CW complex) G in Σ
$\lvert X \rvert$	the number of elements in a finite set X
$\lvert p \cap q \rvert$	the geometric intersection number of two 1-submanifolds p and q
$\cong$	isotopy
$int(S)$	the interior of a compact manifold S
$[s]$	isotopy class of simple closed curves s
$I(\alpha, \beta)$	$\inf\{\lvert a \cap b \rvert \mid a \in \alpha, b \in \beta\}$, the geometric intersection number of α and β in $\mathcal{S}$
$i(s, t)$	the algebraic intersection number of two oriented curves s and t
$i([s], [t])$	the algebraic intersection number of the isotopy classes of oriented curves s and t
D_s	the positive Dehn twist about a simple closed curve s

2.2. *Reductions.* We begin by introducing several useful concepts. A *cutting arc* c for a simple closed curve p is a closed embedded arc in Σ with end points in p so that $c \cap p = \partial c \cap p = \partial c$ and that c approaches both end points in ∂c from different sides of p. It is clear that p has a cutting arc if and only if p is non-separating. If two embedded 1-submanifolds p and q of Σ intersect in more than one point, then an *adjacent arc* for q in p is a segment in p which is bounded by two adjacent intersection points and which does not intersect q except at the end points. We will be mainly interested in finding adjacent cutting arcs. This is the same as finding adjacent arcs whose end points have the same intersection signs. In particular, if p is a simple closed curve and $\lvert p \cap q \rvert$ is an odd number, then by checking parity, one finds an adjacent cutting arc for q in p. Again by checking the parity, we have the following very useful lemma.

2.3. LEMMA (alternating principle). *Suppose q_1 and q_2 are two simple closed curves and p is an arc in Σ so that $q_1 \cap p = q_2 \cap p$. Let P and Q be the two outermost intersection points of $q_1 \cap p$ in p. If the intersection signs of P and Q in $q_1 \cap p$ are the same, and the intersection signs of P and Q in $q_2 \cap p$ are different, then there exists an adjacent cutting arc for q_1 or q_2 in p.*

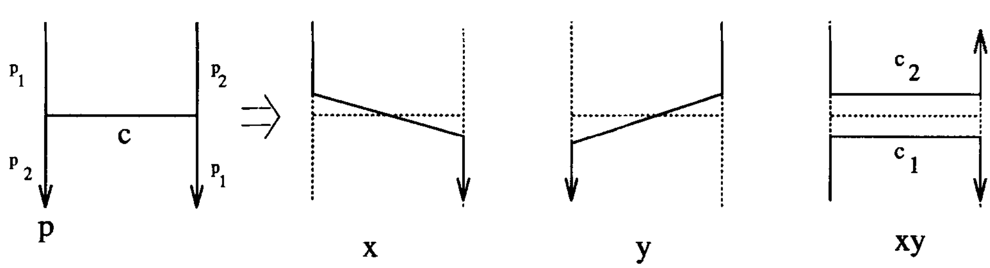

Fig. 4. In the figure, the orientation is the right handed orientation in the plane.

2.4. *Definition* (*The* H-*operation*). Suppose c is a cutting arc for a non-separating simple closed curve p. By changing the graph $c \cup p$ into simple closed curves as in Fig. 4, we obtain three simple closed non-separating curves x, y, xy. They are said to be the curves *spanned by* p and c, or obtained by the H-*operation* on p and c. The curve xy is said to be the *third curve* spanned by s and c.

Note that x, y and xy are all in a regular neighborhood of $c \cup p$. Suppose p is decomposed into $p_1 \cup p_2$ by $c \cap p$. Then the curves x and y are isotopic to $p_1 \cup c$ and $p_2 \cup c$. Take two parallel copies c_1 and c_2 of c with end points in p so that c_1 and c_2 are in different sides of p. The curve xy is isotopic to $(p - int(N(c))) \cup (c_1 \cup c_2)$. By the construction, we have $[x] \perp [p]$, $[y] \perp [p]$, $[x] \perp [y]$, $[xy] \perp [x]$, and $[xy] \perp [y]$. Furthermore, xy is isotopic to $D_x y$ and p is isotopic to $D_y x$. Thus $[p]$ is spanned by $[x]$, $[y]$ and $[xy]$.

2.5. Lemma. *Let G be a graph and s be a simple closed curve in* Σ *so that s avoids the vertices of G and s intersects G transversely. Suppose c is a cutting arc for s and x, y and xy are the curves spanned by c and s.*

(a) *If* $\partial c = c \cap (G \cap s) = c \cap G$ *and a parallel copy* c' *of c with end points in s is disjoint from G (as in Fig.* 5(*a*) *but not in Fig.* 5(*b*)), *then* $|x \cap G| \leqslant |s \cap G| - 1$, $|y \cap G| \leqslant |s \cap G| - 1$ *and* $|xy \cap G| \leqslant |s \cap G|$.

(b) *If G is a* 1-*submanifold and* $c \subset G$ *is an adjacent cutting arc, then* $|x \cap G| \leqslant |s \cap G| - 1$, $|y \cap G| \leqslant |s \cap G| - 1$ *and* $|xy \cap G| \leqslant |s \cap G| - 2$.

The proof is evident from the definition and the fact that x, y and xy are in a regular neighborhood of $c \cup S$. See Fig. 5.

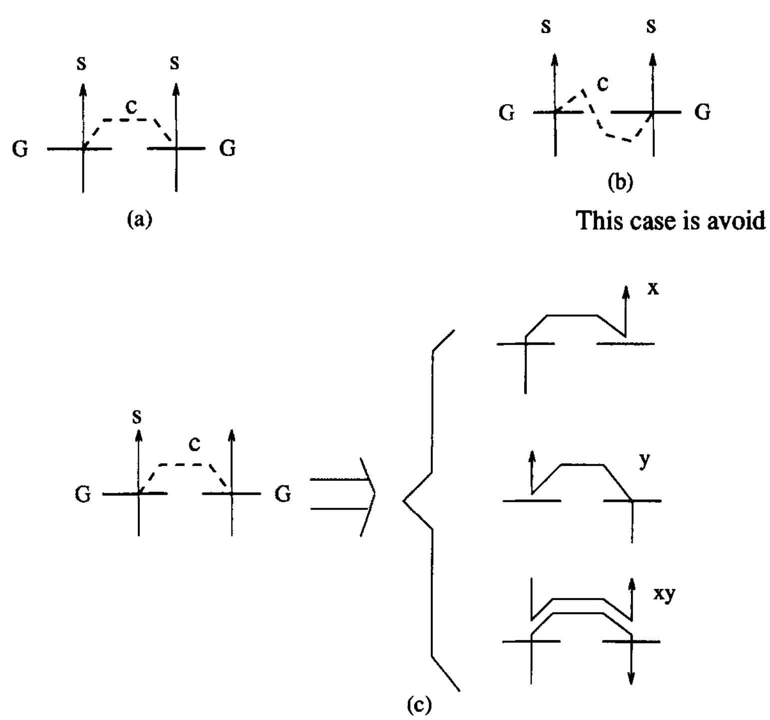

Fig. 5.

2.6. *Orientations.* Although curves are regarded as unoriented in the paper, it is convenient to orient them arbitrarily in the proof so that one may compare the intersection signs easily. To this end, one adopts the following convention in the paper. If p is oriented, we orient x and y so that they induce the same orientations on p_1 and p_2 (with respect to the orientation induced from p). Any orientation on xy will induce orientations on p_1, p_2, c_1 and c_2 so that exactly one of the orientations on p_1 and p_2 is reversed (with respect to the orientation induced from p) and the orientations on c_1 and c_2 are the same with respect to the parallel. We will orient p and xy arbitrarily.

2.7. *Marking and torus with holes.* Given a compact oriented surface Σ of genus g ($g \geqslant 1$) so that Σ is not a torus nor a one-holed torus, choose a maximal collection of simple closed curves a_i and b_i for $i = 1, 2, \ldots, g$ on Σ so that $a_i \perp b_i$ and $a_i \cap a_j = a_i \cap b_j = b_i \cap b_j = \emptyset$ for $i \neq j$. For each index i, let c_i be a simple closed curve so that $c_i \perp a_i, c_i \perp b_i, c_i \cap a_i \cap b_i \neq \emptyset$, and c_i is not isotopic into a regular neighborhood of $a_i \cup b_i$. Our goal is to show the following crucial proposition.

2.8. PROPOSITION. *Let $\mathscr{X}$ be the subset $\{\alpha \in \mathscr{S} \mid$ there is an index i so that either $I(\alpha, [a_i]) = I(\alpha, [b_i]) = 0$, or $0 \leqslant I(\alpha, [a_i]), I(\alpha, [b_i]), I(\alpha, [c_i]) \leqslant 1\}$. Then $\mathscr{X}^\infty = \mathscr{S}$.*

The proposition will be proved by inductions on a norm $|\alpha| = \sum_{i=1}^{g}(I(\alpha, [a_i]) + I(\alpha, [b_i]))$ of α in $\mathscr{S}$ and a semi-norm $\| \alpha \|_i = I(\alpha, [a_i]) + I(\alpha, [b_i]) + I(\alpha, [c_i])$ indexed by i. Since the norms involve the geometric intersection numbers, we begin with a discussion of intersection of arcs with $a_i \cup b_i$ and with $a_i \cup b_i \cup c_i$.

Given two simple closed curves a and b with $a \perp b$, a regular neighborhood $N(a \cup b)$ of $a \cup b$ is a one-holed torus. We will represent the figure eight $a \cup b$ as in Fig. 6(a) and its regular neighborhood $N(a \cup b)$ as in Fig. 6(b). The intersection of a simple closed curve with $a \cup b$ will be represented as in Fig. 6(c) where all intersection points are drawn to be near $a \cap b$.

Finally, the intersection as in Fig. 6(d) may be changed to the one in Fig. 6(e) by an isotopy which moves the intersection points along the open arcs a–b or b–a. The graph $a \cup b$ is said to be a *standard spine* of the one-holed torus.

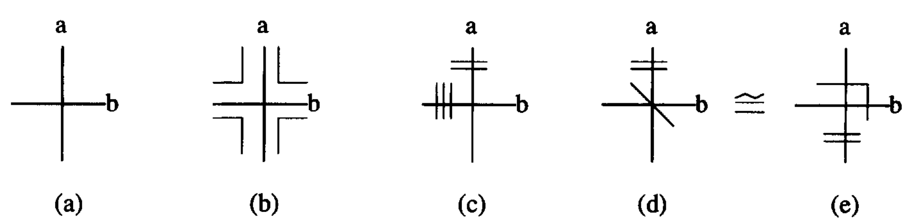

Fig. 6.

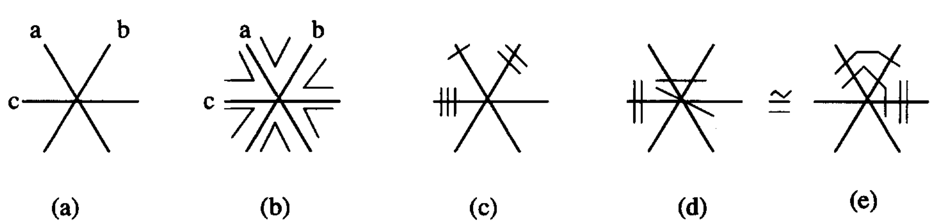

Fig. 7.

If a, b and c are three simple closed curves so that $a \perp b$, $b \perp c$, $c \perp a$, $a \cap b \cap c \neq \emptyset$, and c is not isotopic into a regular neighborhood of $a \cup b$, then a regular neighborhood $N(a \cup b \cup c)$ of $a \cup b \cup c$ is a two-holed torus. The figure $a \cup b \cup c$ will be represented as in Fig. 7(a), and the regular neighborhood $N(a \cup b \cup c)$ will be represented as in Fig. 7(b).

The intersection of the curve s with $a \cup b \cup c$ will be represented as in Fig. 7(c). We may also change the intersection $s \cap (a \cup b \cup c)$ as in Fig. 7(d) by an isotopy which moves the intersection points along the open arcs a–b, b–c or c–a as in Fig. 7(e). The graph $a \cup b \cup c$ is said to be a *standard spine* of the two-holed torus $N(a \cup b \cup c)$.

2.9. We begin the proof of Proposition 2.8 by showing the following lemma.

2.10. LEMMA. *Let $\mathscr{Y}$ be the subset $\{\alpha \in \mathscr{S} \mid$ there is an index i so that $0 \leqslant I(\alpha, [a_i])$ and $I(\alpha, [b_i] \leqslant 1\}$. Then $\mathscr{Y}^\infty = \mathscr{S}$.*

Proof. Given α in $\mathscr{S}$ of norm $|\alpha| = \sum_{i=1}^{g} (I(\alpha, [a_i]) + I(\alpha, [b_i])) = m$, we will use induction on m to show that $\alpha \in \mathscr{Y}^\infty$. Choose a representative s in α, so that $|s \cap a_i| = I(\alpha, [a_i])$, and $|s \cap b_i| = I(\alpha, [b_i])$ for all indices i. In particular, the norm $|\alpha| = \sum_{i=1}^{g} (|s \cap a_i| + |s \cap b_i|)$. We will assume that s avoids all the intersection points of a_i and b_i. Since s is non-separating, there is an index i so that one of the numbers $|s \cap a_i|$ or $|s \cap b_i|$ is an odd number. Thus $|\alpha| \geqslant 1$. If the equality holds, then α is in $\mathscr{Y}$. Suppose now that the result holds for all α is $\mathscr{S}$ of norm less than or equal to $m - 1$. Given $\alpha = [s]$ of norm m, let us assume without loss of generality that either $|s \cap a_1|$ or $|s \cap b_1|$ is an odd number.

REDUCTION I. *If there is an adjacent cutting arc c for s in $a_i - b_i$, or in $b_i - a_i$, then α is in $\mathscr{Y}^\infty$.*

Indeed, let x, y and xy be the curves spanned by s and c. By applying Lemma 2.5 to the graph $G = \bigcup_{i=1}^{g} (a_i \cup b_i)$, we have that the norms of $[x]$, $[y]$ and $[xy]$ are all less than m. Thus $[x]$, $[y]$ and $[xy]$ are in $\mathscr{Y}^\infty$ by the induction hypothesis. Since $[s]$ is spanned by $[x]$, $[y]$ and $[xy]$, $[s]$ is in $\mathscr{Y}^\infty$.

We assume in the following that no such adjacent cutting arcs exist for s, i.e. s is *irreducible* with respect to Reduction I.

The proof is given in Figs 8–11. It consists of simplifying a chain of diagrams. Each diagram represents a case of the simplification process and is indexed by an ordered finite sequence of numbers. The following notations will be used in the figures.

$\Rightarrow$	the H-operation on the dotted cutting arc and the curve
$\{$	breaking into cases
$\cong$	isotopy of the curve in the diagram
AP (123)	by applying the alternating principle to compare the curve in the current case and the curve in case (123), we conclude that one of the two curves is reducible. This ends the proof
AP	the same as AP (123) when case (123) is the preceding diagram
IND	this ends the proof by the induction hypothesis
Goto (123)	go to case (123)
SAA (123)	the same argument used in reducing case (123) applies, and this ends the proof
RED	Reduction I applies and this ends the proof by the induction hypothesis

The basic reduction process to simplify a simple closed curve p is the H-operation on a cutting arc c for p. Take the graph G to be $\bigcup_{i=1}^{g}(a_i \cup b_i)$. The cutting arc c is chosen so that $c \cap G = \partial c$ and a parallel copy c' of c with end points in p is disjoint from G. Let x, y, and xy be the curves spanned by p and c. Then by Lemma 2.5, the norms of $[x]$ and $[y]$ are less than that of $[p]$ and the norm of $[xy]$ is at most that of $[p]$. By the induction hypothesis, both $[x]$ and $[y]$ are in $\mathscr{Y}^\infty$. Thus showing that $[p] \in \mathscr{Y}^\infty$ is equivalent to showing $[xy] \in \mathscr{Y}^\infty$. Due to this, we will only draw the third curve xy in the diagram following the H-operation. We label the curves in the diagrams according to the following rule: (1) isotopic curves will be labeled by the same latter; (2) the three curves spanned by the curve s and a cutting arc c are labelled by x, y, and xy; and the three curves spanned by $x^i y^i$ and a cutting arc are labelled by x^{i+1}, y^{i+1}, and $x^{i+1}y^{i+1}$ where $x^0 = x$, $y^0 = y$. We also assume that at each (non-final stage) diagram, the curve involved are irreducible with respect to Reduction I since otherwise Reduction I applies and the proof is finished. When we apply the alternating principle to two diagrams, the outmost intersection points are marked in the diagram.

Case 1. $|s \cap a_1|$ or $|s \cap b_1| = 1$. We will simplify s in the one-holed torus $N(a_1 \cup b_1)$ by reducing the intersection number $|s \cap a_1| + |s \cup b_1|$. Let us assume without loss of generality that $|s \cap a_1| = 1$. If $|s \cap b_1| \leqslant 1$, then $[s] \in \mathscr{Y}$. If $|s \cap b_1| \geqslant 2$, we simplify s according to the following chain of diagrams (see Fig. 8).

Here are the details. We represent the figure $a_1 \cup b_1$ according to the scheme in Fig. 6(a). In case (1), the three intersection points in $s \cap (a_1 \cup b_1)$ adjacent to $a_1 \cap b_1$ are shown in the diagram. Case (1) breaks into two subcases (11) and (12) according to the intersection signs as indicated. In case (12), we move $s \cap a_1$ to the other side of b_1 along the open arc $a_1 - b_1$ by an isotopy. This can be achieved since $s \cap a_1$ consists of only one point by the assumption. Thus case (121) is covered in case (11). The reduction from the (11) to (111) is an H-operation on the indicated cutting arc c. By the remark above, among the three curves x, y and xy spanned by s and c, $[x]$, $[y]$ are in $\mathscr{Y}^\infty$. The third curve xy is drawn in case (111) since its norm could reach m. Assume the norm of $[xy]$ is m and that xy is irreducible, since otherwise we are done by either the induction hypothesis or Reduction I. We proceed arc c^1 as indicated and use the H-operation on c^1 and xy. The third curve x^1y^1 spanned by c^1 and xy is shown in (1^5) and x^1y^1 is isotopic to the curve in (1^6) which has at most $m-1$ intersection points with the graph G.

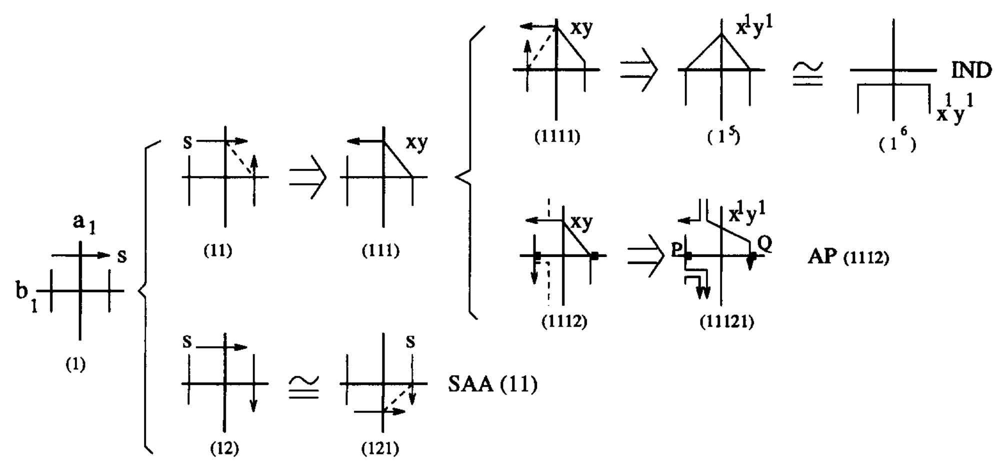

Fig. 8.

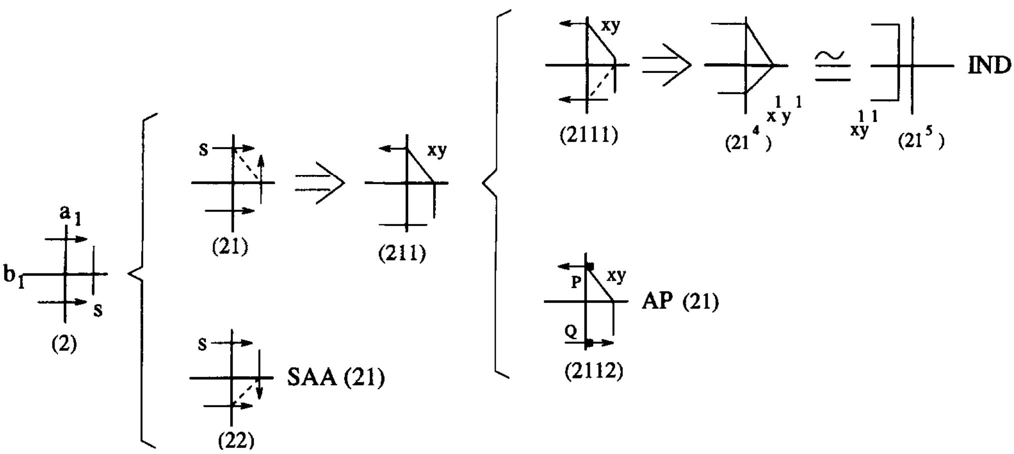

Fig. 9. Case (2) subcase 2.1.

Thus by the induction hypothesis, $[x^1y^1]$ is in $\mathscr{Y}^\infty$. This in turn implies that $\alpha = [s]$ is in $\mathscr{Y}^\infty$. In the other case (1112), we find a cutting arc c^1 as indicated (c^1 is parallel to an arc in a_1) and use the H-operation on the cutting arc c^1 and the curve xy. The resulting third curve x^1y^1 is drawn in (11121). We apply the alternating principle to x^1y^1 and b_1–a_1 in case (11121) and xy and b_1–a_1 in case (1112). The intersection signs of P and Q in $x^1y^1 \cap b_1$ are the same, whereas the intersection signs of P and Q in $xy \cap b_1$ are different. Since xy is irreducible, and $xy \cap b_1 = x^1y^1 \cap b_1$, by the alternating principle, there is an adjacent cutting arc for x^1y^1 in b_1–a_1. Thus, by Reduction I, $[x^1y^1]$ is in $\mathscr{Y}^\infty$. This finishes the proof of case (1).

Case 2. Neither $|s \cap a_1|$ nor $|s \cap b_1|$ is one. We will simplify s again in the one-holed torus $N(a_1 \cup b_1)$. Since one of $|s \cap a_1|$ or $|s \cap b_1|$ is an odd number, let us assume without loss of generality that $|s \cap a_1|$ is an odd number bigger than or equal to three. This implies that there is an adjacent cutting arc for s in a_1 by a simple parity checking. Because the curve s is irreducible, the cutting arc c must intersect b_1, i.e. the two intersection points of $s \cap a_1$ which are adjacent to $a_1 \cap b_1$ have the same intersection signs.

Subcase 2.1. $|s \cap b_1| \geqslant 2$. We reduce s according to the chain of diagrams as in Fig. 9.

Here are the details. Case (2) breaks into two subcases (21) and (22) according to the intersection signs of $s \cap b_1$ at the point near $a_1 \cap b_1$. In case (22), the same argument used in reducing case (21) applies. Thus, it suffices to consider case (21). Find a cutting arc c as indicated in (21) and use the H-operation on s and c. The third curve xy spanned by c and s indicated in (211) consists of two subcases (2111) and (2112) according to the intersection sign. In case (2112), we compare the curve xy and a_1–b_1 with the curve s and a_1–b_1 in case (21). Since s irreducible and $s \cap a_1 = xy \cap a_1$ and the intersection signs of P and Q in $xy \cap a_1$ are different, by the alternating principle, there is an adjacent cutting arc for xy in a_1–b_1. Thus, xy is reducible with respect to Reduction I and $[xy]$ is in $\mathscr{Y}^\infty$. In case (2111), we find a cutting arc c^1 as indicated in (2111) and use H-operation on xy and c^1. The three curves spanned by xy and c^1 all have norm less than m by Lemma 2.5 and by an isotopy which reduces case (21^4) to case (21^5). Thus, by the induction hypothesis, the isotopy classes of all of them are in $\mathscr{Y}^\infty$. This in turn implies that α is in $\mathscr{Y}^\infty$.

Subcase 2.2. $|s \cap b_1| = 0$. Let x, y and xy be the curves spanned by s and c as indicated in Fig. 10. Then the norms of $[x]$, $[y]$ and $[xy]$ are at most m. Since $|x \cap b_1| = |y \cap b_1| = 1$, by case 1 or by the induction hypothesis if the norm is less than m, $[x]$ and $[y]$ are in $\mathscr{Y}^\infty$. If the

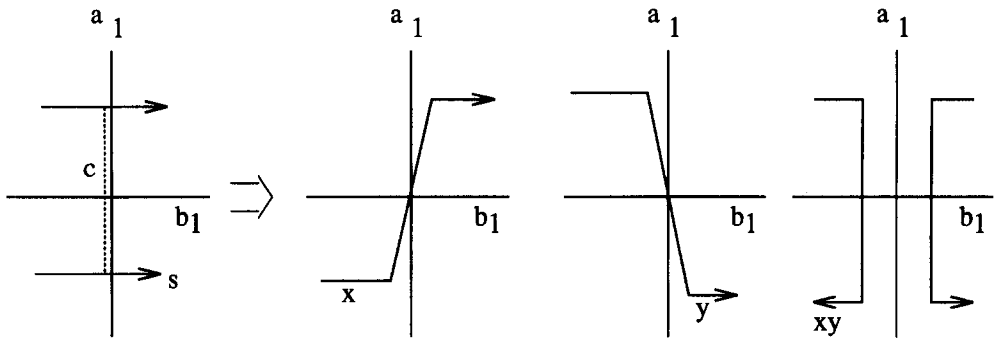

Fig. 10. The H-operation on the arc c.

norm of $[xy]$ is less than m, we are done. If the norm of $[xy]$ is m, since $|xy \cap a_1| - |s \cap a_1|$ is an even number, the curve xy satisfies that $|xy \cap a_1|$ is an odd number and $|xy \cap b_1| = 2$. By case 2.1 or case 1, $[xy]$ is in $\mathscr{Y}^\infty$. This finishes the proof of Lemma 2.10 by induction.

2.11. For each index i, let $\mathscr{Y}_i = \{\alpha \in \mathscr{S} \mid 0 \leqslant I(\alpha, [a_i]),\ I(\alpha, [b_i]) \leqslant 1\}$, and let $\mathscr{X}_i = \{\alpha \in \mathscr{S} \mid \text{either } I(\alpha, [a_i]) = I(\alpha, [b_i]) = 0, \text{ or } 0 \leqslant I(\alpha, [a_i]),\ I(\alpha, [b_i]),\ I(\alpha, [c_i]) \leqslant 1\}$. Clearly by the definition, $\mathscr{X} = \bigcup_{i=1}^{g} \mathscr{X}_i$ and $\mathscr{Y} = \bigcup_{i=1}^{g} \mathscr{Y}_i$. We will show below that $\mathscr{Y}_i \subset \mathscr{X}_i^\infty$. Thus, $\mathscr{Y} \subset \mathscr{X}^\infty$. It follows from Lemma 2.10 that $\mathscr{S} \subset \mathscr{Y}^\infty$, and Proposition 2.8 follows.

2.12. LEMMA. *For each index i,* $\mathscr{Y}_i \subset \mathscr{X}_i^\infty$.

Proof. Let us assume without loss of generality that the index $i = 1$. Consider the graph $G = a_1 \cup b_1 \cup c_1$ and the norm $\| \alpha \| = I(\alpha, [a_1]) + I(\alpha, [b_1]) + I(\alpha, [c_1])$ for α is $\mathscr{S}$. We will also refer $\| \alpha \|$ as the norm of s when $s \in \alpha$. We will use induction on the norm $m = \| \alpha \|$ to prove the lemma. Clearly if $m = 1$, then α is in $\mathscr{X}_1^\infty$.

Suppose the result holds for all $\alpha \in \mathscr{Y}_1$ of the norm $\| \alpha \|$ at most $m - 1$. Given α of the norm m, choose $s \in \alpha$ so that $|s \cap a_1| = I(\alpha, [\alpha_1])$, $|s \cap b_1| = I(\alpha, [b_1])$ and $|s \cap c_1| = I(\alpha, [c_1])$. In particular, $\| \alpha \| = |s \cap a_1| + |s \cap b_1| + |s \cap c_1|$. We will reduce s according to $(|s \cap a_1|, |s \cap b_1|) = (0, 0)$, $(1, 0)$, $(0, 1)$ or $(1, 1)$. If it is $(0, 0)$, then by definition α is in $\mathscr{X}_1^\infty$. Assume also that $|s \cap c_1| \geqslant 2$ since otherwise α is in $\mathscr{X}_1$.

The curve s will be simplified in the two-holed torus $N(a_1 \cup b_1 \cup c_1)$ by reducing the intersection number $|s \cap (a_1 \cup b_1 \cup c_1)|$ through a sequence of H-operations and isotopies. To begin the proof, we assume that the curve s is irreducible with respect to Reduction I, i.e. there is no cutting arc c for s in a_1–b_1, or b_1–c_1, or c_1–a_1. Otherwise, the H-operation on s and the cutting arc will produce three curves x, y, and xy of norm less than m by Lemma 2.5. Clearly $[x]$, $[y]$ and $[xy]$ are all in $\mathscr{Y}_1$. Thus by the induction hypothesis, $[x]$, $[y]$ and $[xy]$ are all in $\mathscr{X}_1^\infty$.

The reduction process below consists of simplifying a chain of diagrams in Figs 11–15. The notations introduced in the proof of Lemma 2.10 will be used again. The graph $G = a_1 \cup b_1 \cup c_1$ will be represented according to the scheme in Fig. 7(a). Only the intersection points of $s \cap G$ adjacent to $a_1 \cap b_1 \cap c_1$ are shown in each diagram. The cutting arcs are the dotted lines in the diagrams. We divide the proof into two cases where case (1) corresponds to $(|s \cap a_1|, |a \cap b_1|) = (1, 1)$ and case (2) corresponds to $(|s \cap a_1|, |s \cap b_1| = (1, 0)$ or $(0, 1)$.

Case (1). $(|s \cap a_1|, |s \cap b_1|) = (1, 1)$. This case is covered in Figs 11 and 12.

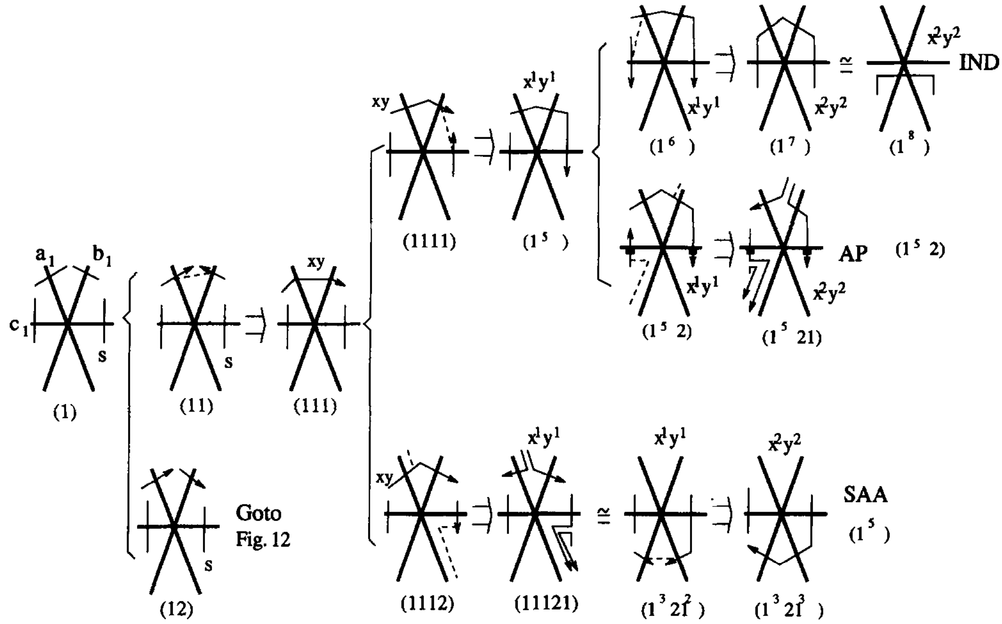

Fig. 11. Case (1).

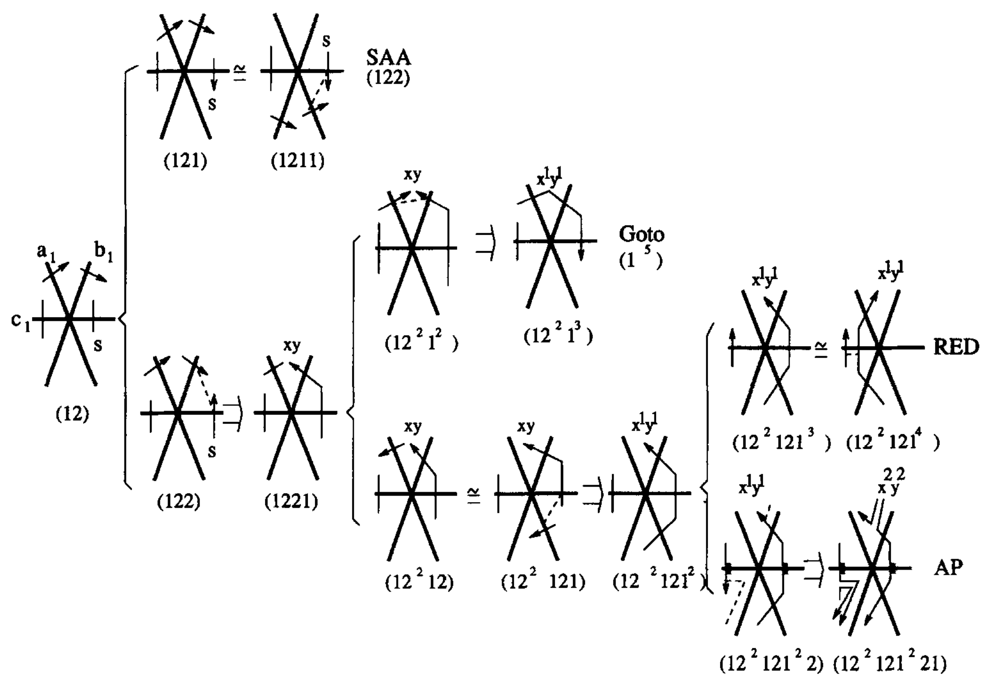

Fig. 12. Case (12).

Here are the details of Fig. 11. Case (1) breaks into two cases (11) and (12) according to the intersection signs of $s \cap a_1$ and $s \cap b_1$ at the points adjacent to $a_1 \cap b_1$. The second case (12) will be discussed later in Fig. 12. In case (11), one finds a cutting arc c as indicated and uses the H-operation on the curve s and the cutting arc c to obtain the third curve xy spanned by s and

c in case (111). Case (111) consists of two subcases (1111) and (1112) according to the intersection signs. In case (1111), we find a cutting arc c^1 as indicated and use the H-operation on the cutting arc c^1 and the curve xy to obtain the curve x^1y^1 in case (1^5). Case (1^5) breaks into two subcases according to the intersection signs.

In case (1^6), one finds a cutting arc as indicated and uses the H-operation on the cutting arc and the curve to obtain case (1^7). The resulting third curve x^2y^2 in (1^7) is isotopic to one as indicated in the diagram (1^8). Thus, the norm of $[x^2y^2]$ is less than m. By the induction hypothesis, $[x^2y^2]$ is in $\mathscr{X}_1^\infty$. In case ($1^5$2), assuming that the curve x^1y^1 has the norm m and is irreducible with respect to the Reduction I, one finds a cutting arc as indicated and uses the H-operation on the cutting arc and the curve to obtain ($1^5$21). The resulting third curve x^2y^2 in ($1^5$21) is reducible with respect to Reduction I by the alternating principle. Indeed, applying the alternating principle to the curve x^1y^1 in case ($1^5$2) and x^2y^2 in ($1^5$21) with respect to the arc c_1–a_1, one finds an adjacent cutting arc for x^2y^2 in c_1–a_1. Thus $[x^2y^2]$ is in $\mathscr{X}_1^\infty$ by Reduction I. This implies that α is in $\mathscr{X}_1^\infty$ in case (1111). Lastly, in case (1112), one finds a cutting arc as indicated and uses the H-operation on the cutting arc and the curve to obtain case (11121). In case (11121), since the curve x^1y^1 intersects both a_1 and b_1 each at one point, it can be isotopic into a position as in ($1^3$2$1^2$). Find a cutting arc in ($1^3$2$1^2$) as indicated and use the H-operation on the cutting arc and the curve to obtain the third curve x^2y^2 in ($1^3$2$1^3$). Case ($1^3$2$1^3$) is already covered in case (1^5). Thus, we have finished the simplification of case (11).

We now come to case (12) as in Fig. 12. Case (12) breaks into two cases according to the intersection signs. In case (121), the curve s is isotopic into the position as in (1211) since $|s \cap a_1| = |s \cap b_1| = 1$. Case (1211) is the same as in case (122). Thus, it suffices to consider case (122). In case (122), we find a cutting arc as indicated and use the H-operation to obtain case (1221). There are two subcases (1$2^2$1^2) and (1$2^2$12) according to the intersection signs. If we are in case (1$2^2$1^2), we find a cutting arc as indicated and use the H-operation to transform (1$2^2$1^2) into (1$2^2$1^3). Case (1$2^2$1^3) is already covered in case (1^5) in Fig. 11. In case (1$2^2$12), the curve xy moved by an isotopy to the position as in (1$2^2$121) since $xy \cap a_1$ consists of one point. One finds a cutting arc as indicated in (1$2^2$121) and uses the H-operation to obtain case (1$2^2$12$1^2$). There are now two cases depending on the intersection signs. If we are in case (1$2^2$12$1^3$), then an isotopy of the curve x^1y^1 as indicated transforms it into (1$2^2$12$1^4$) without increasing the norm. Case (1$2^2$12$1^4$) is reducible. Thus, by the Reduction I, $[x^1y^1]$ is in $\mathscr{X}_1^\infty$. This implies that α is in $\mathscr{X}_1^\infty$. If we are in case (1$2^2$12$1^2$2) and the curve x^1y^1 is irreducible, we find a cutting arc as indicated and use the H-operation to obtain case (1$2^2$12$1^2$21). Now the two curves x^1y^1 in (1$2^2$12$1^2$2) and x^2y^2 in (1$2^2$12$1^2$21) satisfy the condition in the alternating principle in the arc $c_1 - a_1$. Thus, x^2y^2 is reducible with respect to Reduction I. Thus the result follows. This finishes the proof of case (1).

Case (2). $(|s \cap a_1|, |s \cap b_1|) = (1, 0)$ or $(0, 1)$. Let us assume without loss of generality that it is $(1, 0)$. We reduce s according to the schemes in Figs 13–15.

Let us consider Fig. 13 first.

Here are the details of Fig. 13. Case (2) breaks into two subcases according to the intersection signs. Starting from case (21), one follows the routine reduction by finding cutting arcs as indicated in (21), (2111) and (21121) and using the H-operation to reduce the curves involved. Case (21111) and its preceding case (2111) satisfy the conditions of the alternating principle at the points marked. Thus, by Reduction I, we are done with case (21111). In case (2112), the curve xy is isotopic into the other side of b_1 as in Fig. (21121). This increases the intersection number of the curve xy with the graph G by two. Using the H-operation on the cutting arc in (21121), we obtain three curves x^1, y^1 and x^1y^1 as in (211211), (211212) and (211213). Due to the increase of the intersection number $|xy \cap G|$, one has to make sure that

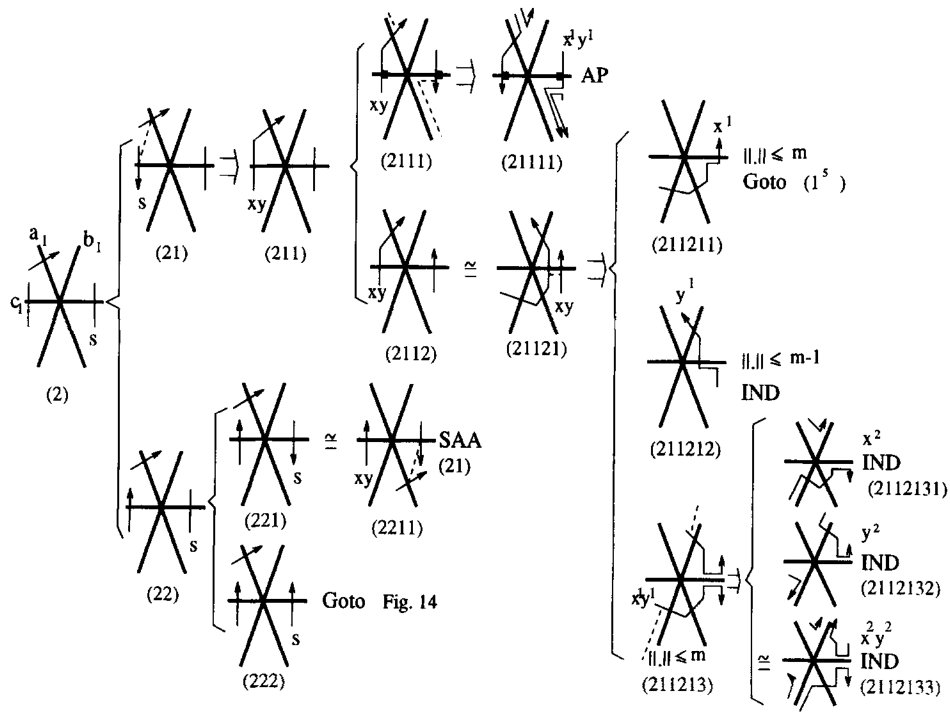

Fig. 13. Case (2).

all the isotopy classes of x^1, y^1, and x^1y^1 are in $\mathscr{X}_1^\infty$. In case (211211), the curve x^1 intersects the graph G in at most m points and $[x^1]$ is in $\mathscr{Y}_1$. If the norm of $[x^1]$ is m, then this case is already covered in case (1^5); if the norm is less than m, $[x^1]$ is in $\mathscr{X}_1^\infty$ by the induction hypothesis. In case (211212), the number of intersection points of y^1 with G is at most $m-1$ and both $|y^1 \cap a_1|$ and $|y^1 \cap b_1|$ are at most one, thus by the induction hypothesis, the isotopy class of the curve is in $\mathscr{X}_1^\infty$. In case (211213), the curve x^1y^1 intersects the graph G in at most m points, x^1y^1 intersects a_1 in one point, and x^1y^1 intersects b_1 at two points of the same intersection signs. Thus Reduction I applies to x^1y^1 at the adjacent cutting arc in b_1–a_1 as indicated in (211213). By Lemma 2.5, the isotopy classes of the three curves x^2, y^2, x^2y^2 spanned by x^1y^1 and the cutting arc have the norms at most $m-1$ and furthermore, they are all in $\mathscr{Y}_1$. Thus by the induction hypothesis, the isotopy classes $[x^2]$, $[y^2]$ and $[x^2y^2]$ are all in $\mathscr{X}_1^\infty$. Now that $[x^1]$, $[y^1]$ and $[x^1y^1]$ are in $\mathscr{X}_1^\infty$, this in turn implies that α is in $\mathscr{X}_1^\infty$. This takes care of case (21). In case (22), we obtain two subcases (221) and (222) according to the intersection signs. In case (221), the curve s is isotopic into the position as in (2211) since $|s \cap a_1| = 1$. The same type of reduction used in case (21) applies to (2211). Thus α is in $\mathscr{X}_1^\infty$.

We consider the most difficult case (222) in Figs 14 and 15 below.

Here are the details of Fig. 14. Find a cutting arc c for s which cuts the graph G in one point as indicated in (222). We obtain three curves x, xy and y spanned by s and c as in (2221), (2^4) and (2223), respectively. We need to discuss each of the three curves $[x]$, $[y]$ and $[xy]$ separately since the norms of $[x]$ and $[y]$ may not be less than m and the curve $[xy]$ is not in the set $\mathscr{Y}_1$. In cases (2221) and (2223), the dotted arrowed line segments indicate that these line segments may be part of the curves x and y. In case (2221), the curve x intersects each of a_1 and b_1 in one point, and x intersects the graph G in at most m points. Thus, if the norm of $[x]$ is less than m, by the induction hypothesis $[x]$ is in $\mathscr{X}_1^\infty$; if the norm is m, this is already

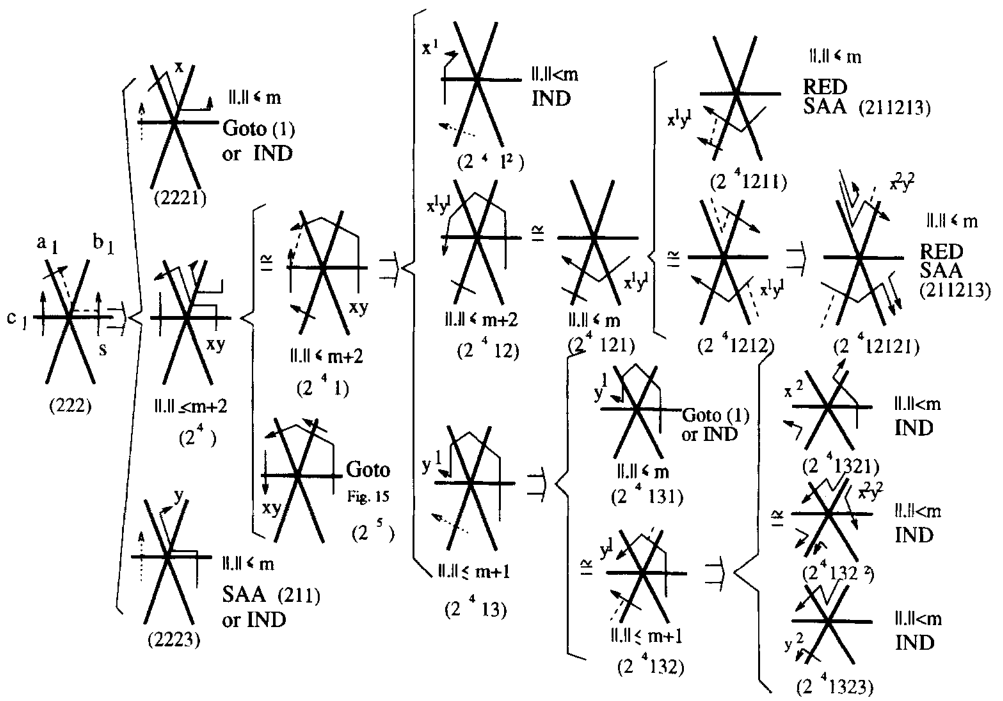

Fig. 14. Case (222).

covered in case (1). Thus $[x]$ is in $\mathscr{X}_1^\infty$. In the case (2223), the curve y intersects b_1 in one point, y does not intersect a_1, and y intersects G in at most m points. If the norm of $[y]$ is less than m, by the induction hypothesis, we are done. If the norm of $[y]$ is m, then the same reduction argument used in (211) applied to y. Thus, $[y]$ is also in $\mathscr{X}_1^\infty$.

This shows that α is in $\mathscr{X}_1^\infty$ if and only if $[xy]$ in case (2^4) is in $\mathscr{X}_1^\infty$. Let us now proceed from (2^4). The curve xy in (2^4) intersects G in at most $m + 2$ points, intersects a_1 in one point, and intersects b_1 in two points of the same intersection signs. There are now two cases (2^41) and (2^5) according to the intersection signs. Case (2^5) will be covered in Fig. 15. In case (2^41), the curve xy is moved by an isotopy to the position indicated. Find a cutting arc c^1 as indicated in (2^41) and use the H-operation on the cutting arc c^1 and the curve xy. We obtain three curves x^1, y^1, and x^1y^1 spanned by c^1 and xy as indicated in cases (2^41^1), (2^413) and (2^412), respectively. In case (2^41^2), the curve x^1 intersects G in at most $m - 1$ points, and x^1 intersects each of a_1 and b_1 in at most one point. Thus by the induction hypothesis, $[x^1]$ is in $\mathscr{X}_1^\infty$. In case (2^412), the curve x^1y^1 can be moved by an isotopy to the position as in (2^4121). The curve x^1y^1 in (2^4121) intersects G in at most m points, intersects b_1 in two points, and a_1 in one point. There are two subcases according to the intersection signs as indicated in (2^4121^2) and (2^41212). In the case (2^4121^2), there is an adjacent cutting arc for the curve x^1y^1 in b_1–a_1. Thus Reduction I applies and we obtain three curves whose isotopy classes are all in $\mathscr{X}_1^\infty$ by the same argument that we use to treat case (211213). Thus we are done with case (2^4121^2). In case (2^41212), we find a cutting arc as indicated. The H-operation on the cutting arc and x^1y^1 produces the curve x^2y^2 in (2^412121). The norm of x^2y^2 is at most m and there is an adjacent cutting arc for it in b_1–a_1. Thus the result follows by the same argument that we use in case (2^4121^2). In case (2^413), $\| y \| \leqslant m + 1$, $| y \cap b_1 | \leqslant 2$, and $| y \cap a_1 | \leqslant 1$. We break case (2^413) into two cases according to whether the dotted arrowed line segment is in y^1 or not. In case (2^4131) the dotted arrowed line segment is not in y^1, the isotopy class of the curve y^1 is in $\mathscr{Y}_1$ and it has either the norm less than m or it has the norm m but $[y^1]$ is in the case (1). In

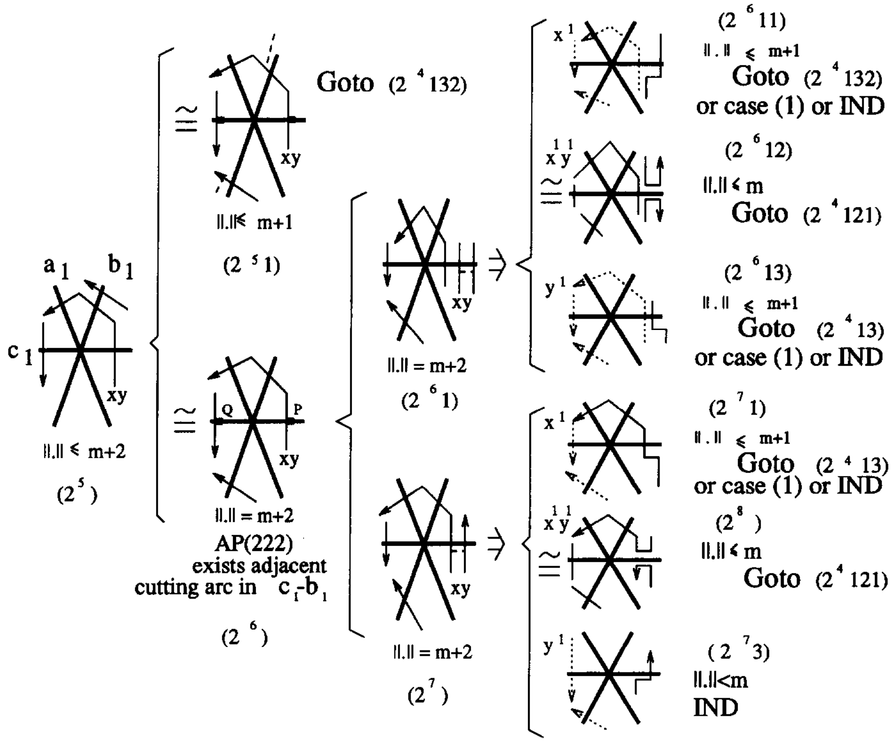

Fig. 15. Case (2^5).

both situations, $[y^1]$ is in $\mathscr{X}_1^\infty$ by either the induction hypothesis or by the proof of case (1). In case $(2^4 132)$, we find a cutting arc in b_1–a_1 as indicated and use the H-operation on the cutting arc and y^1 to obtain three curves x^2, y^2 and x^2y^2 in cases $(2^4 1321)$, $(2^4 1323)$, and $(2^4 132^2)$, respectively. Clearly all isotopy classes of x^2, y^2 and x^2y^2 are in $\mathscr{Y}_1$ and their norms are less than m. Thus, by the induction hypothesis, all of them are in $\mathscr{X}_1^\infty$. This finishes the proof that α is in $\mathscr{X}_1^\infty$ in case $(2^4 1)$.

We now consider the last case (2^5) in Fig. 15.

Here are the details. We first break (2^5) into two cases according to $\|[xy]\| \leqslant m+1$ or $\|[xy]\| = m+2$. Case $(2^5 1)$ is already covered in $(2^4 132)$. In case (2^6), compare the curves xy in (2^6) with the curve s in (222). Since $xy \cap (c_1–a_1) = s \cap (c_1–a_1)$, and the intersection signs at P and Q of $xy \cap (c_1–a_1)$ are different, by the alternating principle, there is an adjacent cutting arc for xy in c_1–a_1. Case (2^6) is divided into two cases according to whether the adjacent cutting arc contains the point P or not. If it does not contain P, we are in $(2^6 1)$. Using the H-operation at the adjacent cutting arc and xy, we obtain three new curves x^1, y^1 and x^1y^1 as in $(2^6 1^2)$, $(2^6 13)$ and $(2^6 12)$, respectively. The curve x^1 in $(2^6 1^2)$ intersects G in at most $m+1$ points, and x^1y^1 is either in case $(2^4 13)$, or in case (1), or having the norm less than m. Thus, $[x^1]$ is in $\mathscr{X}_1^\infty$. The situation for y^1 is the same as x^1. Thus $[y^1]$ are in $\mathscr{X}_1^\infty$.

In case $(2^6 12)$, the curve x^1y^1 has norm at most m. This case is already covered in case $(2^4 121)$. Thus, $[x^1y^1]$ is in $\mathscr{X}_1^\infty$. This takes care of $(2^6 1)$. In the last case (2^7) that the adjacent cutting arc contains P, let x^1, y^1 and x^1y^1 be the curves spanned by xy and the adjacent cutting arc as in cases $(2^7 1)$, $(2^7 3)$, and (2^8), respectively. Case $(2^7 1)$ is already covered in $(2^4 13)$. Thus $[x^1]$ is in $\mathscr{X}_1^\infty$. In the case (2^8), the norm of $[x^1y^1]$ is at most m and x^1y^1 intersects a_1 in one point and b_1 in two points. Thus, it is already covered in case $(2^4 121)$. Thus $[x^1y^1]$ is in

$\mathscr{X}_1^\infty$. In the case ($2^7 3$), the norm of $[y^1]$ is less than m and $[y^1]$ is in $\mathscr{Y}_1$. Thus, by the induction hypothesis, $[y^1]$ is in $\mathscr{X}_1^\infty$.

This ends the proof of Lemma 2.12. Thus Proposition 2.8 is also proven.

2.13. *Curves in the two-holed torus.* The following lemma shows that the curves a_1, b_1 and c_1 forming a standard spine are unique up to self-homeomorphism of the two-holed torus $N(a_1 \cup b_1 \cup c_1)$.

2.14. LEMMA. *Suppose S is a two-holed torus with two standard spines $a_1 \cup b_1 \cup c_1$ and $a_2 \cup b_2 \cup c_2$. Then there is a self-homeomorphism of S sending a_1 to a_2, b_1 to b_2, and c_1 to c_2. In particular, there is a Z_3 action on S which permutes a_1, b_1 and c_1.*

Proof. Take a small regular neighborhood $N(a_i \cup b_i \cup c_i)$ of $a_i \cup b_i \cup c_i$ in S for $i = 1, 2$. There is a diffeomorphism f from $N(a_1 \cup b_1 \cup c_1)$ to $N(a_2 \cup b_2 \cup c_2)$ sending a_1 to a_2, b_1 to b_2, and c_1 to c_2, respectively. We can easily extend f to a diffeomorphism of S since by the assumption the complement $S - int(N(a_i \cup b_i \cup c_i))$ consists of two disjoint annuli for $i = 1, 2$.

To show the last assertion of the lemma, we consider the two-holed torus together with the standard spine drawn in Fig. 7(a). It has the required 3-fold symmetry induced by the $2\pi/3$-rotation in the plane. Furthermore, the 3-fold symmetry leaves each component of ∂S invariant.

2.15. PROPOSITION. *Suppose S is an oriented two-holed torus with a standard spine $a \cup b \cup c$. The order of a, b, and c are so chosen that $ab \cap c = bc \cap a = ca \cap b = \emptyset$ as in Fig.* 16.

Let p be a proper 1-submanifold in S so that p intersects each of the three curves a, b and c in at most one point and $p \cap (a \cup b \cup c) \neq \emptyset$. Then:

(a) *If p is a simple closed curve, p is isotopic to one of the nine curves a, b, c, ab, bc, ca, $\bar{a}\bar{b}$, $\bar{b}\bar{c}$, or $\bar{c}\bar{a}$ as in Fig.* 16. *Furthermore, $\mathscr{S}(S)$ is generated by $\{a, b, c, ab, bc, ca\}$;*

(b) *if p is a proper arc, then p is isotopic in S to one of the six proper arcs as in Fig.* 17 *or to one of their images under the 3-fold symmetry which permutes a, b, c.*

Proof. Take a small regular neighborhood $N = N(a \cup b \cup c)$ of the standard spine $a \cup b \cup c$ in S. The complement $S - int(N)$ consists of two annuli A_1 and A_2 as in Fig. 18(a). The boundary ∂A_i of A_i consists of a boundary component of ∂S and a boundary component n_i of ∂N for $i = 1, 2$. After an isotopy, we may assume that the intersection points of p with N are located exactly the same as intersection points of the model curves in Figs 16 and 17 (here we use the fact that N is a very small regular neighborhood) except when p is an arc and $|p \cap a| = |p \cap b| = |p \cap c| = 1$ with which we will deal later.

We will use the following easy fact about proper arcs in an annulus. The proof of the lemma is omitted.

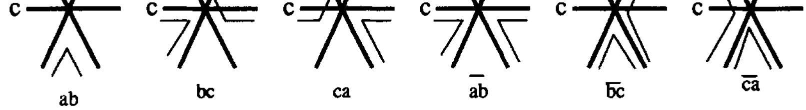

Fig. 16. In the figure, the orientation is the right-handed orientation in the plane.

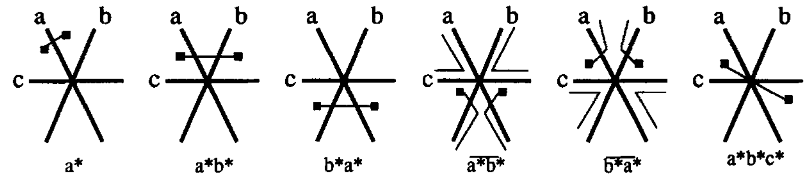

Fig. 17.

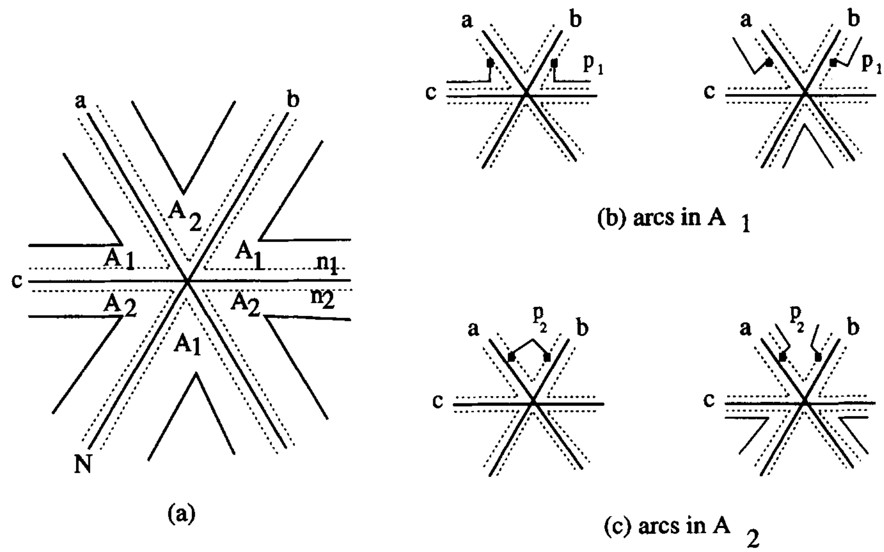

Fig. 18.

2.16. Lemma. *Suppose q is a proper arc in an annulus A with ∂A consisting of two components ∂_1 and ∂_2.*

(a) *If $|q \cap \partial_1| = 1$, then there is only one isotopy class of such arcs in A with respect to isotopies leaving ∂_1 pointwise fixed.*

(b) *If $|q \cap \partial_1| = 2$, then there are exactly two isotopy classes of such arcs in A with respect to isotopies leaving ∂_1 pointwise fixed. Furthermore, if $[q_1]$ and $[q_2]$ are the two distinct isotopy classes and q_3 is a proper arc joining ∂_1 to ∂_2, then $q_3 \cap (q_1 \cup q_2) \neq \emptyset$.*

We now continue the proof of the Proposition 2.15. The triple of geometric intersection numbers $(I(p, a), I(p, b), I(p, c))$ are given by $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(1,1,0)$, $(1,0,1)$, $(0,1,1)$, and $(1,1,1)$. By Lemma 2.14, it suffices to consider cases where $(I(p, a), I(p, b), I(p, c))$ is $(1,0,0)$, $(1,1,0)$ or $(1,1,1)$.

In the first case (a) that p is a simple closed curve, the triples $(1,0,0)$ and $(1,1,1)$ cannot occur by a simple Z_2-homology argument.

If $(I(p, a), I(p, b), I(p, c))$ is $(1,1,0)$, we claim that p is either isotopic to c or ab or $\bar{a}\bar{b}$. The intersection $p \cap N$ cuts p into two arcs p_1 and p_2 where p_i is a proper arc in A_i and $\partial p_i \subset n_i$. By Lemma 2.16 (b), there are exactly two isotopy classes of p_i in A_i with respect to isotopies leaving n_i pointwise fixed. These classes are listed in Fig. 18(b) and (c).

Combing these arcs in various ways, we obtain all four possible isotopy classes of p in N as in Fig. 19.

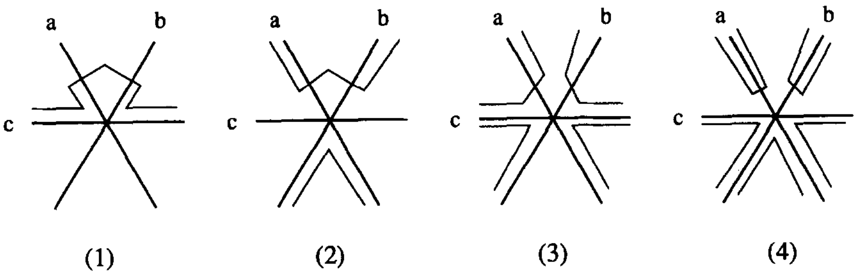

Fig. 19.

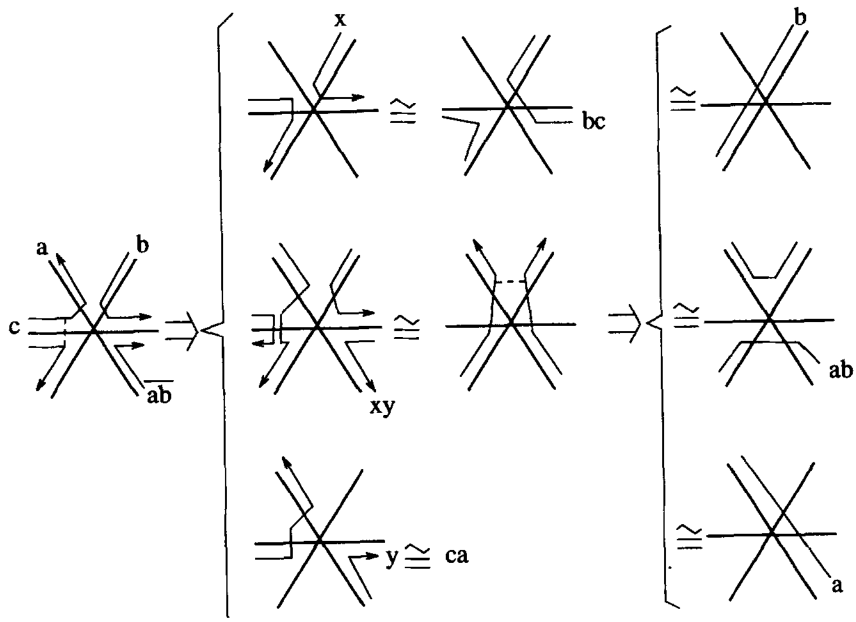

Fig. 20.

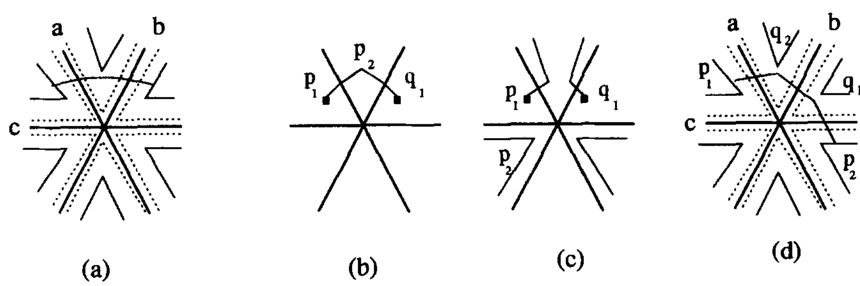

Fig. 21.

Now in case (1), $p \cong c$; in case (2), $p \cong ab$; in case (3), $p \cong \bar{a}\bar{b}$; and in the last case (4), $p \cong c$ by an isotopy which moves $p \cap a$ and $p \cap b$ to the position below c.

To show the last assertion in part (a), by Proposition 2.8, it suffices to prove that $\bar{a}\bar{b}$ is in $\{a, b, ab, bc, ca\}^2$. This in turn follows from the chains of diagrams in Fig. 20. This finishes the proof of case (a).

We now consider the case that p is a proper arc in S. There are three cases according to the triple $(|p \cap a|, |p \cap b|, |p \cap c|)$ being (1,0,0), (1,1,0), and (1,1,1).

In case $(|p\cap a|, |p\cap b|, |p\cap c|) = (1,0,0)$, the intersection $p\cap N$ cuts p into two arcs p_1 and p_2 with $p_i \subset A_i$ for $i = 1, 2$. Each curve p_i joints the two boundary components of ∂A_i. Hence by Lemma 2.16 part (a), there is only one isotopy class of p_i with respect to isotopies leaving n_i pointwise fixed. Thus p is unique up to isotopy in S. In particular, $p \cong a^*$ as in Fig. 17.

In the second case $(|p\cap a|, |p\cap b|, |p\cap c| = (1,1,0)$, there are two types of arcs according to $p\cap(\partial A_1 - n_1) \neq \emptyset$ or $p\cap(\partial A_2 - n_2) \neq \emptyset$. Let us assume without loss of generality that $p\cap(\partial A_1 - n_1) \neq \emptyset$. In this case, the intersection $p\cap N$ cuts p into three arcs p_1, q_1 and p_2 where p_1 and q_1 are proper arcs in A_1 joining the two components of ∂A_1 and p_2 is a proper arc in A_2 with $\partial p_2 \subset n_2$ as in Fig. 21. By Lemma 2.16, there is only one isotopy class of p_1 (and q_1 respectively) in A_1 with repect to isotopies leaving n_1 pointwise fixed and there are two isotopy classes for p_2 in A_2 with respect to isotopies leaving n_2 pointwise fixed. We now assemble these isotopy classes of p_1, q_1 and p_2 as in Fig. 21(b) and (c) to recover p in S. We obtain two isotopy classes of p in S, namely a^*b^* and $\bar{a}^*\bar{b}^*$ as in Fig. 17. To see that the two curves are not isotopic, we observe that $I([a^*b^*], [ab]) = 0$ and $I([\bar{a}^*\bar{b}^*], [ab]) = 2$.

In the final case that $(|p\cap a|, |p\cap b|, |p\cap c|) = (1, 1, 1)$, there are three possibilities according to the order of $p\cap a$, $p\cap b$ and $p\cap c$ in p. Let us assume without loss of generality that $p\cap b$ is between $p\cap a$ and $p\cap c$ in p. We also assume that the three intersection points $p\cap a$, $p\cap b$ and $p\cap c$ are fixed during the discussion. The arc p is cut by $p\cap N$ into four arcs p_1, q_1, p_2 and q_2 where p_i and q_i are proper arcs in A_i as in Fig. 21(d).

The arc p_i joins the two boundary components of ∂A_i and there is only one such isotopy class of p_i in A_i with respect to isotopies leaving n_i pointwise fixed. The end points of the arc q_i are in n_i and $q_i \cap p_i = \emptyset$. By Lemma 2.16, the disjointness of q_i from p_i forces q_i to be in only one isotopy class of such arcs with respect to isotopies leaving n_i pointwise fixed. Thus, there is at most one isotopy class of such p in S. The existence is demonstrated by Fig. 17.

2.17. Suppose $\Sigma = \Sigma_{g,n}$ is a compact oriented surface of positive genus g with n boundary components. We define $\| \Sigma \|$ to be $3g + n$ and use induction on $\| \Sigma \|$ to prove Theorem I for the surfaces in the class $\mathscr{L} = \{\Sigma_{g,n} \mid (g, n) \neq (2, 0), g \geqslant 1\}$. The remaining case that the surface is $\Sigma_{2,0}$ will follow easily.

To begin with, let us consider subsurfaces Σ' in Σ. By convention, all subsurfaces are assumed to be connected with the orientation induced from Σ. Given a subsurface $\Sigma' \subset \Sigma$, the inclusion map $i: \Sigma' \to \Sigma$ induces a map $i_*: \mathscr{S}(\Sigma') \to \mathscr{S}(\Sigma)$ by sending $[a]$ to $[i(a)]$. The induced map i_* preserves the disjointness, the orthogonality and the product. A subsurface Σ' in Σ is said to be *incompressible* if the inclusion map $i: \Sigma' \to \Sigma$ induces a monomorphism in the fundamental group. An equivalent definition of incompressibility is that each component of $\partial\Sigma'$ is an essential curve in Σ. If Σ' is incompressible in Σ, and $i: \Sigma' \to \Sigma$ is not a homotopy equivalence, then $\| \Sigma' \| < \| \Sigma \|$. Furthermore, the induced map $i_*: \mathscr{S}(\Sigma') \to \mathscr{S}(\Sigma)$ is an injective map. We will identify from now on $\mathscr{S}(\Sigma')$ with $i_*(\mathscr{S}(\Sigma'))$ for incompressible subsurfaces Σ' in Σ. A subsurface Σ' in Σ is said to be *separating* if each boundary component of $\partial\Sigma'$ is a separating simple closed curve in Σ. Each separating subsurface in Σ is obtained as the closure of the connected component of $\Sigma - \bigcup_{i=1}^k s_i$ where $\{s_1, \dots, s_k\}$ is a collection of disjoint separating simple closed curves in Σ. If Σ' is a separating subsurface with $\partial\Sigma'$ consisting of k circles $s_1, \dots, s_k$, then there are k disjoint subsurfaces $\Sigma_1, \dots, \Sigma_k$ so that $\Sigma = \Sigma' \cup \bigcup_{i=1}^k \Sigma_i$ and $\Sigma_i \cap \Sigma' = s_i$. It follows that if Σ' is a separating subsurface of Σ and s is a simple closed curve in Σ', then s is separating in Σ' if and only if s is separating in Σ.

The goal now is to construct a finite collection of essential non-boundary parallel separating simple closed curves $\{s_1, \dots, s_k\}$ in Σ so that $\mathscr{X} = \{\alpha \in \mathscr{S}(\Sigma) \mid \alpha \cap [s_i] = \phi$ for some

$i\}$ satisfies $\mathscr{Z}^{\infty} = \mathscr{S}(\Sigma)$. Assuming this is achieved, we let Σ_{2i-1} and Σ_{2i} be the two incompressible separating subsurfaces obtained as the closure of the connected components of s_i in Σ, then $\mathscr{Z} = \bigcup_{i=1}^{2k} \mathscr{S}(\Sigma_i)$. Since $\| \Sigma_i \| < \| \Sigma \|$ by the construction, and Σ_i is in $\mathscr{L}$ unless its genus is zero, the induction hypothesis applies to Σ_i. We find that $\mathscr{S}(\Sigma_i)$ is finitely generated for each i. Putting these finitely many generators together, we obtain a finite set generating $\mathscr{S}(\Sigma)$.

To begin the proof, if $\| \Sigma \|$ is 3 or 4, then Σ is either a torus or a 1-holed torus. The result is well known as explained in the introduction (note that the map induced by the inclusion $i_* : \mathscr{S}(\Sigma_{1,1}) \to S(\Sigma_{1,0})$ is a bijective map). In both cases, $\mathscr{S}(\Sigma)$ is generated by three elements. If $\| \Sigma \| = 5$, then Σ is a 2-holed torus. By Propositions 2.8 and 2.15(a), $\mathscr{S}(\Sigma_{1,2})$ is generated by six elements. Assume now that Theorem I holds for Σ in $\mathscr{L}$ of $\| \Sigma \| < m$. Given a compact surface $\Sigma = \Sigma_{g,n}$ in $\mathscr{L}$ with $\| \Sigma \| = m \geqslant 6$, we choose a collection of simple closed non-separating curves a_i, b_i, and c_i for $i = 1, 2, \ldots, g$ in Σ so that:

(1) $a_i \perp b_i$, $b_i \perp c_i$, $c_i \perp a_i$, $a_i \cap b_i \cap c_i \neq \emptyset$ and c_i is not isotopic into a regular neighborhood of $a_i \cup b_i$ for each i;

(2) $a_i \cap a_j = a_i \cap b_j = b_i \cap b_j = 0$ when $i \neq j$;

(3) each of the two boundary components d_i and e_i of a small regular neighborhood $N(a_i \cup b_i \cup c_i)$ of $a_i \cup b_i \cup c_i$ is an essential separating curve. Furthermore, one of the curves, say d_i, is the boundary of a subsurface P_i in Σ so that $P_i \cap N(a_i \cup b_i \cup c_i) = d_i$ with $\partial P_i - d_i \subset \partial \Sigma$ and P_i is either a 3-holed sphere (in case $n \geqslant 2$) or a 1-holed torus (in case $n \leqslant 1$) as in Fig. 22.

The properties (1), (2) and (3) can easily be realized since $3g + n \geqslant 6$, $g \geqslant 1$, and $(g, n) \neq (2, 0)$. Note that the boundary component e_i is non-boundary parallel unless $\Sigma = \Sigma_{1,3}$ or $\Sigma_{2,1}$.

For simplicity, given a simple closed curve s in Σ, let $Com(s)$ be $\{\alpha \in \mathscr{S} \mid \alpha \cap [s] = \phi\}$. If S is a finite set of simple close curves $\{s_1, \ldots, s_k\}$, let $Com(S)$ be $\bigcup_{i=1}^{k} Com(s_i)$.

By Proposition 2.8, $\mathscr{S} = \mathscr{X}^{\infty}$ where $\mathscr{X} = \bigcup_{i=1}^{g} \mathscr{X}_i$ with $\mathscr{X}_i = \{\alpha \in \mathscr{S} \mid$ there is an index i so that either $I(\alpha, [a_i]) = I(\alpha, [b_i]) = 0$, or $0 \leqslant I(\alpha, [a_i]), I(\alpha, [b_i]), I(\alpha, [c_i]) \leqslant 1\}$. On the other hand, if $\mathscr{X} \subset \mathscr{Z}^{\infty}$, then $\mathscr{X}^{\infty} \subset \mathscr{Z}^{\infty}$ and also $\mathscr{X}^{\infty} \cup \mathscr{Y}^{\infty} \subset (\mathscr{X} \cup Y)^{\infty}$. Thus to finish the proof of Theorem I, it suffices to show the following.

2.18. Proposition. *For each index $i = 1, 2, \ldots, g$, there is a finite collection $\mathscr{C}_i$ of non-boundary parallel essential separating simple closed curves in Σ so that $\mathscr{X}_i \subset (Com(\mathscr{C}_i))^{\infty}$.*

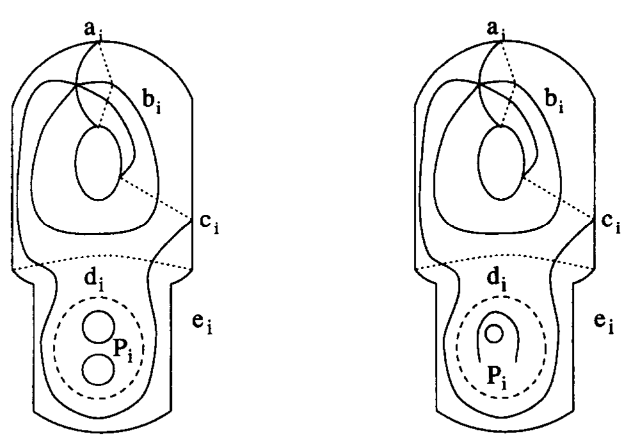

Fig. 22.

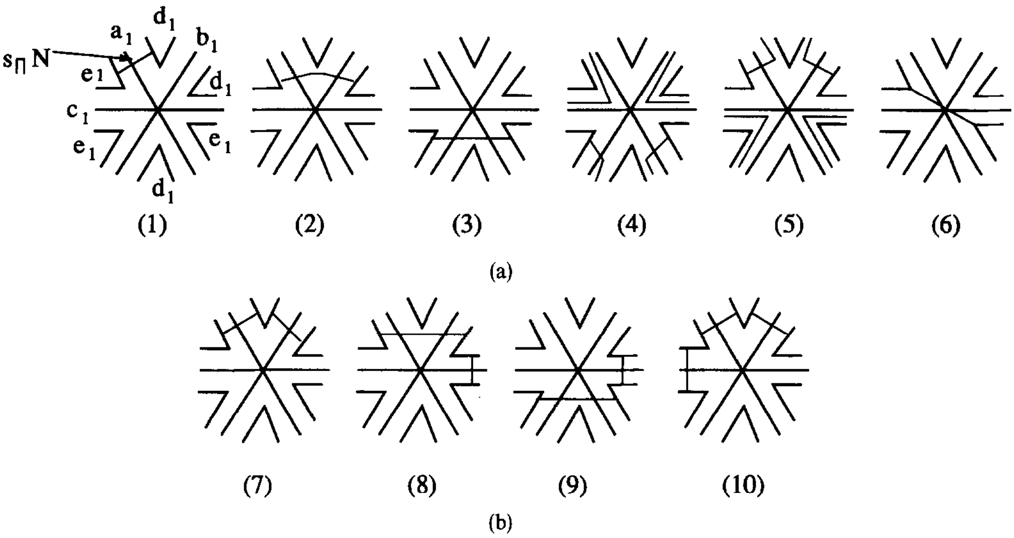

Fig. 23. In (a) the intersection is connected and in (b) the intersection is not connected.

By taking $\mathscr{C} = \bigcup_{i=1}^{g} \mathscr{C}_i$, we have that $\mathscr{X} \subset (Com(\mathscr{C}))^{\infty}$. This completes the proof of Theorem I for $\Sigma \neq \Sigma_{2,0}$.

Proof. For simplicity, we assume that the index $i = 1$ below and let $N = N(a_1 \cup b_1 \cup c_1)$ be the 2-holed torus. By the construction, $\partial N = d_1 \cup e_1$ consists of two essential separating simple closed curves and $d_1 = P_1 \cap N$ with $\partial P_1 - d_1 \subset \partial\Sigma$ where P_1 is either a 3-holed sphere (if $n \geqslant 2$) or a 1-holed torus (if $n \leqslant 1$). The curve d_1 is non-boundary parallel and the other curve e_1 is also non-boundary parallel unless $\Sigma = \Sigma_{1,3}$ or $\Sigma_{2,1}$. By the construction, both subsurfaces $N \cup P_1$ and N are separating in Σ.

Given $\alpha \in \mathscr{X}_1$, by Proposition 2.15, we may choose a representative $s \in \alpha$ so that either (1) $s \subset N$, or (2) $I(\alpha, [a_1]) = |s \cap a_1|$, $I(\alpha, [b_1]) = |s \cap b_1|$, $I(\alpha, [c_1]) = |s \cap c_1|$, and $s \cap N$ is isotopic to one of the ten types of arcs and their images under the 3-fold symmetry as in Fig. 23. Note that we represent N as in Fig. 7(b) and the 3-fold symmetry is induced by the $2\pi/3$ rotation in the plane. Figure 23 is divided into two parts according to the connectivity of $s \cap N$.

Now if α has a representative s in N, then $\alpha \in Com(d_1)$. Thus, by taking $[d_1]$ to be in the collection $\mathscr{C}_1$, we may assume that α is not in $\mathscr{S}(N)$, i.e. $s \cap N$ consists of arcs. In this case, since both d_1 and e_1 are separating curves, $|s \cap d_1|$ and $|s \cap e_1|$ are even numbers. Thus, types (1), (6), (8), (9), (10) and their images under the 3-fold symmetry in Fig. 23 cannot occur.

We now divide the proof into four cases where case 1 corresponds to $(g, n) \neq (1, 3)$ and $(2, 1)$ and P_1 being a 3-holed sphere, case 2 corresponds to $(g, n) = (1, 3)$ and P_1 being a 3-holed sphere, case 3 corresponds to $(g, n) \neq (1, 3)$ and $(2, 1)$ and P_1 being a 1-holed torus, and case 4 corresponds to $(g, n) = (2, 1)$.

2.19. *Case* 1. The surface $\Sigma = \Sigma_{g,n}$ has $(g, n) \neq (1, 3)$, $(2, 1)$ and P_1 is a 3-holed sphere. Then e_1 is non-boundary parallel. Take $[e_1]$ to be in the collection $\mathscr{C}_1$. Then if $s \cap N$ is of types (2), (3), (4) or (5), it follows that α is disjoint from one of the curves d_1 or e_1, i.e., $\alpha \in Com(\mathscr{C}_1)$. Thus, it remains to consider the only situation that $s \cap N$ is of type (7) in Fig. 23.

In the 3-holed sphere P_1, since $|s \cap d_1| = 2$ and $\partial P_1 - d_1 \subset \partial\Sigma$, the intersection $s \cap P_1$ is an arc. It follows that the intersection $s \cap (N \cup P_1)$ is also an arc. By the Dehn–Thurston classification of curve systems [3, 5, 8], the curve system $s \cap (N \cup P_1)$ in $N \cup P_1$ is obtained by

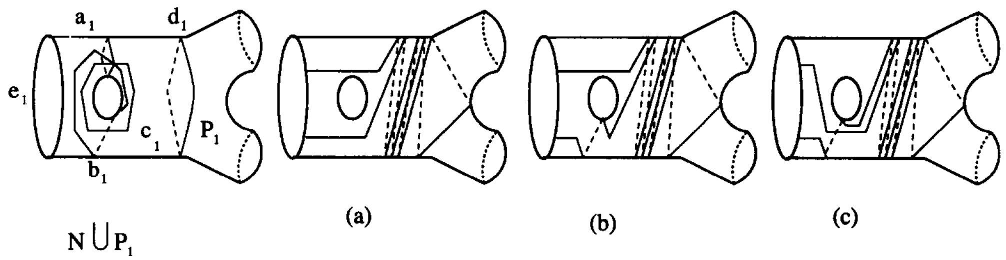

Fig. 24. In the figure, positive, negative, and half Dehn-twists are within the same class.

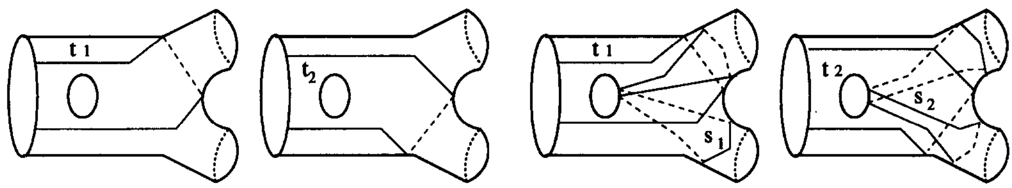

Fig. 25.

taking a copy of $s \cap N$ in N (there are three isotopy classes of them) and a copy of $s \cap P_1$ in P_1 (there is one isotopy class of it) and gluing them along d_1 with a twist. Therefore, there are three classes of $s \cap (N \cup P_1)$ as listed in Fig. 24. The classes (a), (b) and (c) are characterized by the triple of geometric intersection numbers ($I(\alpha, [a_1])$, $I(\alpha, [b_1])$, $I(\alpha, [c_1])$ being $(1, 1, 0)$, $(1, 0, 1)$ and $(0, 1, 1)$).

The three classes are induced by the 3-fold symmetry in N. In particular, there are two self-homeomorphisms ϕ_1 and ϕ_2 of Σ leaving each point in $\Sigma - (N \cup P_1)$ fixed which sends class (b) curves and class (c) curves to class (a) curves, respectively. Each curve system in class (a) is of the form $D^k_{d_1}(t_1)$ or $D^k_{d_1}(t_2)$ for some integer k where t_1 and t_2 are as in Fig. 25. Those curve systems of class (a) with positive (or negative) exponent k are said to have positive (or negative, respectively) Dehn–Thurston coordinate at d_1. Within the class (a) curves, there is a self-(orientation reversing) homeomorphism ϕ_3 of Σ leaving $N \cup P_1$ invariant which sends curves of negative Dehn–Thurston coordinates to curves of positive coordinates. Note that each curve t_1 and t_2 does not intersect a non-boundary parallel essential separating simple closed curve s_1 and s_2, respectively, as in Fig. 25.

Thus to finish the proof of case 1, it suffices to enlarge the collection that we constructed so far to a finite collection $\mathscr{C}'_1$ of non-boundary parallel essential separating curves so that $\mathscr{X}'_1 = \{\alpha \in \mathscr{X}_1 \mid \alpha$ has a representative s so that $s \cap (N \cup P_1)$ is a class (a) curve of positive Dehn–Thurston coordinate at d_1 in Fig. 24$\} \subset Com(\mathscr{C}'_1)^\infty$. Indeed, having constructed the finite collection $\mathscr{C}'_1$ we simply take $\mathscr{C}_1$ to be $\bigcup_{i=1}^{3} \phi_i(\mathscr{C}'_1)$.

To construct $\mathscr{C}'_1$, we introduce eight families of simple closed curves and arcs $A_k, B_k, C_k, D_k, E_k, F_k, G_k$, and H_k in $N \cup P_1$ of positive Dehn–Thurston coordinates at d_1 as in Fig. 26 where the index k is the sum of numbers of intersection points of the curve or arc with the marked arc l_1 and l_2. Note that the two intersection points of the arc F_k or G_k or H_k with e_1 have different intersection signs. This will be used to find cutting arcs for them.

For simplicity, given a simple closed curve or a proper arc C in $N \cup P_1$. we use C^* to denote the set of all isotopy classes of simple closed non-separating curves in Σ so that each element has a representative s with $s \cap (N \cup P_1)$ isotopic to C. In particular, if C is a simple closed non-separating curve, then $C^* = [C]$. Also $\mathscr{X}'_1 = \bigcup_{i=1}^{\infty} F^*_k$.

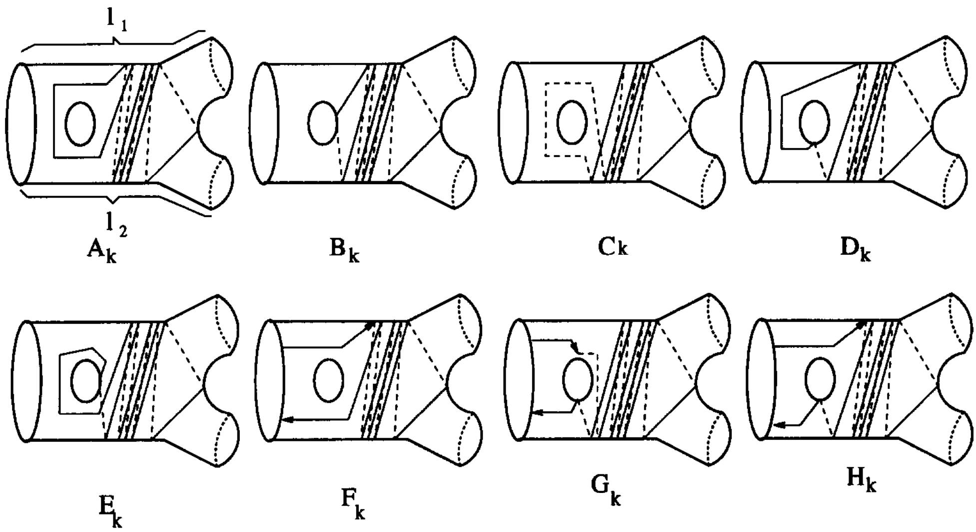

Fig. 26.

2.20. LEMMA. *We have the following recurrent relations where* $\mathscr{F} = Com(d_1)$.

(1) $A_k^* \subset (B_{k-1}^* \cup C_{k-2}^* \cup D_{k-1}^* \cup \mathscr{F})^\infty$;

(2) $B_k^* \subset (A_{k-1}^* \cup A_{k-3}^* \cup B_{k-4}^* \cup E_{k-2}^* \cup E_{k-4}^* \cup \mathscr{F})^\infty$;

(3) $C_k^* \subset (A_{k-2}^* \cup B_{k-1}^* \cup D_{k-2}^* \cup \mathscr{F})^\infty$;

(4) $D_k^* \subset (A_{k-1}^* \cup A_{k-3}^* \cup B_{k-2}^* \cup E_{k-4}^* \cup \mathscr{F})^\infty$;

(5) $E_k^* \subset (B_k^* \cup D_k^* \cup \mathscr{F})^\infty$;

(6) $F_k^* \subset (B_{k-1}^* \cup G_{k-2}^* \cup H_{k-1}^* \cup \mathscr{F})^\infty$;

(7) $G_k^* \subset (B_{k-1}^* \cup F_{k-2}^* \cup H_{k-1}^* \cup \mathscr{F})^\infty$;

(8) $H_k^* \subset (B_{k-2}^* \cup B_{k-4}^* \cup F_{k-3}^* \cup G_{k-3}^* \cup H_{k-2}^* \cup H_{k-4}^* \cup \mathscr{F})^\infty$.

Proof. The proof of the lemma is given by the chain of diagrams in Fig. 27. The notations are the same as before. The surface $N \cup P_1$ (drawn as in Fig. 24) is abbreviated as N–d_1 since all H-operations and isotopies are supported in the surface N. The cutting arcs are the darkened line segments in the diagrams.

By Lemma 2.20, we conclude that the set $\mathscr{X}_1' \subset (\bigcup_{i=1}^5 (A_i^* \cup B_i^* \cup \cdots \cup H_i^*) \cup \mathscr{F})^\infty$. On the other hand, $A_i^* \cup B_i^* \cup C_i^* \cup D_i^* \cup E_i^* \subset Com(e_1)$; for each i there is an integer k so that $F_i^* \subset Com(\{D_{d_1}^k(s_1), D_{d_1}^k(s_2)\})$ as in Fig. 25; similarly, for each G_i^*(and each H_i^*) there is an essential non-boundary parallel separating simple closed curve r so that G_i^* is a subset of $Com(r)$. This finishes the proof of case 1.

2.21. *Case* 2. The surface Σ is $\Sigma_{1,3}$. This case is actually covered in Lemma 2.20(1)–(5). Indeed, by the reduction before, it suffices to show that the subset $\mathscr{X}_1' = \{\alpha \in \mathscr{X}_1 \mid$ therefore is a representative s of α so that $s \cap N$ is of type (2) or (5) in Fig. 23 or their images under the 3-fold symmetry$\}$ is finitely generated. However, $\mathscr{X}_1'$ is the union of the images of $\mathscr{X}_1'' = \bigcup_{k=1}^\infty (A_k^* \cup B_k^* \cup C_k^* \cup E_k^*)$ under the homeomorphism ϕ_1, ϕ_2 and ϕ_3. Lemma 2.20 shows that $\mathscr{X}_1''$ is generated by $\bigcup_{k=1}^5 (A_k^* \cup B_k^* \cup C_k^* \cup D_k^* \cup E_k^*) \cup \mathscr{F}$, and $\mathscr{F}$ is finitely generated by Proposition 2.15(a). Thus the result follows.

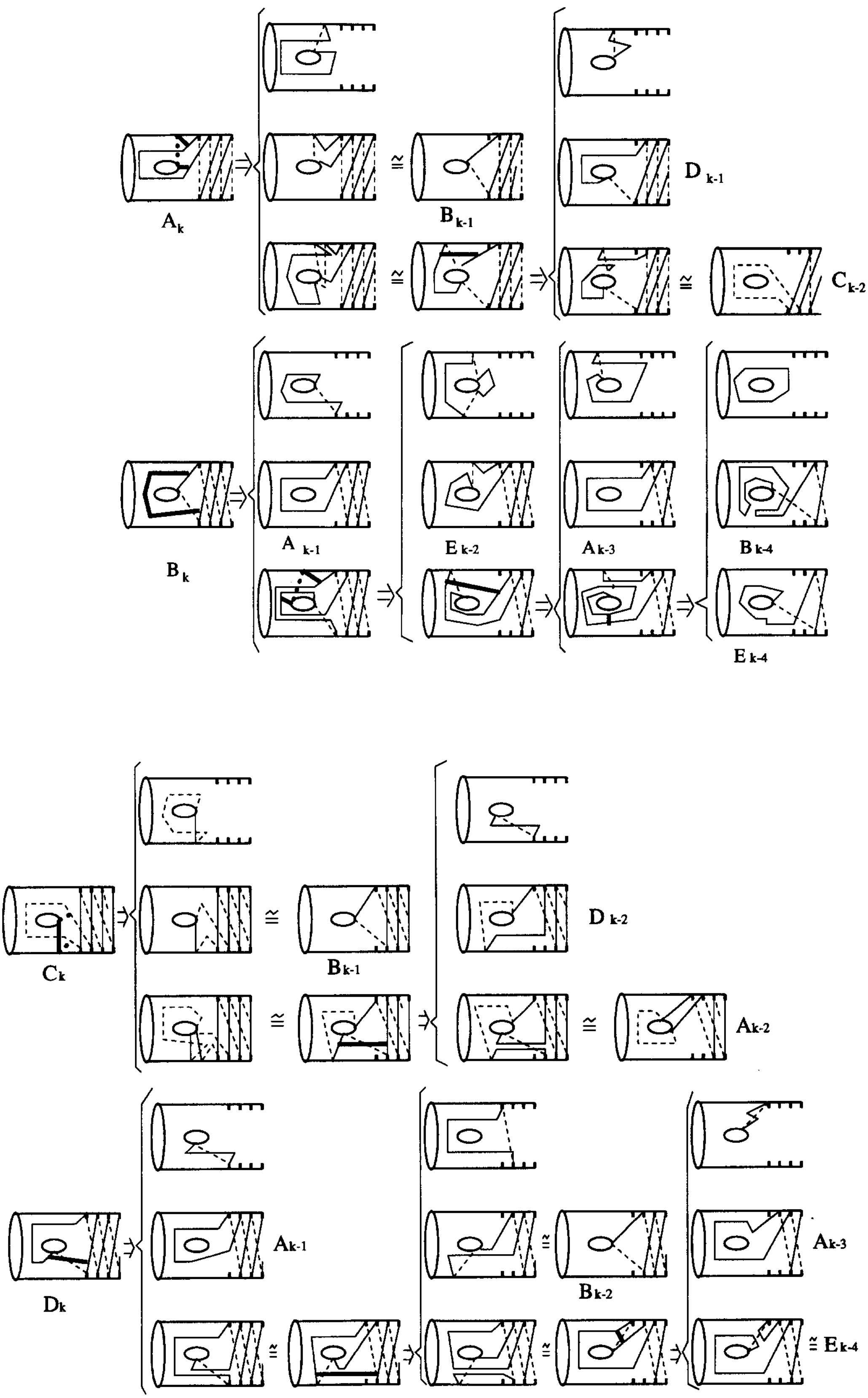

Fig. 27. *Continued on following two pages.*

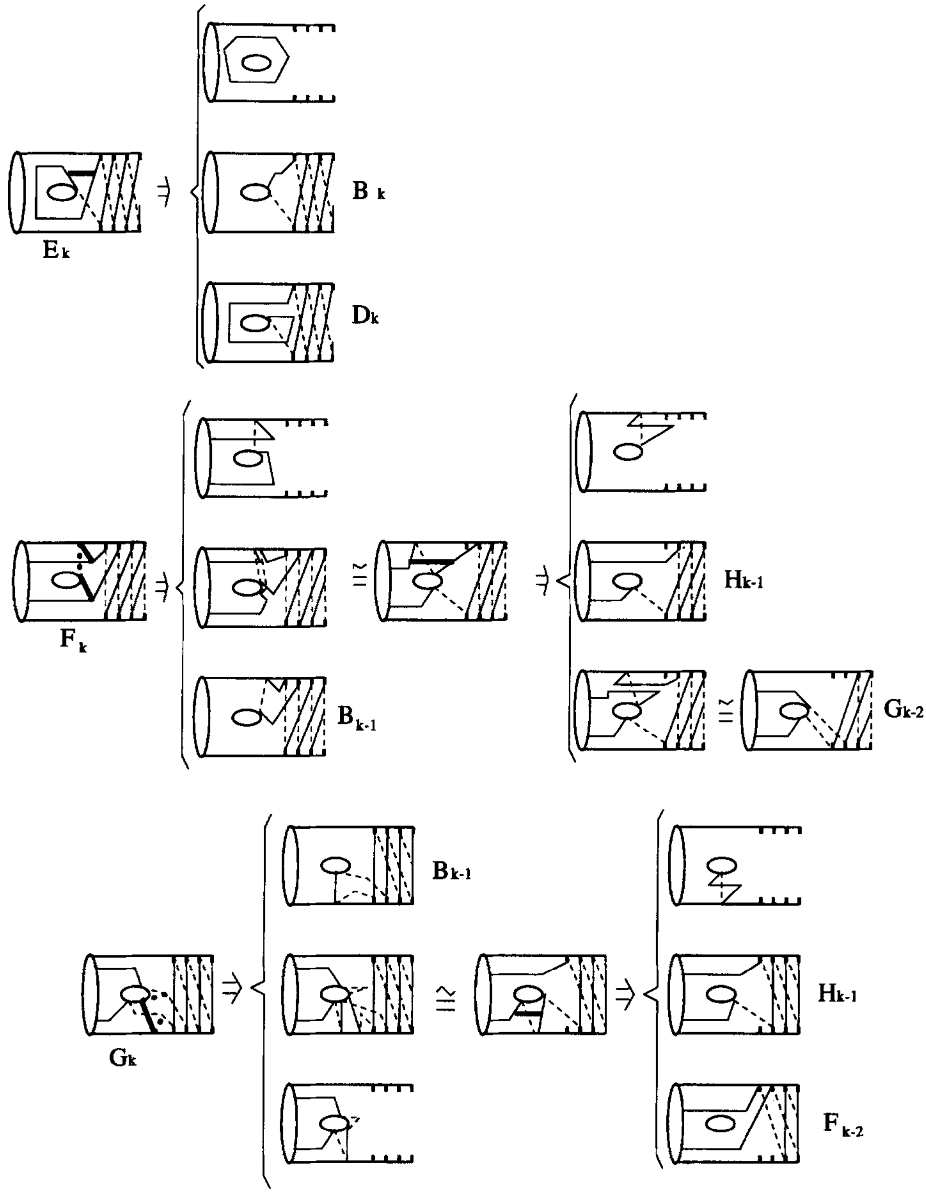

Fig. 27. *continued*

2.22. *Cases* 3 *and* 4. P_1 is a 1-holed torus. Choose two simple closed curves m_1 and l_1 in P_1 so that $m_1 \perp l_1$. For each α in $\mathscr{X}_1$, since $I(\alpha, [d_1]) = 2$, choose a representative s in α so that $|s \cap d_1| = 2$ and $I(\alpha, [m_1]) = |s \cap m_1|$, $I(\alpha, [l_1]) = |s \cap l_1|$, and $s \cap P_1$ is an arc. By the classification of curve systems in the 1-holed torus P_1, if one of the two numbers $|s \cap m_1|$ or $|s \cap l_1|$ is bigger than 1, then there exists an adjacent cutting arc for s in either $m_1 - l_1$ or $l_1 - m_1$. This is due to the fact that any two intersection points of s with m_1 (or l_1, respectively) have the same intersection signs. Thus the set $Y_1' = \{\alpha \in \mathscr{X}_1 \mid 0 \leqslant I(\alpha, [m_1]), I(\alpha, [l_1]) \leqslant 1\}$ generates $\mathscr{X}_1$. By the reduction before, it suffices to show that the subset $Y_1'' = \{\alpha \in Y_1' \mid \text{there is a representative } s \text{ for } \alpha \text{ so that } s \cap N \text{ is of type (2), or (5), or (7) in Fig. 23, or their images under the 3-fold symmetry}\}$ is finitely generated. Now each element α in Y_1'' has a representative s which does not intersect one of the four non-separating curves

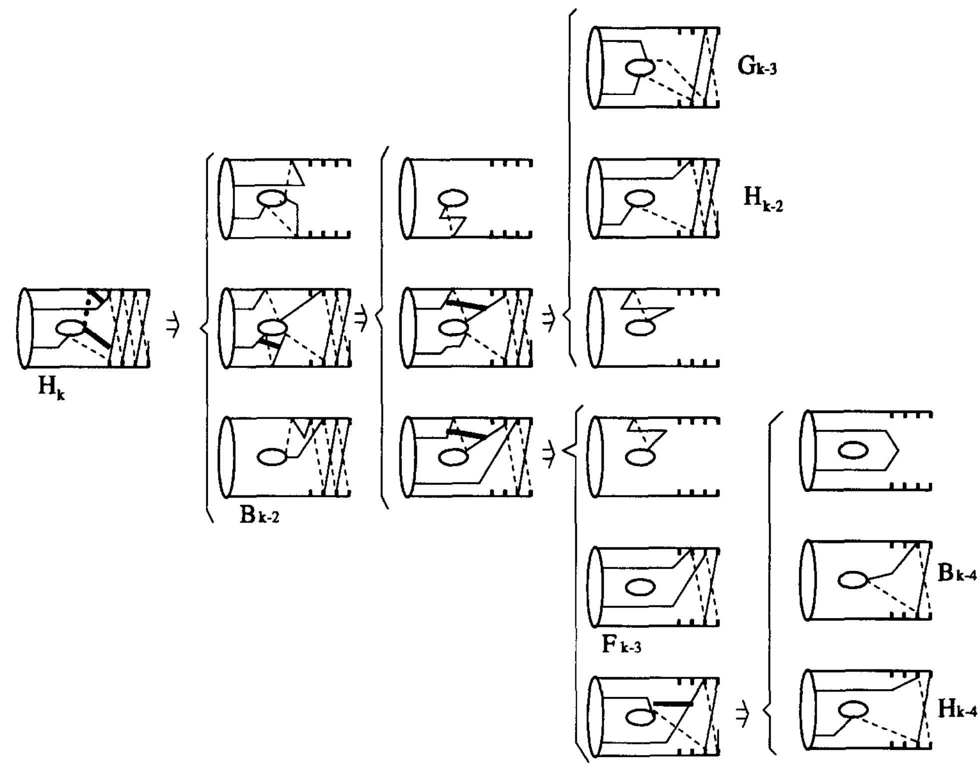

Fig. 27. *continued.*

$m_1, l_1, m_1 l_1$, or $l_1 m_1$, and s is orthogonal to one of the three curves a_1, b_1 or c_1 in N. By cutting Σ open along one of the four curves $m_1, l_1, m_1 l_1$, or $l_1 m_1$ which does not intersect s, we obtain a new surface Σ' with $\| \Sigma' \| < \| \Sigma \|$ in $\mathscr{L}$ so that s is still non-separating in Σ'. Note that Σ' is incompressible, not separating in Σ, and $\| \Sigma' \| < \| \Sigma \|$. Since the image of a non-separating simple closed curve in Σ' is still non-separating in Σ, by applying the induction hypothesis to the four possible surfaces Σ', we obtain a finite collection of isotopy classes of non-separating simple closed curve in Σ which generates Y_1''. This in turn shows that $\mathscr{S}(\Sigma)$ is finitely generated.

This finishes the proof of Theorem I for surfaces in $\mathscr{L}$. To see the result for $\Sigma_{2,0}$, we note that there is a natural inclusion i of $\Sigma_{2,1}$ into $\Sigma_{2,0}$ which sends the boundary curve to a null homotopic loop. This induces a surjective map from $\mathscr{S}(\Sigma_{2,1})$ to $\mathscr{S}(\Sigma_{2,0})$ which preserves the orthogonality. Thus, Theorem I holds for $\Sigma_{2,0}$.

This completes the proof of the Theorem I.

Acknowledgements—I would like to thank X.-S. Lin for many helpful discussions, and Peter Landweber and the referee for many helpful comments. I also thank Professors Boju Jiang and Shicheng Wang for their hospitality while part of the paper was done at Peking University. This work is supported in part by the NSF.

REFERENCES

1. F. Luo: On the mapping class groups of compact surfaces, preprint.
2. W. Thurston: On the geometry and dynamics of diffeomorphisms of surfaces, *Bull. AMS* **19**(2) (1988), 417–438.
3. A. Fathi, F. Laudenbach and V. Poènaru: Travaux de Thurston sur les surfaces, *Astérisque* (1979), 66–67.

4. F. Bonahon: Geometry of Teichmüller space via geodesic currents, *Invent. Math.* **92** (1988), 139–162.
5. M. Dehn: in *Papers on Group Theory and Topology*, J. Stillwell, Ed., Springer, New York (1987).
6. W. Lickorish: A representation of oriented combinatorial 3-manifolds, *Ann. Math.* **72** (1962), 531–540.
7. S. Humphries: Generators for the mapping class group, *LNM* **722** (1979), 44–47.
8. R. Penner and J. Harer: Combinatorics of train tracks, *Ann. Math. Stud.* **125** (1992).
9. A. Hatcher: Measured lamination spaces for surfaces, *Topology Appl.* **30** (1988), 63–88.
10. F. Luo: Non-separating simple closed curves in a compact surface, *ERA*, American Math. Soc., **1**(1) (1995), 18–25.
11. F. Luo: Geodesic Length Functions and Teichmüller spaces, preprint.
12. N. Ivanou: Automorphisms of complexes of curves and of Teichmüller spaces, preprint.

Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
U.S.A.
E-mail: fluo@math.rutgers.edu.