Topology Vol. 19, pp. 221–237
© Pergamon Press Ltd., 1980. Printed in Great Britain

0040-9383/80/0901-0221/$02.00/0

A PRESENTATION FOR THE MAPPING CLASS GROUP OF A CLOSED ORIENTABLE SURFACE

A. HATCHER and W. THURSTON

(*Received* 18 *April*, 1978; *in revised form* 20 *March* 1979)

THE CENTRAL objects of study in this paper are collections $\{C_1, \ldots, C_g\}$ of g disjoint circles on a closed orientable surface M of genus g, whose complement M-$(C_1 \cup \ldots \cup C_g)$ is a $2g$-punctured sphere. We call an isotopy class of such collections a *cut system*. Of course, any two cut systems are related by a diffeomorphism of M, by the classification of surfaces. We show that any two cut systems are also joined by a finite sequence of *simple moves*, in which just one C_i changes at a time, to a circle intersecting it transversely in one point and disjoint from the other C_j's. Furthermore, we find a short list of relations between sequences of simple moves, sufficient to pass between any two sequences of simple moves joining the same pair of cut systems.

From these properties of cut systems it is a routine matter to read off a finite presentation for the mapping class group of M, the group of isotopy classes of orientation preserving self-diffeomorphisms of M. Unfortunately, the presentation so obtained is rather complicated, and stands in need of considerable simplification before much light will be shed on the structure of the mapping class group. Qualitatively, one can at least deduce from the presentation that all relations follow from relations supported in certain subsurfaces of M, finite in number, of genus at most two. This may be compared with the result of Dehn [3] and Lickorish [4] that the mapping class group is generated by diffeomorphisms supported in finitely many annuli.

A finite presentation in genus two was obtained by Birman–Hilden[2], completing a program begun by Bergau–Mennicke[1]. For higher genus the existence of finite presentations was shown by McCool[10], using more algebraic techniques. For another approach to finite presentations, see [12], and for general background on mapping class groups, see [11].

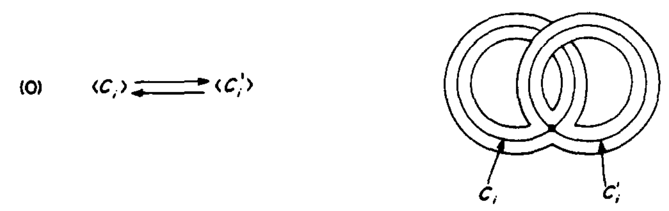
Our methods apply also to maximal systems of disjoint, non-contractible, non-isotopic circles on M. This is discussed briefly in an appendix.

§1. CUT SYSTEMS

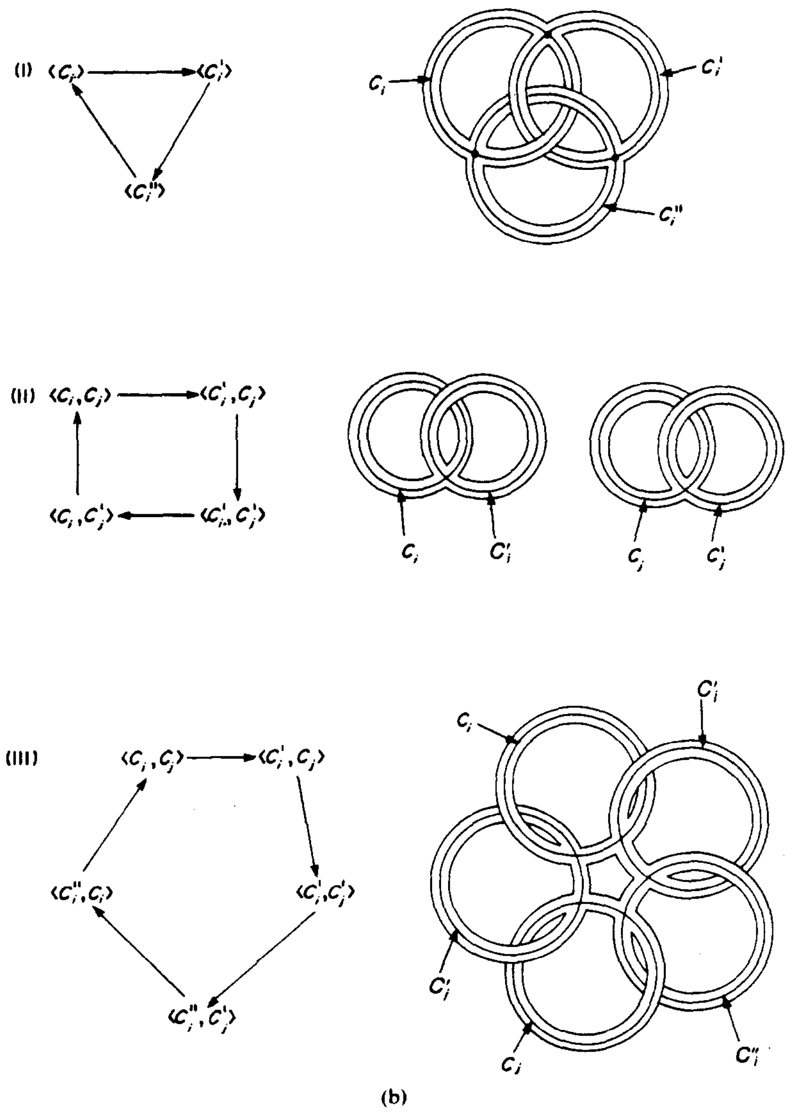
Let M be a closed orientable surface of genus g. We shall be considering unordered collections of g disjoint smoothly embedded circles $C_1, \ldots, C_g$ in M, whose complement M-$(C_1 \cup \ldots \cup C_g)$ is a $2g$-punctured sphere. Equivalently, $\{C_1, \ldots, C_g\}$ is a maximal non-separating system of disjoint circles on M. (In other contexts, $\{C_1, \ldots, C_g\}$ is termed a Heegaard diagram.) An isotopy class of such systems $\{C_1, \ldots, C_g\}$ we call a *cut system* $\langle C_1, \ldots, C_g \rangle$.

Let $\langle C_1, \ldots, C_g \rangle$ be a cut system, and suppose that for some i, C_i' is a circle in M intersecting C_i transversely in one point and disjoint from C_j for $j \neq i$. Then if C_i is replaced by C_i' in $\langle C_1, \ldots, C_g \rangle$, we obtain another cut system. The replacement $\langle C_1, \ldots C_i, \ldots, C_g \rangle \to \langle C_1, \ldots, C_i', \ldots, C_g \rangle$ is called a *simple move*. For brevity we will often drop the symbols for unchanging circles, e.g. writing $\langle C_i \rangle \to \langle C_i' \rangle$ for a simple move.

Certain sequences of simple moves are cycles, ending at the same cut system they began with. The simplest of these cycles are the following:



(a)



(b)

In each case, the only assumption on the circles involved is that the cut systems and simple moves written down are defined.

Let X_g^1 be the graph whose vertices are the cut systems on M, two vertices being joined by an (unoriented) edge if the corresponding cut systems are related by a simple move. (Thus the two arrows in (0) above determine a single edge of X_g^1.) Form a two-dimensional cell complex X_g from X_g^1 by attaching a 3-, 4- or 5-gon to each cycle of type (I), (II) or (III) above, respectively. For example, if M is the torus, then X_1 can be drawn as an open disc with its "rational" boundary points (Fig. 1).

Our main result is:

THEOREM 1.1. *X_g is connected and simply-connected.*

Remark. Without the pentagons, there would be a map of $\pi_1 X_g$ onto the permutation group of a fixed cut system $\langle C_1, \ldots, C_g \rangle$, since the cycle in (III) transposes C_i and C_j while the cycles in (I) and (II) do not permute $C_1, \ldots C_g$.

The proof of the Theorem will occupy the remainder of this section.

Let F: $M \to \mathbb{R}$ be a C^∞ function. Define an equivalence relation, $\sim_f$, on M by $x \sim_f y$ if x and y lie in the same component of a level set $f^{-1}(a)$ for some $a \in \mathbb{R}$. The quotient space $M/\sim_f$ we denote by $\Gamma(f)$, with quotient map $\bar{f}$: $M \to \Gamma(f)$. If f is generic, that is, if all the critical points of f are nondegenerate with distinct critical values, then $\Gamma(f)$ is a finite graph, whose vertices correspond to the critical points of f, as the pictures in Fig. 2 show.

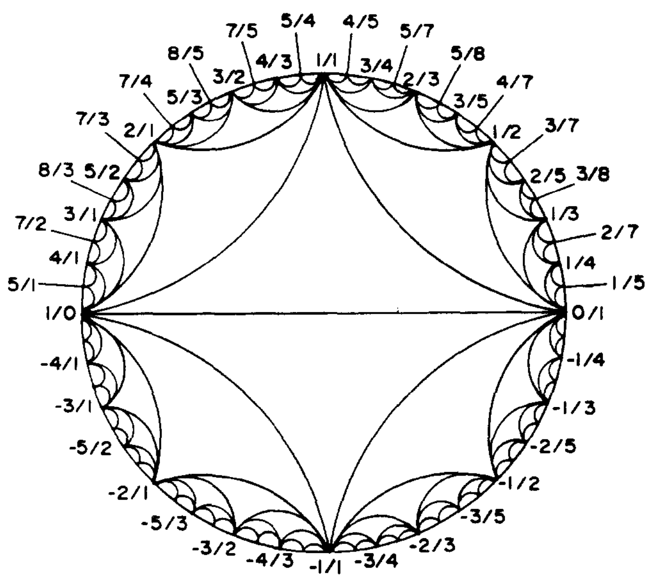


Fig. 1.

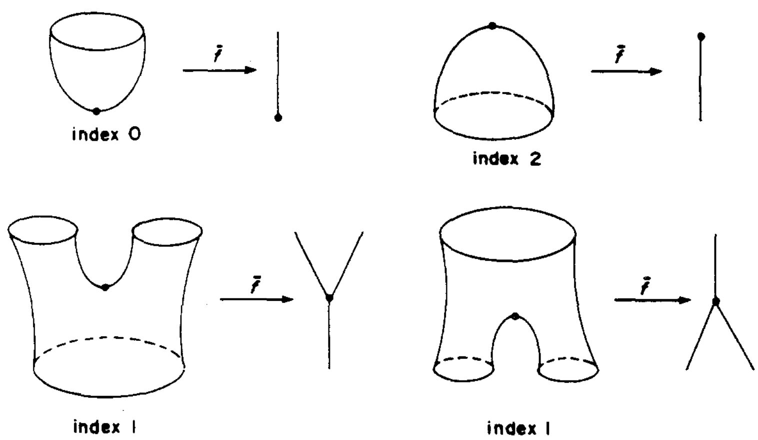


Fig. 2.

For generic f, M can be recovered from $\Gamma(f)$ as the boundary of an oriented 3-dimensional regular neighborhood of $\Gamma(f)$. This can be seen abstractly, by glueing together copies of the local pictures in Fig. 2 according to the edges of $\Gamma(f)$. Or more concretely, M can be embedded in $\mathbb{R}^3$ so that f becomes the height function.

Now choose a maximal tree T in $\Gamma(f)$, and label the edges of $\Gamma(f)$-T, $e_1, \ldots, e_g$, with interior points $\hat{e}_i \in e_i$. If we cut M open along all the circles $\bar{f}^{-1}(\hat{e}_i)$ we obtain a sphere with $2g$ holes, since this corresponds to cutting open the regular neighborhood of $\Gamma(f)$ along 2-discs until it becomes a regular neighborhood of T, hence a 3-ball. Thus to the pair $(f, T \subset \Gamma(f))$ there is associated the cut system $\langle \bar{f}^{-1}(\hat{e}_1), \ldots, \bar{f}^{-1}(\hat{e}_g)\rangle$.

Every cut system $\langle C_1, \ldots, C_g\rangle$ arises in this fashion, for suitable f and $T \subset \Gamma(f)$. Just let f be any generic function on M having $C_1, \ldots, C_g$ as non-critical level curves, and choose $T \subset \Gamma(f)$ to be the complement of the edges containing the points $\bar{f}(C_1), \ldots, \bar{f}(C_g)$.

Now let $(f_0, T_0 \subset \Gamma(f_0))$ and $(f_1, T_1 \subset \Gamma(f_1))$ have as their associated cut systems any two preassigned cut systems. Let $f_t : M \to \mathbb{R}$, $0 \le t \le 1$ be a generic path of C^∞ functions joining f_0 to f_1. Then for each t, f_t is generic with the following isolated exceptions:

(a) f_{t_0} has exactly one degenerate critical point, of the form $f_t(x, y) = x^3 \pm (t - t_0)x \pm y^2$. As t passes t_0, a pair of non-degenerate critical points of adjacent indices are born or die (birth–death point).

(b) f_t has two non-degenerate critical points whose critical values reverse order as t passes t_0 (crossing point).

A helpful picture is the graphic of f_t, $\{(f_t(x), t) | x$ is a critical point of $f_t\}$. For example, see Fig. 3. At a birth–death point, $\Gamma(f_t)$ changes as in Fig. 4.

It is not hard to see that $\Gamma(f_t)$ changes at a crossing point only if both critical points are of index one, and they belong to the same component of the level set of f_{t_0} which contains them. There are five essentially distinct possibilities, corresponding to the various ways of performing two surgeries on a collection of circles so that the resulting 2-dimensional cobordism (the trace of the surgeries) is connected. These are pictured in Figs. 5 and 6, along with the relevant portions of the graphs $\Gamma(f_t)$ for $t < t_0$, $t = t_0$, and $t > t_0$. It will turn out that only the case of Fig. 6 is really interesting, so we call this an *essential crossing*.

LEMMA 1.2. *Let $t = t_0$ be the parameter value of a birth–death or crossing point. Then maximal trees $T_t \subset \Gamma(f_t)$ can be chosen for t near t_0 so that, as t passes t_0, the*

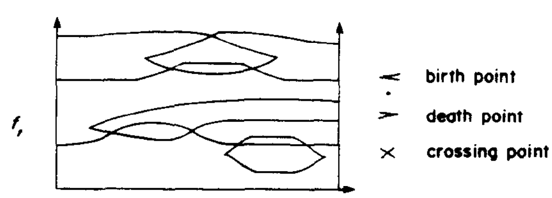


Fig. 3.

Fig. 4.

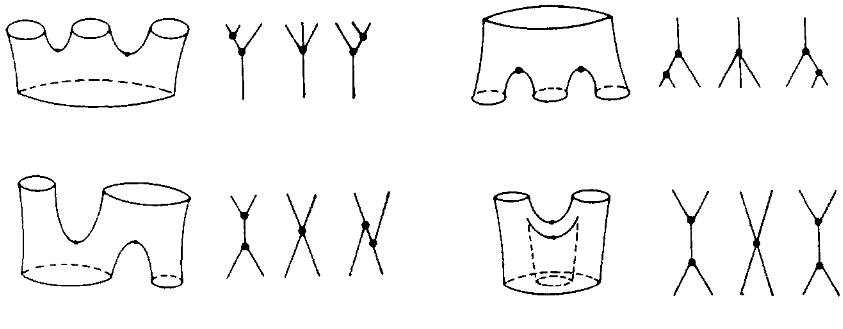


Fig. 5.

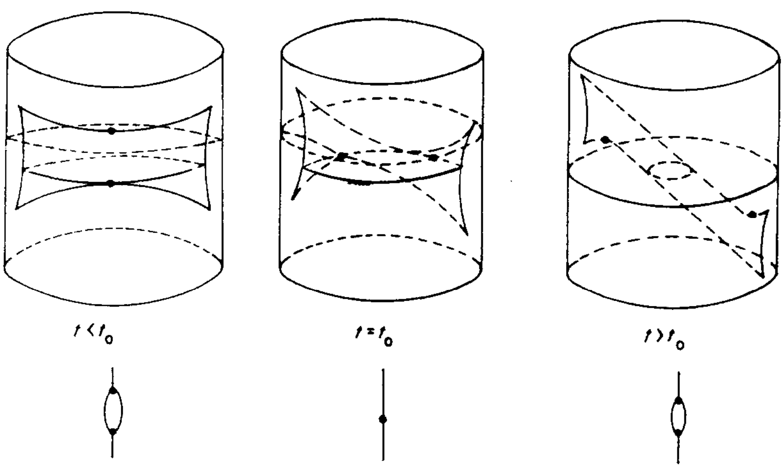


Fig. 6.

circles $\bar{f}^{-1}$ $(\hat{e}_i)$ vary only by isotopy in M, except in the case of an essential crossing, when just one $\bar{f}^{-1}$ $(\hat{e}_i)$ is replaced by a circle intersecting it transversely in one point.

Proof. Near a birth point, we can choose T_t so that the edge of $\Gamma(f_t)$ where the new branch grows belongs to T_t. For $t > t_0$, the new branch is added to T_t. A death point is similar. Near a crossing point, we need consider only the cases in Figs. 5 and 6. For inessential crossings $\Gamma(f_{t_0})$ is obtained from $\Gamma(f_t)$ for nearby t by collapsing one edge to a point. In these cases, we first choose an arbitrary tree T_{t_0} at time t_0, and pull it back to T_t for nearby t by adding the collapsed edge. In the case of an essential crossing, $\Gamma(f_{t_0})$ is obtained from $\Gamma(f_t)$ for nearby t be collapsing two edges to a point. We choose an arbitrary tree T_{t_0}, and then to obtain T_t for t on either side of t_0 we adjoin either one of the collapsed edges. The assertion about the circles $\bar{f}^{-1}$ $(\hat{e}_i)$ is then clear. □

Now we consider what happens in a t-interval between two successive birth–death or crossing points. The graph $\Gamma(f_t)$ is unchanged in this interval, but the maximal trees T_t chosen near the two ends of the interval may not be the same.

Definition. Let T and T' be maximal trees in a graph Γ, such that $T' - T$ is an edge a and $T - T'$ is an edge b. Then we say T' is obtained from T by an *elementary move*, and we write $T' = T + a - b$.

LEMMA 1.3. *Let T and T' be maximal trees in a finite graph Γ. Then we can pass from T to T' by a finite sequence of elementary moves, $T' = T + \sum_{j=1}^{n} (a_i - b_i)$.*

Proof. Let a be an edge of $T' - T$. Then $T + a$ contains a closed loop. This loop cannot be contained in T', so there exists an edge b in this loop, with $b \subset T - T'$. Thus $T + a - b$ is a maximal tree, and $T \to T + a - b$ is an elementary move which lessens the difference between T and T'. □

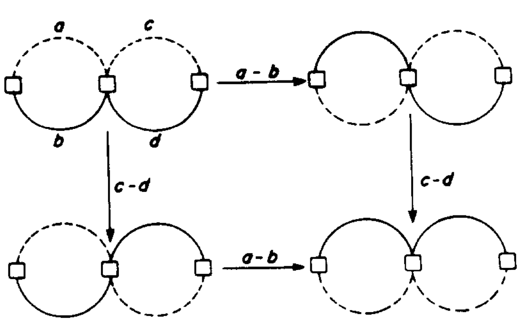
An elementary move $T \to T + a - b$ changes the associated cut system $\langle C_1, \ldots, C_g \rangle$ to one of the form $\langle C_1, \ldots, C_i', \ldots, C_g \rangle$, where $\bar{f}(C_i)$ and $\bar{f}(C_i')$ are interior points of b and a respectively. The unique circuit in $T + a$ lifts to a "linking" circle C_i'' on M which intersects each of C_i and C_i' transversely in one point and is disjoint from the other C_j's. So we have a pair of simple move $\langle C_i \rangle \to \langle C_i'' \rangle \to \langle C_i' \rangle$ realizing the elementary move $T \to T + a - b$.

Summarizing, we have associated an edge-path in X_g to a generic one-parameter family f_t together with appropriate choices of maximal trees $T_t \subset \Gamma(f_t)$. Since the cut systems associated to $(f_0, T_0 \subset (f_0))$ and $(f_1, T_1 \subset \Gamma(f_1))$ were arbitrary, this shows that X_g is connected.

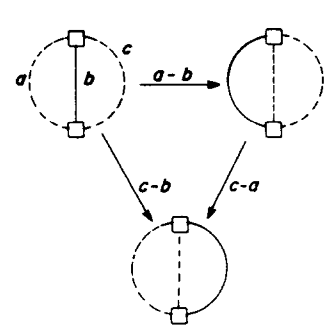
LEMMA 1.4. *For each edge-path in X_g there is a one-parameter family $(f_t, T_t \subset \Gamma(f_t))$ whose associated edge-path is the given edge-path, each edge (simple move) arising from an essential crossing.*

Proof. Let $(f_0, T_0 \subset \Gamma(f_0))$ be associated with the cut system $\langle C_1, \ldots, C_g \rangle$. It suffices to realize the simple move $\langle C_i \rangle \to \langle C_i' \rangle$ by a path $(f_t, T_t \subset \Gamma(f_t))$ with a single essential crossing and no elementary moves in the maximal trees T_t. First deform f_0 to $f_{1/2}$, changing it only in a neighborhood of C_i' (but staying fixed near C_i) so that $f_{1/2}|C_i'$ has one local maximum and one local minimum, both at saddles of $f_{1/2}$, and so that $f_{1/2}$ has no other critical points between the levels of these two saddles. For the resulting (generic) path f_t from f_0 to $f_{1/2}$ there is a natural choice of maximal trees $T_t \subset \Gamma(f_t)$, namely the complement of the edges containing the points $\bar{f}_t(C_i), \ldots, \bar{f}(C_g)$. Thus no essential crossings or elementary moves are involved. Now deform $f_{1/2}$ to f_1 by simply interchanging the levels of the two saddles on C_i', an essential crossing realizing the simple move $\langle C_i \rangle \to \langle C_i' \rangle$. □

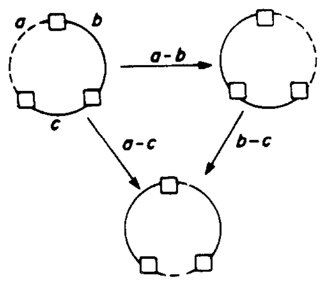
LEMMA 1.5. *All relations between sequences of elementary moves are consequences of the relations* (1)–(4), *below.* (*Boxes represent subgraphs in which the maximal tree is unchanged.*)



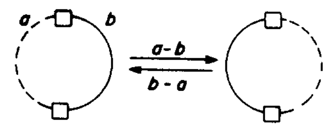
(1) $(a - b) + (c - d) = (c - d) + (a - b)$



(2) $(a - b) + (c - a) = (c - b)$



(3) $(a-b)+(b-c)=(a-c)$

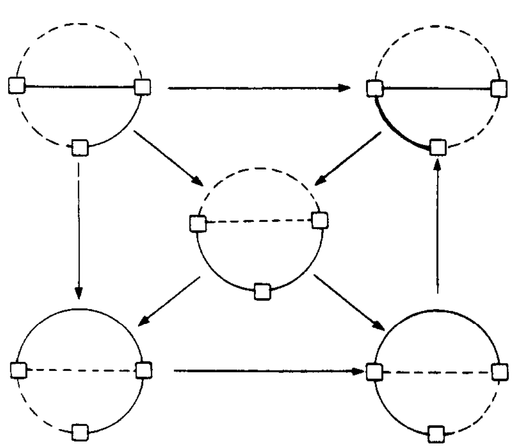


(4) $(a-b)+(b-a)=0$

Proof. Since (4) provides inverses, it suffices to show a relation $(a_1-b_1)+\ldots+(a_n-b_n)=0$ follows from (1) to (4). In such a relation the b_i's are a permutation of the a_i's. Let $m=\min\{|p-q|\colon a_p=b_q\}$ and suppose $a_j=b_k$ realizes this minimum, $m=|j-k|$. If $m>1$ and $j<k$, then applying Lemma 1.6 below to $(a_j-b_j)+(a_{j+1}-b_{j+1})$ produces a new relation of length n with smaller m. Similarly, if $k<j$ apply the lemma to $(a_{j-1}-b_{j-1})+(a_j-b_j)$ to decrease m. Eventually we reach $m=1$ where by using (2), (3) or (4) we can reduce to a smaller n. So by induction on n the lemma is proved. □

LEMMA 1.6. *A pair of elementary moves $(a-b)+(c-d)$ with a, b, c, d distinct can be transformed into one of the pairs of elementary moves $(c-d)+(a-b)$, or $(c-b)+(a-d)$, using relations* (1) *to* (4).

Proof. There are two possible configurations, the one in relation (1) and the one pictured below.



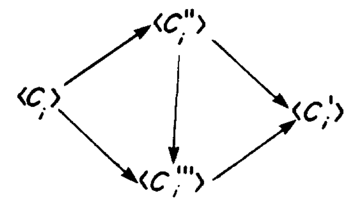
According to how the four edges are labelled, $(a-b)+(c-d)$ corresponds to going from one corner to the diagonally opposite corner, via any of the three intermediate stages. It is easy to verify in all three cases that one of the other two routes is a pair of elementary moves $(c-d)+(a-b)$ or $(c-b)+(a-d)$. □

LEMMA 1.7. *Let $T_t\subset\Gamma(f_t)$ and $T'_t\subset\Gamma(f_t)$ be maximal trees for the generic one-parameter family f_t, chosen according to the rules given above. Then the edge-paths associated to $(f_t, T_t\subset\Gamma(f_t))$ and $(f_t, T'_t\subset\Gamma(f_t))$ are homotopic in X_g.*

Proof. The choices were: (a) The choice of "linking" circle for an elementary move; (b) The choice of representation of a change of maximal trees (in a fixed $\Gamma(f_t)$)

as a sequence of elementary moves; (c) The choice of maximal trees $T_t \subset \Gamma(f_t))$ in t-slices near birth–death and crossing points.

(a) Let C''_i and C'''_i be two circles linking C_i and C'_i, and intersecting each other transversely in n points. If $n = 1$ there are simple moves so the two ways of going from C_i to C'_i are homotopic via type (I) 2-cells of X_g. If $n = 0$ we may find a linking circle C''''_i



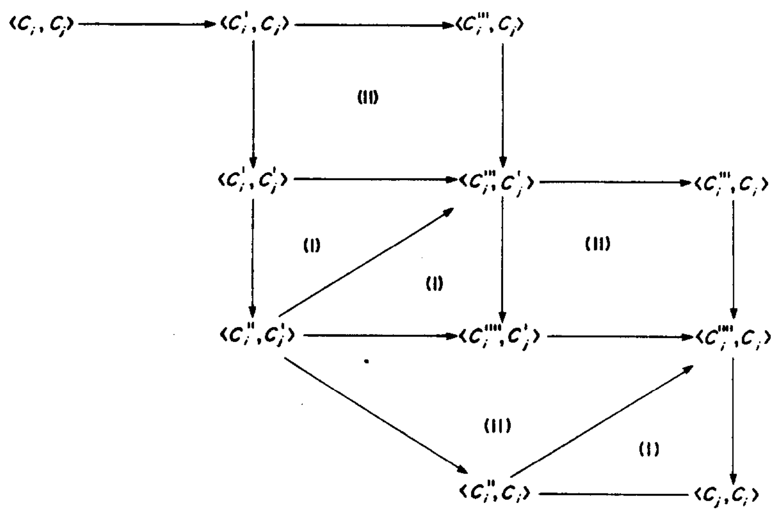
which intersects both C''_i and C'''_i in one point transversely (see Fig. 7), so this case reduces to the previous case. If $n > 1$, we can find a linking circle C''''_i meeting C'''_i in one point and C''_i in fewer than n points (transversely). See Fig. 8. So by induction on n we are done.

(b) We check what homotopies between associated edge-paths are induced by relations (1)–(4) of Lemma 1.5.

(1) Here we can choose the linking circles for the two elementary moves to be disjoint. So this commutation relation among elementary moves follows from the commutation relation (II) for simple moves.

(2) The configuration here is shown in Fig. 9 (with suitable choice of linking circles C'_i, C'_j and C''''_i).

Consider the diagram of simple moves:



The path from $\langle C_i, C_j\rangle$ to $\langle C_j, C_i\rangle$ across the top of the diagram corresponds to the cycle of three elementary moves in relation (2). The path across the bottom is the cycle (III). These paths are homotopic using (I) and (II), as shown.

(3) We may choose the same linking circle for all three elementary moves. Then the edge-path is

$$\langle C_i\rangle \to \langle C'''_i\rangle \to \langle C'_i\rangle \to \langle C'''_i\rangle \to \langle C''_i\rangle \to \langle C'''_i\rangle \to \langle C_i\rangle$$

which is null-homotopic in the 1-skeleton of X_g.

(4) This is just $\langle C_i\rangle \to \langle C''_i\rangle \to \langle C'_i\rangle \to \langle C''_i\rangle \to \langle C_i\rangle$, which is again null-homotopic in X_g^1.

(c) Clearly, only the case of an essential crossing is of interest here. Near an essential crossing in the slice $t = t_0$, $T_t \subset \Gamma(f_t)$ depended first on choosing $T_{t_0} \subset \Gamma(f_{t_0})$,

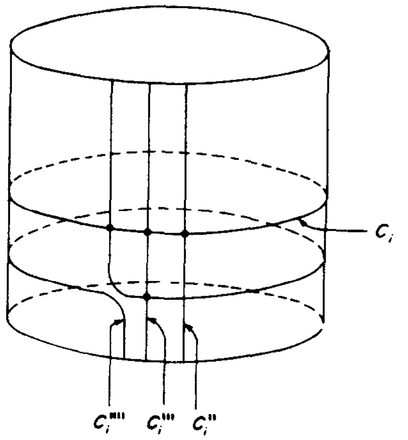


Fig. 7.

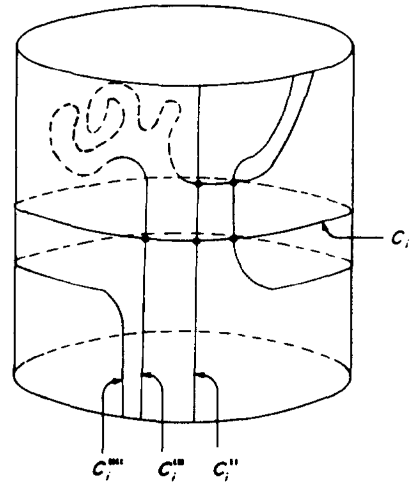


Fig. 8.

and second on choosing for nearby $t \neq t_0$ one of the two collapsing edges of $\Gamma(f_t)$ to belong to T_t.

Consider first the choice of $T_{t_0} \subset \Gamma(f_{t_0})$. A new choice differs by elementary moves in $\Gamma(f_{t_0})$, so it suffices to consider one such elementary move $T_{t_0} \to T_{t_0} + a - b$. In the ambient one-parameter family this can be realized by first introducing inverse elementary moves $T_t \to T_t + a - b \to T_t + (a - b) + (b - a)$ just to one side of $t = t_0$ (the effect of this on sequences of cut systems was discussed in (b) above), and then commuting one of these elementary moves with the given essential crossing. Thus simple moves $\langle C_j \rangle \to \langle C_j'' \rangle \to \langle C_j' \rangle$ corresponding to the elementary move are to commute with the simple move $\langle C_i \rangle \to \langle C_i' \rangle$ corresponding to the essential crossing. Here C_j and C_j' correspond to a and b, edges of $\Gamma(f_{t_0})$, so C_j and C_j' are disjoint from $C_i \cup C_i'$. Also, C_j'' can be chosen disjoint from $C_i \cup C_i'$. So the desired commutation relation follows from (II).

And second, there is the choice of which collapsing edge of $\Gamma(f_t)$ to put in T_t, for t on either side of t_0. Two choices differ by an elementary move, so only a homotopy in X_g^1 can be involved. □

Proof that $\pi_1 X_g = 0$: By Lemma 1.4, we can realize a given closed edge-path in X_g by a one-parameter family $(f_t, T_t \subset \Gamma(f_t))$. Since the edge-path is closed, we may take $(f_0, T_0 \subset \Gamma(f_0)) = (f_1, T_1 \subset \Gamma(f_1))$ by splicing in a one-parameter family $(f_t, T_t \subset \Gamma(f_t))$ without essential crossings or elementary moves, as in the proof of Lemma 1.4.

Let f_{tu}, $(t, u) \in I \times I$ be a generic two-parameter family realizing a homotopy from

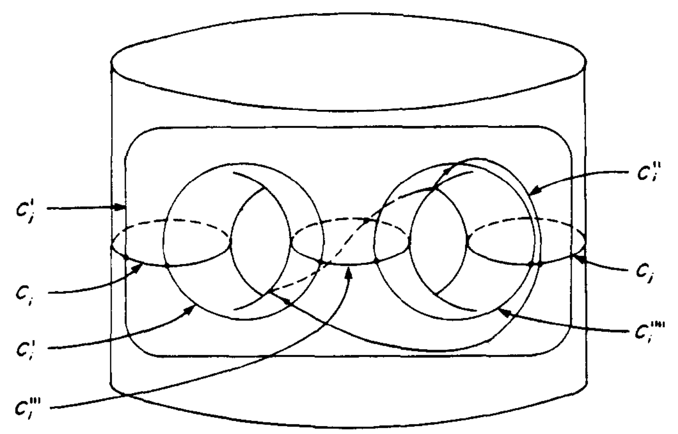


Fig. 9.

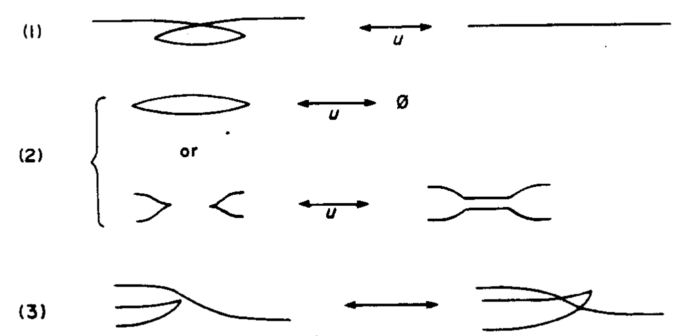


Fig. 10.

the loop f_t to the constant loop f_0. Except for the isolated phenomena described below in (1)–(6), f_{tu_0} will be a generic one-parameter family for each u_0 (see [5, 6]).

(1) A co-dimension two singularity, of the form

$$f_{tu}(x, y) = \pm x^4 \pm (u - u_0)x^2 \pm (t - t_0)x \pm y^2.$$

(2) A birth and death point are cancelled or introduced.

(3) A birth–death point crosses a non-degenerate critical point.

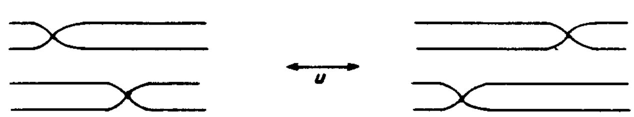
The changes in the graphic from $u < u_0$ to $u > u_0$ are shown in Fig. 10.

In each case the portion of M between levels just above and below the phenomenon in question is a genus zero surface, so no essential crossings are involved, and maximal trees can be chosen for nearby t and u so that no elementary moves are involved. Hence the associated path of simple moves is unchanged.

(4) Two crossings cancel or are introduced:

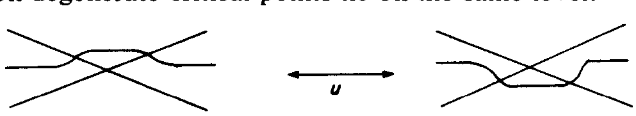
This is non-trivial only for essential crossings, where it is a null-homotopy of a path $\langle C_i\rangle \to \langle C_i'\rangle \to \langle C_i\rangle$.

(5) Two birth–death crossing points occur simultaneously. This is non-trivial only for two essential crossings.



This is a commutation cycle (II).

(6) Three non-degenerate critical points lie on the same level:



The only non-trivial cases occur when all three critical points are saddles, and at least one essential crossing is involved. To enumerate the possibilities, we regard the saddles as one-handles attached to level circles just below the level of the three saddles. From this viewpoint an essential crossing looks like:

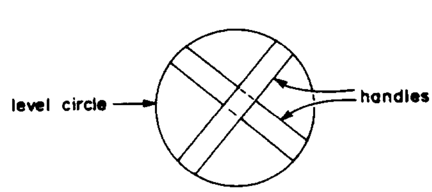


Fig. 11.

Up to sliding handles (i.e. isotopy), the essentially distinct ways of involving a third saddle are listed in Fig. 12.

The configurations of each pair $(a - a')$, $(b - b')$, $(c - c')$ are obtained from each other by replacing f_{tu} by $-f_{tu}$, so it suffices to consider only (a), (b) and (c).

(a) Consider the one-parameter family obtained by restricting (t, u) to a small circle about the point where all three saddles are on the same level. The graphs $\Gamma(f_{tu})$ go through the cycle in Fig. 13. Choose maximal trees to include the edges parallelled by dotted lines. There are three essential crossings, giving simple moves $\langle\alpha\rangle \to \langle\beta\rangle \to \langle\gamma\rangle \to \langle\alpha\rangle$, a cycle of type (I).

(b) The pictures here are shown in Fig. 14. There is an elementary move required. This corresponds to two simple moves $\langle\gamma_3\rangle \to \langle\gamma_1\rangle \to \langle\gamma_2\rangle$ which exactly cancel the two simple moves $\langle\gamma_2\rangle \to \langle\gamma_1\rangle \to \langle\gamma_3\rangle$ coming from the two essential crossings.

(c) Here only one essential crossing is involved, and no elementary moves, so the path of simple moves is unaffected. □

§2. GENERATORS AND RELATIONS FOR THE MAPPING CLASS GROUP

As a standard model for M we choose a sphere with g handles attached, as in Fig. 15.

Let H be the subgroup of $G = \pi_0 \operatorname{Diff}^+(M)$ represented by diffeomorphisms which preserve the standard cut system $\langle\alpha_1, \ldots, \alpha_g\rangle$. (Elements of H can permute the α_i's and reverse their orientations.) At the end of this section we shall derive by well-known methods:

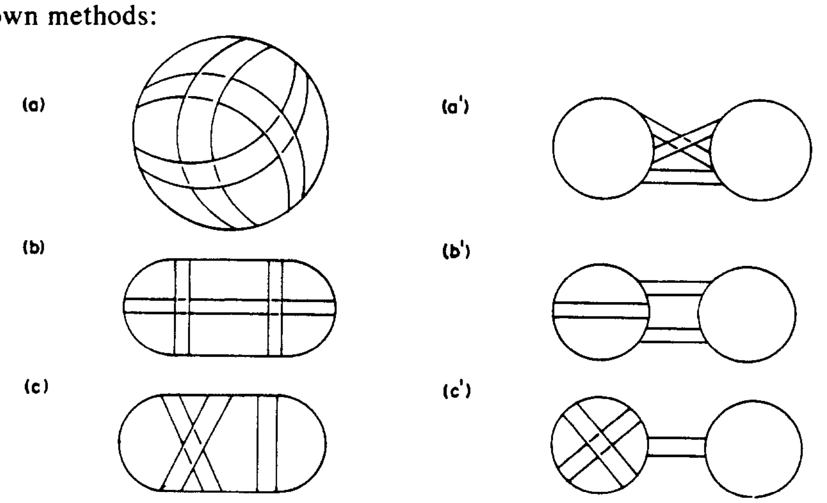


Fig. 12.

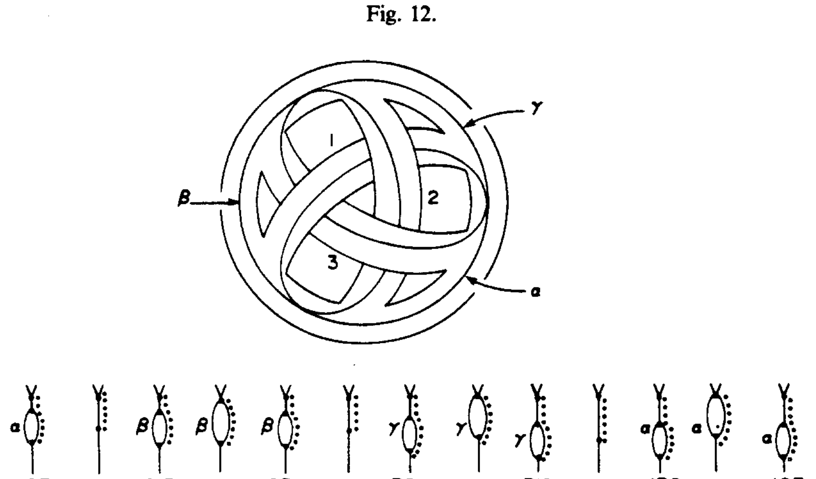


Fig. 13.

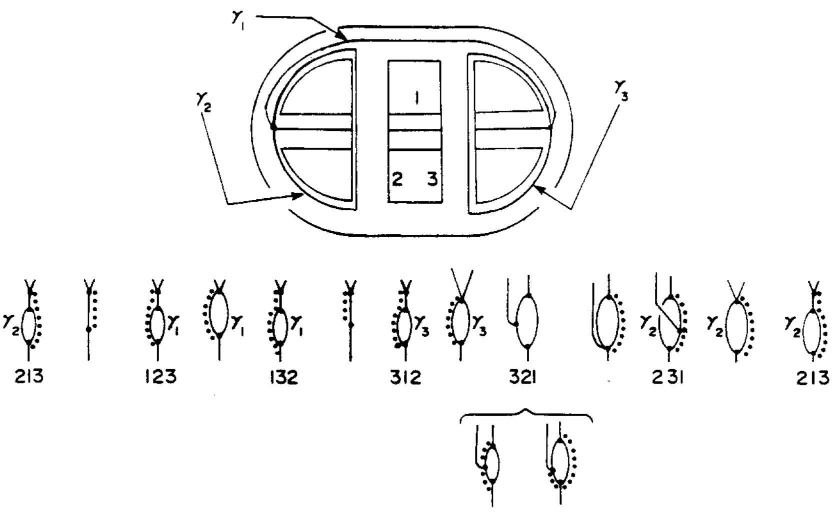


Fig. 14.

PROPOSITION 2.1. *There is an exact sequence*

$$Z \to Z^g \oplus P_{2g-1} \to H \to \pm \sum_g \to 0$$

where P_{2g-1} *is the pure braid group on* $2g-1$ *strands and* $\pm\Sigma_g$ *is the group of signed permutations of g objects.*

From this one can write down a presentation for H.

Let $\sigma \in G$, supported near $\alpha_1 \cup \beta_1$, be a "90° rotation" of the first handle of M, as shown in Fig. 16.

THEOREM 2.2. *Every element of G can be expressed in the form* $\rho_{r+1}\,\sigma\rho_r \ldots \sigma\rho_1$, $\rho_i \in H$. *All relations between such words follow from the relations* (A)–(E) *below.*

Proof. For $\varphi \in \mathrm{Diff}^+(M)$, there is by Theorem 1.1 a path of cut systems $\langle\alpha_1,\ldots,\alpha_g\rangle = \langle C_1^0,\ldots,C_g^0\rangle$, $\langle C_1^1,\ldots,C_g^1\rangle,\ldots,\langle C_1^r,\ldots,C_g^r\rangle = \langle\varphi^{-1}(\alpha_1),\ldots,\varphi^{-1}(\alpha_g)\rangle$ related by simple moves $\langle C_{i_k}^{k-1}\rangle \to \langle C_{i_k}^{k}\rangle$. To this path we associate a word $\sigma\rho_r \ldots \sigma\rho_1$ as follows. Suppose inductively that $\sigma\rho_{k-1}\ldots\sigma\rho_1$ takes $\langle C_1^{k-1},\ldots,C_g^{k-1}\rangle$ to $\langle\alpha_1,\ldots,\alpha_g\rangle$. Then $\sigma\rho_{k-1}\ldots\sigma\rho_1(C_{i_k}^{k-1}) = \alpha_i$ for some i, and $\sigma\rho_{k-1}\ldots\sigma\rho_1(C_{i_k}^{k})$ is a circle meeting α_i

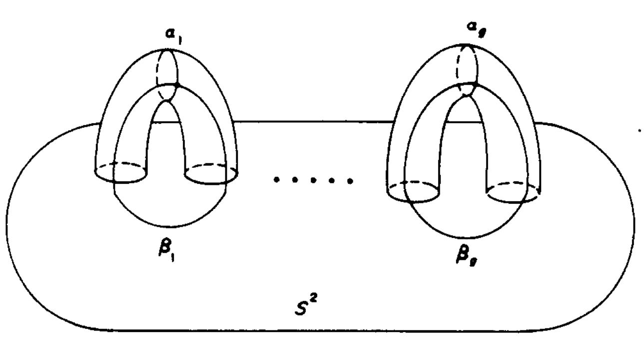


Fig. 15.

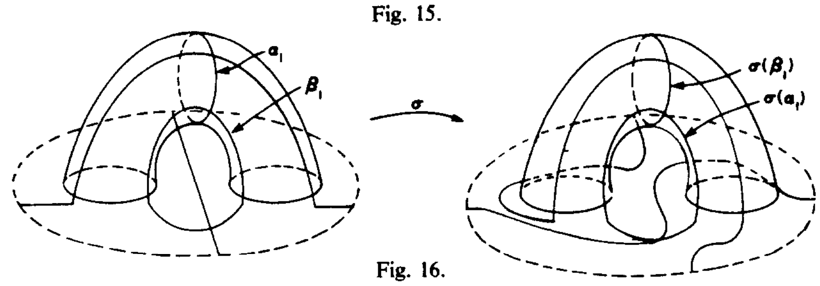


Fig. 16.

transversely in one point and disjoint from α_j, $j \neq i$. So there exists $\rho_k \in H$ with the property that $\rho_k\sigma\rho_{k-1}\ldots\sigma\rho_1(C_{i_k}^k) = \beta_1$. Hence $\rho_k\sigma\rho_{k-1}\ldots\sigma\rho_1(C_{i_k}^{k-1}) = \alpha_1$, and $\sigma\rho_k\ldots\sigma\rho_1$ takes $\langle C_1^k, \ldots, C_g^k\rangle$ to $\langle\alpha_1, \ldots, \alpha_g\rangle$. When $k = r$, both $\sigma\rho_r\ldots\sigma\rho_1$ and φ take $\langle\varphi^{-1}(\alpha_1), \ldots, \varphi^{-1}(\alpha_g)\rangle$ to $\langle\alpha_1, \ldots, \alpha_g\rangle$, so they differ by an element of H, that is, φ has the form $\rho_{r+1}\sigma\rho_r\ldots\sigma\rho_1$ with $\rho_i \in H$. This is the first statement of Theorem 2.1.

Suppose now that $\varphi \in G$ has two representations as words in σ and H,

$$\varphi = \rho_{r+1}\sigma\rho_r\ldots\sigma\rho_1 = \rho'_{s+1}\sigma\rho'_s\ldots\sigma\rho'_1.$$

Both words are associated to paths of simple moves; for example $\rho_{r+1}\sigma\rho_r\ldots\sigma\rho_1$ is associated to

$$\langle C_1^k, \ldots, C_g^k\rangle = \langle(\sigma\rho_k\ldots\sigma\rho_1)^{-1}(\alpha_1), \ldots, (\sigma\rho_k\ldots\sigma\rho_1)^{-1}(\alpha_g)\rangle,\ k = 0, 1, \ldots, r.$$

LEMMA 2.3. *Two words $\rho_{r+1}\sigma\rho_r\ldots\sigma\rho_1 = \rho'_{r+1}\sigma\rho'_r\ldots\sigma\rho'_1$ associated to the same path of simple moves may be obtained from each other using only the relation*

(A) σ commutes with $H(\alpha_1, \beta_1)$, the subgroup of H represented by diffeomorphisms leaving both α_1 and β_1 invariant.

Proof. Suppose inductively the lemma is known when $\rho_1 = \rho'_1, \ldots, \rho_k = \rho'_k$. We show that it is then also true assuming only $\rho_1 = \rho'_1, \ldots, \rho_{k-1} = \rho'_{k-1}$. Let $\rho''_k = \rho'_k\rho_k^{-1}$. This is in $H(\alpha_1, \beta_1)$ since

$$\rho'_k\rho_k^{-1}(\beta_1) = \rho'_k\sigma\rho_{k-1}\ldots\sigma\rho_1(C_{i_k}^k) = \rho'_k\sigma\rho'_{k-1}\ldots\sigma\rho'_1(C_{i_k}^k) = \beta_1.$$

So

$$\begin{aligned}\rho'_{r+1}\sigma\rho'_r\ldots\sigma\rho'_{k+1}\sigma\rho'_k\sigma\rho_{k-1}\ldots\sigma\rho_1 &= \rho'_{r+1}\sigma\rho'_r\ldots\sigma\rho'_{k+1}\sigma\rho''_k\rho_k\sigma\rho_{k-1}\ldots\sigma\rho_1\\ &= \rho'_{r+1}\sigma\rho'_r\ldots\sigma(\rho'_{k+1}\rho''_k)\sigma\rho_k\sigma\rho_{k-1}\ldots\sigma\rho_1 \text{ by (A).}\end{aligned}$$

This word is also associated to the given path of simple moves since

$$\rho'_{k+1}\rho''_k\sigma\rho_k\ldots\sigma\rho_1(C_{i_k}^{k+1}) = \rho'_{k+1}\sigma\rho'_k\ldots\sigma\rho_1(C_{i_k}^{k+1}) = \beta_1.$$

So the induction hypothesis applies. □

Since $H(\alpha_1, \beta_1)$ is finitely generated, (A) reduces to a finite number of relations.

The remaining relations among words $\rho_{r+1}\sigma\rho\ldots\sigma\rho_1$ arise from homotopies between associated edge-paths. These are of four types arising from null-homotopies of the cycles:

(0) $$\langle C_{i_k}^{k-1}\rangle \to \langle C_{i_k}^k\rangle \to \langle C_{i_k}^{k-1}\rangle$$

Since we may choose $\rho_{k+1} = 1$, this yields the relation

$$\sigma^2 \in H. \tag{B}$$

(I) $$\langle C_{i_k}^{k-1}\rangle \to \langle C_{i_k}^k\rangle \to \langle C_{i_k}^{k+1}\rangle \to \langle C_{i_k}^{k-1}\rangle$$

The images of these three circles under $\rho_k\sigma\rho_{k-1}\ldots\sigma\rho_1$ are α_1, β_1, and some circle γ. Rotation of the subsurface of M in Fig. 17 by 120° clockwise, damped off to the

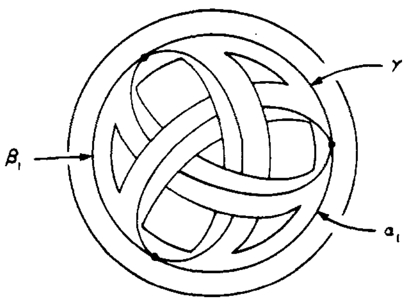


Fig. 17.

identify at the boundary, has the form $\rho_\gamma\sigma$ for some $\rho_\gamma \in H$, uniquely determined by γ.

For the word associated to the three simple moves in question we may choose $\sigma\rho_r \ldots (\rho_\gamma\sigma)^3\rho_k\sigma\rho_{k-1} \ldots \sigma\rho_1$. So the cycle (I) yields:

$$(\rho_\gamma\sigma)^3 \in H. \tag{C}$$

There are only finitely many choices of γ, modulo the choice of ρ_k. So in the presence of (A), there are just a finite number of relations in (C).

(II) $\langle C^{k-1}_{i_k}\rangle \to \langle C^k_{i_k}\rangle$ commuting with $\langle C^k_{i_{k+1}}\rangle \to \langle C^{k+1}_{i_{k+1}}\rangle$.

We may choose ρ_k so that $\rho_k\sigma\rho_{k-1} \ldots \sigma\rho_1$ carries C^k_{ik+1} to α_2 and $C^{k+1}_{i_{k+1}}$ to β_2. Then (II) becomes:

σ commutes with $\rho^{-1}\sigma\rho$, where $\rho \in H$ satisfies $\rho(\alpha_2) = \alpha_1$, $\rho(\beta_2) = \beta_1$. (D)

In the presence of (A), a single choice of such a ρ suffices.

(III) $\langle C^{k-1}_{i_k}, C^k_{i_{k+1}}\rangle \to \langle C^k_{i_k}, C^k_{i_{k+1}}\rangle \to \langle C^k_{i_k}, C^{k+1}_{i_{k+1}}\rangle \to \langle C^{k+2}_{i_k}, C^{k+1}_{i_{k+1}}\rangle \to \langle C^{k+2}_{i_k}, C^{k-1}_{i_k}\rangle \to \langle C^k_{i_{k+1}}, C^{k-1}_{i_k}\rangle$.

Suppose these five simple moves give the subword $\sigma\rho_{k+4} \ldots \sigma\rho_k$. The desired relation is then:

$$\sigma\rho_{k+4} \ldots \sigma\rho_k \in H. \tag{E}$$

We may choose ρ_k so that the cycle of five circles $C^{k-1}_{i_k}, C^k_{i_k}, C^{k+2}_{i_k}, C^k_{i_{k+1}}, C^{k+1}_{i_{k+1}}$ is carried by $\rho_k\sigma\rho_{k-1} \ldots \sigma\rho_1$ to the chain of circles α_1, β_1, α_2, γ in Fig. 18.

Modulo the choice of ρ_k, there are just finitely many possibilities for the position of β and γ on M-$(\alpha_3 \cup \ldots \cup \alpha_g)$. For each of these finitely many choices a single choice of $\rho_k, \rho_{k+1}, \ldots, \rho_{k+4}$ suffices. □

Proof of Proposition 2.1: Let $\mathrm{Diff}^+(M;\{\alpha_i\}) \subset \mathrm{Diff}^+(M)$ be the subgroup of (orientation preserving) diffeomorphisms which leave $\alpha_1 \cup \ldots \cup \alpha_g$ invariant.

LEMMA 2.4. *$\pi_0\mathrm{Diff}^+(M;\{\alpha_i\}) \to \pi_0\mathrm{Diff}^+(M)$ is injective. Hence $H \approx \pi_0\mathrm{Diff}^+(M;\{\alpha_i\})$.*

Proof. Restriction of elements of $\mathrm{Diff}^+(M)$ to $\alpha_1 \cup \ldots \cup \alpha_g$ gives a fibration

$$\mathrm{Diff}^+(M;\{\alpha_i\}) \to \mathrm{Diff}^+(M) \to B$$

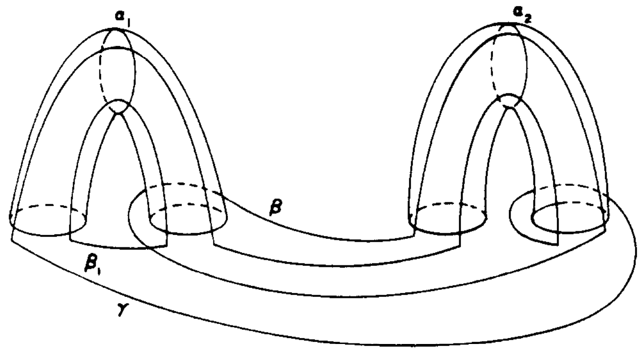


Fig. 18.

where B is the space of smooth submanifolds of M consisting of g disjoint circles whose union does not separate M. If $g > 1$, the components of B are contractible (simply-connected is all we need here; see [8]) and the result follows. □

Restriction of elements of $\mathrm{Diff}^+(M;\{\alpha_i\})$ to $\alpha_1 \cup \ldots \cup \alpha_g$ gives a fibration

$$\mathrm{Diff}^+(M \text{ rel } \{\alpha_i\}) \to \mathrm{Diff}^+(M;\{\alpha_i\}) \to \mathrm{Diff}(\{\alpha_i\})$$

whose exact sequence of homotopy groups ends with

$$Z^g \to \pi_0 \,\mathrm{Diff}\,(M \text{ rel } \{\alpha_i\}) \to H \to \pm \textstyle\sum_g \to 0.$$

LEMMA 2.5. $\pi_0 \,\mathrm{Diff}\,(M \text{ rel } \{\alpha_i\}) \approx Z^{2g-1} \oplus P_{2g-1}$.

Proof. Cutting open M along the circles α_1 produces a sphere with $2g$ holes, or a disc with $2g-1$ holes, say $D^2 - (\mathring{D}_1 \cup \ldots \cup \mathring{D}_{2g-1})$, where α_i corresponds to $\partial D_{2i-1} \cup \partial D_{2i}$, $i < g$, and α_g corresponds to $\partial D_{2g-1} \cup \partial D^2$. Diff $(M$ rel $\{\alpha_i\})$ can then be identified with the fiber of the fibration

$$\mathrm{Diff}\,(D^2 \text{ rel } \{D_1, \ldots, D_{2g-1}, \partial D^2\}) \to \mathrm{Diff}\,(D^2 \text{ rel } \partial D^2) \to E$$

where E is the space of orientation-preserving embeddings of $2g-1$ disjoint discs $D_1, \ldots, D_{2g-1}$ in $\mathring{D}^2$. Since $\mathrm{Diff}\,(D^2 \text{ rel } \partial D^2)$ is contractible, we obtain π_0 $\mathrm{Diff}\,(M \text{ rel } \{\alpha_i\}) \approx \pi_1 E$.

There is a map $d\colon E \to [SL(2,\mathrm{R})]^{2g-1} \times P_{2g-1}(\mathring{D}^2)$, obtained by taking the differential of $f\colon D_1 \cup \ldots \cup D_{2g-1} \to \mathring{D}^2$ at the centerpoints of the D_i's. Here $P_{2g-1}(\mathring{D}^2)$ is the configuration space of distinct $(2g-1)$-tuples of points in $\mathring{D}^2$. By definition, $\pi_1 P_{2g-1}(\mathring{D}^2) = P_{2g-1}$, the pure braid group. Standard methods of differential topology show that the map d is a (weak) homotopy equivalence. Hence $\pi_1 E \approx Z^{2g-1} \oplus P_{2g-1}$, as desired. □

We now have an exact sequence

$$Z^g \xrightarrow{\theta} Z^g \xrightarrow{\theta} Z^{2g-1} \oplus P_{2g-1} \to H \to \pm \sum_g \to 0.$$

It remains to describe the map θ. A basis $\{s_i\}$ for Z^g corresponds to rotations of the circles α_i through 360°, while a basis $\{t_i\}$ for Z^{2g-1} corresponds to Dehn–Lickorish twists along the circles ∂D_i. In these bases, θ sends s_i to $t_{2i}\, t_{2i-1}^{-1}$ for $i < g$. In particular, $g-1$ summands of Z^g can be split off from the exact sequence above. Also, $\theta(s_g) = t\, t_{2g-1}^{-1}$ where t is the twist along ∂D^2. □

REFERENCES

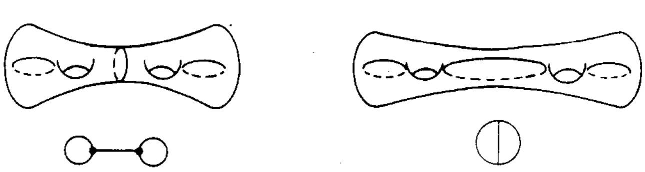
1. P. BERGAU and J. MENNICKE, Über topologische Abbildungen der Bretzelfläche vom Geschlect 2. *Math. Z.* **74** (1960), 414–435.
2. J. BIRMAN and H. HILDEN, On mapping class groups of closed surfaces as covering spaces. In: *Advances in the theory of Riemann surfaces. Ann. Math. Stud.* **66** (1971), 81–115.
3. M. DEHN, Die Gruppe der Abbildungsklassen. *Acta Math.* **69** (1938), 135–206.
4. W. B. R. LICKORISH, A finite set of generators for the homotopy group of a 2-manifold. *Proc. Camb. Phil. Soc.* **60** (1964), 769–778.
5. J. CERF, Sur les difféomorphismes de la sphère de dimension trois ($\Gamma_4 = 0$), *Springer Lecture Notes* No. 53 (1968).
6. J. CERF, La stratification naturelle ... *Publ. Math. I.H.E.S.* **39** (1970), 5–173.
7. A. F. MÖBIUS, Theorie der elementaren Verwandschaft. *Werke*, Bd. 2.
8. A. GRAMAIN, Sur le type d'homotopie du groupe de difféomorphismes d'une surface compact. *Ann. Scient. Econ. Norm. Sup.* **6** (1973), 53–66.
9. S. HUMPHRIES, Generators for the mapping class group (to appear).
10. J. MCCOOL, Some finitely presented subgroups of the automorphism group of a free group. *J. Algebra* **35** (1975), 205–213.
11. J. BIRMAN, Braids, links and mapping class groups. *Ann. Math. Stud.* **82** (1975).
12. C. EARLE and A. MARDEN, (to appear).

University of California
Los Angeles, CA 90024, U.S.A.

APPENDIX

Markings

A maximal collection of disjoint, non-contractible, pairwise non-isotopic smooth circles on M we call a *marking*. Each complementary component must then be a *trinion*†, or thrice-punctured sphere. So by Euler characteristic considerations, there are $3g-3$ circles in the marking and $2g-2$ complementary trinions. To a marking is associated a finite graph, whose vertices correspond to the complementary trinions, and whose edges correspond to the circles of the marking. This graph is trivalent, three edges (not necessarily distinct) meeting at each vertex. There is a one-to-one correspondence between such finite trivalent graphs and (ambient) diffeomorphism classes of markings. For example, in genus two there are just two non-diffeomorphic markings:



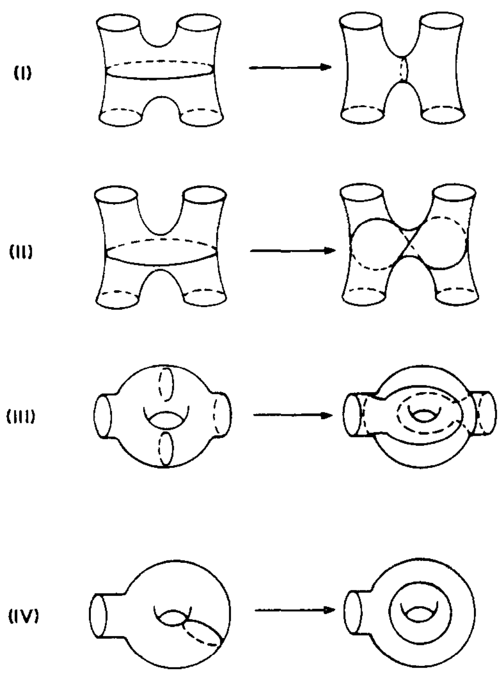
We can change markings by the *simple moves* (I)–(IV) below.

Note that only (I) produces a net change in the associated graph.

PROPOSITION. *Any two markings of M are obtainable one from the other by a finite sequence of simple moves of these four types (up to isotopy).*

Sketch of Proof. To a generic function $f: M \to \mathbb{R}$ we associate a marking $\mu(f)$ as follows. In the graph $\Gamma(f)$ let $\gamma(f)$ be the unique smallest subgraph to which $\Gamma(v)$ collapses (i.e. $\gamma(f)$ has no univalent vertices. Regarding bivalent vertices of $\gamma(f)$ not as vertices at all, then $\gamma(f)$ is trivalent). Lifting a point of each edge of $\gamma(f)$ to a level curve of f gives the marking $\mu(f)$. This is uniquely determined by f, up to isotopy. Note that $\gamma(f)$ is the graph associated to $\mu(f)$.

†This odd term is due to Möbius [7], following his use of "union" for "disc" and "binion" for "annulus".



Every marking arises in this fashion—just define f on disjoint product neighborhoods of the circles of the marking to be projection onto the normal direction, then extend over the complement of these neighborhoods in any generic way. Thus any two given markings have the form $\mu(f_0)$ and $\mu(f_1)$. Join f_0 to f_1 by a generic path $f_t\colon M \to \mathbf{R}$. Clearly, birth–death points and crossing points in the graphic of f_t, other than those in Figs. 5 and 6, have no effect on $\mu(f_t)$. Examining the crossings in Figs. 5 and 6, one can easily check that the markings $\mu(f_t)$ are changed by the simple moves (I)–(IV).

It seems that the rest of the program of §1 can also be carried over to markings, to obtain again a complete set of relations between simple moves. However, to apply this to the mapping class group there is now the added difficulty that there is not one standard model for markings, but finitely many (for each genus), one for each type of associated trivalent graph.

Remark. Moves (I) and (IV) in fact suffice.