AN ALGORITHM FOR SIMPLE CURVES ON SURFACES

JOAN S. BIRMAN AND CAROLINE SERIES

1. *Introduction*

Let M be a compact orientable surface with non-empty boundary and with $\chi(M) < 0$, and let $\Gamma = \pi_1 M$. Let $\hat{C}$ be the free homotopy class of a closed loop on M and let $W = W(\hat{C})$ be a word in a fixed set of generators $\bar{\Gamma}$ which represents $\hat{C}$. In this paper we give an algorithm to decide, starting with W, whether $\hat{C}$ has a simple representative, that is a representative without self-intersections. Such a word will be said to be *simple*. As an application, we begin a study of simple words in $\bar{\Gamma}$. Our results also apply to infinite geodesics on M, corresponding to biinfinite words in $\bar{\Gamma}$, where now we ask which finite blocks appear in such a word when the corresponding infinite homotopy class has a simple representative.

For finite words there are, of course, other such algorithms, see for example [**8**, **9**, **2**, **3**, **4**]. Our algorithm most resembles that in [**2**] in that it is purely mechanical and combinatorial. It is simpler than that in [**2**] but what is more important is that it reveals the underlying mechanism which determines whether self-intersections occur; the combinatorics of that mechanism seem quite interesting and non-trivial.

We represent M as U/Γ where $U \subseteq \mathbb{D}$ is the universal covering space of M, and where $\mathbb{D}$ is the unit disc with the Poincaré metric and Γ is a discrete group of hyperbolic isometries. Poincaré showed in [**7**] that $\hat{C}$ contains a simple representative if and only if the unique smooth geodesic representative C of $\hat{C}$ is simple, and that C is simple if and only if for each lift γ of C to $\mathbb{D}$ the curves in the infinite family $\{f\gamma\}_{f\in\Gamma}$ are pairwise disjoint. Now, to see if geodesics $\gamma_1, \gamma_2 \in \{f\gamma\}$ are disjoint in $\mathbb{D}$ it is enough to know whether the ideal endpoints of γ_1 on $\partial\mathbb{D}$ separate those of γ_2. Crucial to our work is a scheme for parametrizing points on $\partial\mathbb{D}$ by infinite words in $\bar{\Gamma}$, first developed by Nielsen in [**6**]. The idea of this paper is to show how information on the order of the points $\partial\gamma_1, \partial\gamma_2$ on $\partial\mathbb{D}$ is encoded in Nielsen's 'boundary expansion' (Theorem A) and then to examine consequences.

When $\partial M \neq \varnothing$ the group Γ is a free group so that each conjugacy class has a unique shortest representative which is obtained by cyclic reduction of any word in the class. However, if $\partial M = \varnothing$ the shortest word in the conjugacy class is in general not unique. If $\partial M = \varnothing$ and $W \in \Gamma$ has a shortest representative which does not contain any pieces which are half of the defining relator in Γ, then the problem of deciding whether W is simple is identical with that on the surface with a disc removed, that is one simply regards Γ as if it were a free group. On the other hand, the exceptional cases when W contains half a relator involve some subtle points which are not without interest, but are somewhat tangential to the main idea in the paper. For that reason, we shall omit the case in which $\partial M = \varnothing$.

Here is an outline of this paper. The tools we need are set up in §§2 and 3 where we prove Theorem A. The algorithm (Theorem B) is given in §4. In §5 we give

Received 11 February, 1983.

The first author was partially supported by N.S.F. grant no. MCS 79-04715.

applications. In the special case of a surface of genus 1 with a single boundary component the group Γ is free of rank 2 and so, as will be shown, the set of simple words in $\bar{\Gamma}$ coincides with the set of all generators of Γ. This situation was studied in a recent paper of Cohen, Metzler and Zimmerman [5], and our Theorem C is a direct generalization of their result, presented here as Theorem 5.1.

The first author acknowledges a travel grant from the Barnard College Faculty Research Committee.

2. *The group* Γ

The problem of whether a word W in the generators of Γ determines a free homotopy class on M which has a simple representative is a problem about Γ, and M, for which we use the techniques of hyperbolic geometry. We are free to choose, for each topological surface M, the most convenient group Γ such that U/Γ is homeomorphic to M. Let us define a convenient class of groups.

Draw $p = 4g+2b-2$ symmetrically placed disjoint geodesic arcs in $\mathbb{D}$ with their endpoints on $\partial\mathbb{D}$, as in Figure 1. Label these $C_1, \ldots, C_p$ in anticlockwise order. Choose transformations $s_j \in \operatorname{Isom} \mathbb{D}$, $j = 1, \ldots, p/2$ which identify these circles in pairs in such a way that $s_j(C_\tau) = C_\rho$; then C_τ is the isometric circle of s_j. Let Γ be the group generated by $s_1, \ldots, s_{p/2}$.

Let U be the convex hull in $\mathbb{D}$ of the limit set of Γ. Then for appropriate choice of $s_1, \ldots, s_{p/2}$ the surface U/Γ will have genus $g \geqslant 0$, and $b > 0$ boundary components. For example, if $g = 0$ we may choose the s_j so that $s_j(C_{2b-j-1}) = C_j$, $1 \leqslant j \leqslant b-1 = p/2$. For $g > 0$, $b > 0$ we could choose $s_j(C_{2g+b+j-1}) = C_j$, for $1 \leqslant j \leqslant 2g-1$ and $s_j(C_{6g+2b-j-2}) = C_j$ for $2g \leqslant j \leqslant 2g+b-1$. These particular choices are illustrated in Figures 1 and 2. The curves representing these generators on M are shown in Figure 3. The region outside the circles C_j is a fundamental region for Γ which we denote by R.

The symbol $\bar{\Gamma}$ will be used to denote the set of generators and their inverses, that is $\bar{\Gamma} = \{s_1, \ldots, s_{p/2}, \bar{s}_1, \ldots, \bar{s}_{p/2}\}$. If $x \in \bar{\Gamma}$, we shall sometimes write x^{-1} and sometimes $\bar{x}$ for the inverse of x. The symbol s_j^ε, $\varepsilon = \pm 1$, will be understood to mean $s_j = s_j^{+1}$ or s_j^{-1}. A *word* in $\bar{\Gamma}$ means a word in the symbols of $\bar{\Gamma}$. The *equivalence class* of a word is the set of all words which represent the same element of the group Γ. The symbol $\Gamma(g, b)$ will always be understood to mean the specific groups illustrated in Figures 1 and 2. The letter O denotes the origin in $\mathbb{D}$. Elements in Γ are composed from right to left, that is if $e_1, e_2 \in \Gamma$ then $e_1 e_2 O$ means $e_1(e_2(O))$.

A finite word $w = e_1 e_2 \ldots e_n$, $e_j \in \bar{\Gamma}$, is said to be *reduced* if $e_j \neq \bar{e}_{j+1}$ for every $j = 1, \ldots, n-1$ and *cyclically reduced* if it is reduced and $e_n \neq \bar{e}_1$. An infinite word $e_1 e_2 \ldots$ or a biinfinite word $\ldots e_{-1} e_0 e_1 e_2 \ldots$ is reduced if each of its finite subwords is reduced.

3. *Boundary expansions and cyclic lexicographical ordering*

Let Γ be one of the groups we are considering, and recall that we defined Γ by specifying isometric circles for the generators of Γ. Label the arc cut off on $\partial\mathbb{D}$ by the isometric circle of $\bar{e}_i$, $e_i \in \bar{\Gamma}$, by $[e_i]$. We call these arcs the first order intervals on $\partial\mathbb{D}$.

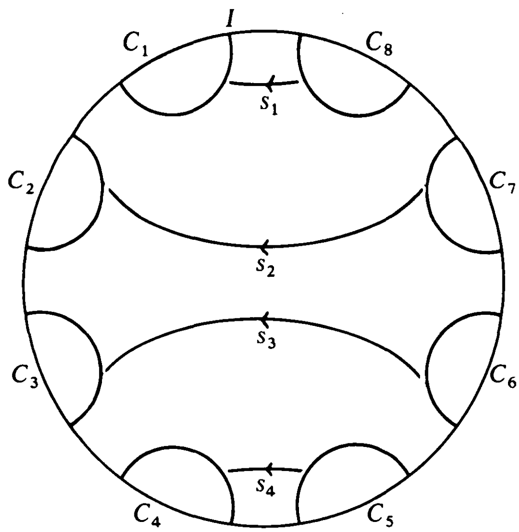

FIG. 1.

$g = 0$
$b = 5$

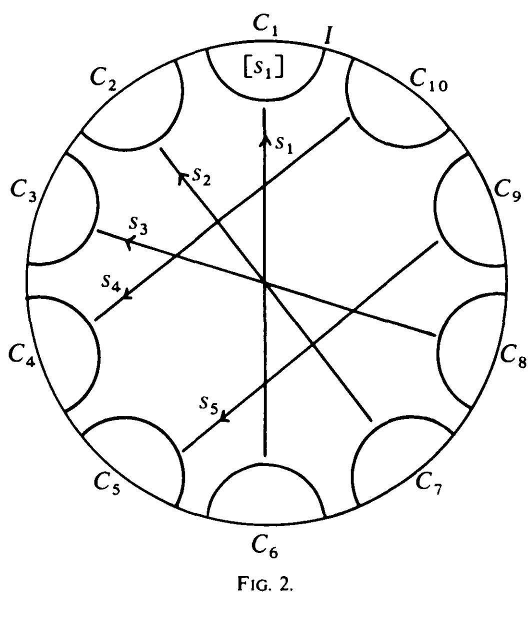

FIG. 2.

$g = b = 2$

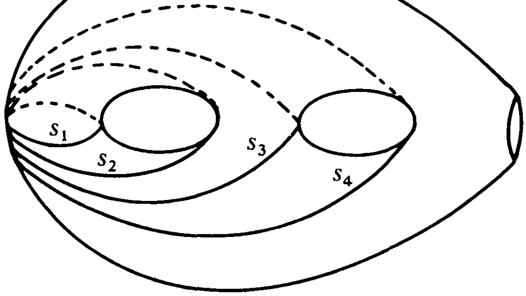

FIG. 3. $M(2, 1)$

For each positive integer m and each reduced word $e_1 \dots e_m$ in $\bar{\Gamma}$ define the m-th order interval $[e_1 \dots e_m]$ by

$$[e_1 \dots e_m] = e_1 \dots e_{m-1}[e_m].$$

We claim that the set of m-th order intervals is disjoint and that

$$[e_1 \dots e_m] \subset [e_1 \dots e_{m-1}] \subset \dots \subset [e_1].$$

Since for $2 < r \leqslant m$,

$$[e_1 \dots e_r] = e_1 \dots e_{r-1}[e_r] = e_1 \dots e_{r-2}[e_{r-1} e_r]$$

and

$$[e_1 \dots e_{r-1}] = e_1 \dots e_{r-2}[e_{r-1}],$$

by applying $(e_1 \dots e_{r-2})^{-1}$ it is enough to see that $e_i[e_j] \subset [e_i]$ whenever $e_i, e_j \in \bar{\Gamma}$ and $e_i \neq \bar{e}_j$. Now since $\bar{e}_j \neq e_i$, $[e_j]$ lies outside the isometric circle of e_i and so is mapped by e_i into the isometric circle of $\bar{e}_i$. Thus $e_i[e_j] \subset [e_i]$.

Assume inductively that all the $(m-1)$-th order intervals are disjoint. Then for fixed $x \in \bar{\Gamma}$, all the m-th order intervals $[xe_2 \dots e_m]$ with $e_2 \neq \bar{x}$ are disjoint. Since by the above $[xe_2 \dots e_m] \subset [x]$, and since $[x] \cap [y] = \varnothing$, for distinct $x, y \in \bar{\Gamma}$, the result follows.

It is not hard to prove, and follows as a special case of 4.9, 4.10 in [**8**], that if $e_1 e_2 \dots$ is an infinite reduced word in $\bar{\Gamma}$ then

(1) $\displaystyle\bigcap_{m=1}^{\infty} [e_1 \dots e_m] = \lim_{m \to \infty} e_1 \dots e_m O,$

(2) the set $\left\{\xi = \lim_{m\to\infty} e_1 \dots e_m O : e_1 e_2 \dots \text{ is an infinite reduced word}\right\}$ is precisely the limit set Λ of Γ on $\partial\mathbb{D}$. This representation of points in Λ is unique.

From now on we write $\xi = e_1 e_2 \dots$ if $\xi = \lim_{m\to\infty} e_1 \dots e_m O$ and refer to $e_1 e_2 \dots$ as the *boundary expansion* of ξ.

An *alphabet* is a finite ordered set of distinct symbols. A *cyclic alphabet* is a cyclically ordered set of distinct symbols. A cyclic alphabet, say $A = \{x_1, \dots, x_n\}$, becomes an alphabet A_{x_j} on choosing one of the symbols $x_j \in A$ as an initial letter. Thus to the cyclic alphabet A, we associate n distinct alphabets $A_{x_1}, \dots, A_{x_n}$.

Assign to Γ the cyclic alphabet whose letters are the symbols in the generating set $\bar{\Gamma}$ arranged in the order in which the first order intervals occur around $\partial\mathbb{D}$ anticlockwise. For example, if $\Gamma = \Gamma(g, b)$ we have

$$A(g,b) = \begin{cases} \{s_1, s_2, \dots, s_{2g+b-1}, \bar{s}_1, \bar{s}_2, \dots, \bar{s}_{2g-1}, \bar{s}_{2g+b-1}, \bar{s}_{2g+b-2}, \dots, \bar{s}_{2g}\}, & g \neq 0, \\ \{s_1, \dots, s_{b-1}, \bar{s}_{b-1}, \dots, \bar{s}_1\}, & g = 0. \end{cases}$$

THEOREM A. *Let P, Q be distinct points on $\partial\mathbb{D}$ with boundary expansions $e_1 e_2 \ldots, f_1 f_2 \ldots$. Then P precedes Q in anticlockwise order around $\partial\mathbb{D}$ starting from the point I (see Figures 1 and 2) if and only if either*

(i) *e_1 precedes f_1 in the alphabet A_{s_1}, or*

(ii) *$e_i = f_i$ for each $i = 1, \ldots, m$ and e_{m+1} precedes f_{m+1} in the alphabet $A_{\bar{e}_m}$.*

Proof. By the definition of boundary expansions we have $P \in [e_1 \ldots e_m]$ and $Q \in [f_1 \ldots f_m]$ for each m. If $e_1 \neq f_1$ then $P \in [e_1]$, $Q \in [f_1]$, and so since A_{s_1} lists the first order intervals anticlockwise round $\partial\mathbb{D}$ starting at the point I we obtain (i).

Suppose now that $e_i = f_i$ for $i \leqslant m$ and $e_{m+1} \neq f_{m+1}$. Then $P, Q \in [e_1 \ldots e_m]$. Let $g = (e_1 \ldots e_m)^{-1}$. Then $g[e_1 \ldots e_m] = e_m^{-1}[e_m]$, and $gP \in [e_{m+1}]$, $gQ \in [f_{m+1}]$. Moreover $e_m^{-1}[e_m]$ is an interval on $\partial\mathbb{D}$ outside $[\bar{e}_m]$. Now the anticlockwise order of P, Q around $\partial\mathbb{D}$ starting from I is the same as the anticlockwise order of P, Q in $[e_1 \ldots e_m]$. This is the same as the anticlockwise order of gP, gQ in $g[e_1 \ldots e_m]$, which by the above observations may be read off as the same as the anticlockwise order of $[e_{m+1}]$, $[f_{m+1}]$ round $\partial\mathbb{D}$ starting at $[\bar{e}_m]$; that is, the order of e_{m+1}, f_{m+1} in the alphabet $A_{\bar{e}_m}$.

The rule for ordering points on $\partial\mathbb{D}$ described in Theorem A we shall call the *cyclic lexicographical ordering*. Obviously it depends on the choice of Γ and A.

We now look at the use of boundary expansions to represent geodesics in $\mathbb{D}$. Suppose that $\mathbf{e} = \ldots e_{-1} e_0 e_1 \ldots$ is a biinfinite reduced word in the generators. Consider the two points x, y whose boundary expansions are $e_1 e_2 \ldots$, $\bar{e}_0 \bar{e}_{-1} \bar{e}_{-2} \ldots$ respectively. Notice that since $x \in [e_1]$, $y \in [\bar{e}_0]$ and $\bar{e}_0 \neq e_1$ because $\mathbf{e}$ is reduced, certainly $x \neq y$. Thus we have the following.

(3.1) Each reduced biinfinite word $\mathbf{e}$ determines a unique oriented geodesic with positive endpoint $e_1 e_2 \ldots$ and negative endpoint $\bar{e}_0 \bar{e}_{-1} \ldots$. This geodesic we denote by $\gamma(\mathbf{e})$.

Since by definition of the boundary expansions $e_1 e_2 \ldots = \lim_{n \to \infty} e_1 \ldots e_n O$, we have the following.

(3.2) If $x \in \partial\mathbb{D}$ has boundary expansion $e_1 e_2 \ldots$, then $\bar{e}_1 x$ has boundary expansion $e_2 e_3 \ldots$. If $f \in \bar{\Gamma}$, $f \neq \bar{e}_1$, then fx has boundary expansion $f e_1 e_2 \ldots$.

As a consequence we obtain the following.

(3.3) Let $\mathbf{e} = \ldots e_{-1} e_0 e_1 e_2 \ldots$ be reduced. Let $\sigma^n(\mathbf{e})$, $n \in \mathbb{Z}$, be the sequence whose j-th entry is in position $j+n$ in $\mathbf{e}$. Then $\gamma(\sigma^n \mathbf{e}) = (e_1 \ldots e_n)^{-1} \gamma(\mathbf{e})$.

Finally, let $\mathbf{W} = \ldots WWW \ldots$ where W is a cyclically reduced word. By (3.2) the endpoints of $\gamma(\mathbf{W})$ are fixed by W and W^{-1}. Thus $\gamma(\mathbf{W})$ is fixed by W, since $W \in \Gamma \subset \operatorname{Isom} \mathbb{D}$. Hence we have the following.

(3.4) The projection of $\gamma(\mathbf{W})$ to M is a closed smooth geodesic with homotopy class W.

4. *The algorithm*

In this section we give our algorithm for deciding whether a cyclically reduced word or a reduced biinfinite word in $\bar{\Gamma}$ is simple. In the latter case, the procedure may involve infinitely many tests. As always, we assume that $b > 0$, and Γ is any of the groups described in §2.

Let C be a closed curve on M. Let W be the reduced word representing the image of C in $\pi_1(M)$ and let U be the cyclic reduction of W. By (3.4), $\gamma(\dots UU \dots)$ projects to a smooth geodesic C' on M with homotopy class U. Since W and U are conjugate in $\pi_1(M)$, C and C' are in the same free homotopy class and therefore W is simple if and only if the same is true of U. Thus it is sufficient to test cyclically reduced words for simplicity.

If C is an infinite geodesic on M then C lifts to a geodesic in $\mathbb{D}$ passing through the fundamental region R defined in §2.

Let the positive and negative endpoints of this geodesic on $\partial\mathbb{D}$ be $e_1 e_2 \dots$ and $\bar{e}_0 \bar{e}_{-1} \dots$ respectively. It is clear that these points lie in distinct first order intervals on $\partial\mathbb{D}$, and thus that $\dots e_{-1} e_0 e_1 \dots$ is a biinfinite reduced word in $\bar{\Gamma}$. Thus every infinite geodesic on M corresponds to a biinfinite reduced word in $\bar{\Gamma}$ which may be tested for simplicity using the algorithm. (The lift of C to a geodesic intersecting R is obviously not unique. As will be apparent from the proof of Theorem B, the different possible lifts all correspond to shifts $\sigma^n(\dots e_{-1} e_0 e_1 \dots)$ of the same sequence $\dots e_{-1} e_0 e_1 \dots$.)

Fix the generating set $\bar{\Gamma}$ to be $\{s_1, \dots, s_{p/2}, \bar{s}_1, \dots, \bar{s}_{p/2}\}$. Our cyclic alphabet is the cyclically ordered set defined in §3, for example if Γ is one of the groups $\Gamma(g, b)$, then A is $A(g, b)$. A word $W = e_1 e_2 \dots$ will be said to precede a word $W' = e'_1 e'_2 \dots$, written $W < W'$, if W precedes W' in the cyclic lexicographic ordering of Theorem A, that is if $e_1 e_2 \dots$ precedes $e'_1 e'_2 \dots$ in anticlockwise order around $\partial\mathbb{D}$ from I.

A *cut* in the reduced biinfinite word $\mathbf{W} = \dots e_{-1} e_0 e_1 e_2 \dots e_{j-1} e_j e_{j+1} \dots$ is a subdivision of $\mathbf{W}$ into a left half and a right half, and will be indicated by the symbol $\dots e_{j-2} e_{j-1} \mid e_j e_{j+1} \dots$. Each cut in $\mathbf{W}$ determines two reduced infinite words $W = e_j e_{j+1} \dots$ and $W' = \bar{e}_{j-1} \bar{e}_{j-2} \dots$. We shall sometimes refer to these as the *right* and *left* words at the cut.

THEOREM B (the algorithm).

Part 1, finite words. *Let $W = e_1 e_2 \dots e_r$ be a finite, non-periodic cyclically reduced word in $\bar{\Gamma}$. Let $W_j, j = 1, \dots, r$, denote the r cyclic permutations of W and let W_j^{-1} be the inverse of W_j. Order these $2r$ words by the cyclic lexicographical ordering rule so that $W_{\mu_1}^{\varepsilon_1} < W_{\mu_2}^{\varepsilon_2} < \dots < W_{\mu_{2r}}^{\varepsilon_{2r}}$. Let $X = W_{\mu_1}^{\varepsilon_1} W_{\mu_2}^{\varepsilon_2} \dots W_{\mu_{2r}}^{\varepsilon_{2r}}$. Then, thinking of X as a word in the free group with free basis $W_1, \dots, W_r$, the word W is simple if and only if X can be cyclically reduced to the empty word.*

Part 2, biinfinite words. *Let $\mathbf{W} = \dots e_1 e_0 e_1 e_2 \dots$ be a reduced biinfinite word. Choose cuts in $\mathbf{W}$, say $\dots e_{j-2} e_{j-1} \mid e_j e_{j+1} \dots$ and $\dots e_{k-2} e_{k-1} \mid e_k e_{k+1} \dots$. Let W_j, W_j^{-1}, W_k, W_k^{-1} be the right and left words associated to these cuts. Then $\mathbf{W}$ is simple if and only if for each pair of cuts the pair W_j, W_j^{-1} does not separate W_k, W_k^{-1} with respect to the cyclic lexicographical ordering on $\partial\mathbb{D}$.*

EXAMPLES. Let $g = 2$, $b = 1$, so that $A = s_1, s_2, s_3, s_4, \bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4$.

1. Let $W = s_1\bar{s}_2s_3$. The six words which we must order are

$$W_1 = s_1\bar{s}_2s_3, \qquad W_1^{-1} = \bar{s}_3s_2\bar{s}_1,$$

$$W_2 = \bar{s}_2s_3s_1, \qquad W_2^{-1} = \bar{s}_1\bar{s}_3s_2,$$

$$W_3 = s_3s_1\bar{s}_2, \qquad W_3^{-1} = s_2\bar{s}_1\bar{s}_3.$$

The ordering gives $W_1 < W_3^{-1} < W_3 < W_2^{-1} < W_2 < W_1^{-1}$ (cf. Figure 4). Since $W_1W_3^{-1}W_3W_2^{-1}W_2W_1^{-1} \sim \varnothing$, W is simple.

2. Let $\mathbf{W} = \dots(s_2s_2\bar{s}_3)(s_2s_2\bar{s}_3)\dots$,

$$W_1 = s_2s_2\bar{s}_3\dots, \qquad W_1^{-1} = s_3\bar{s}_2\bar{s}_2\dots,$$

$$W_2 = s_2\bar{s}_3s_2\dots, \qquad W_2^{-1} = \bar{s}_2s_3\bar{s}_2\dots,$$

$$W_3 = \bar{s}_3s_2s_2\dots, \qquad W_3^{-1} = \bar{s}_2\bar{s}_2s_3\dots$$

and $W_2 < W_1 < W_1^{-1} < W_2^{-1} < W_3^{-1} < W_3$. (Note: W_2 precedes W_1 because $\bar{s}_3$ precedes s_2 in the alphabet $A_{\bar{s}_2}$; similarly, W_2^{-1} precedes W_3^{-1} because s_3 precedes $\bar{s}_2$

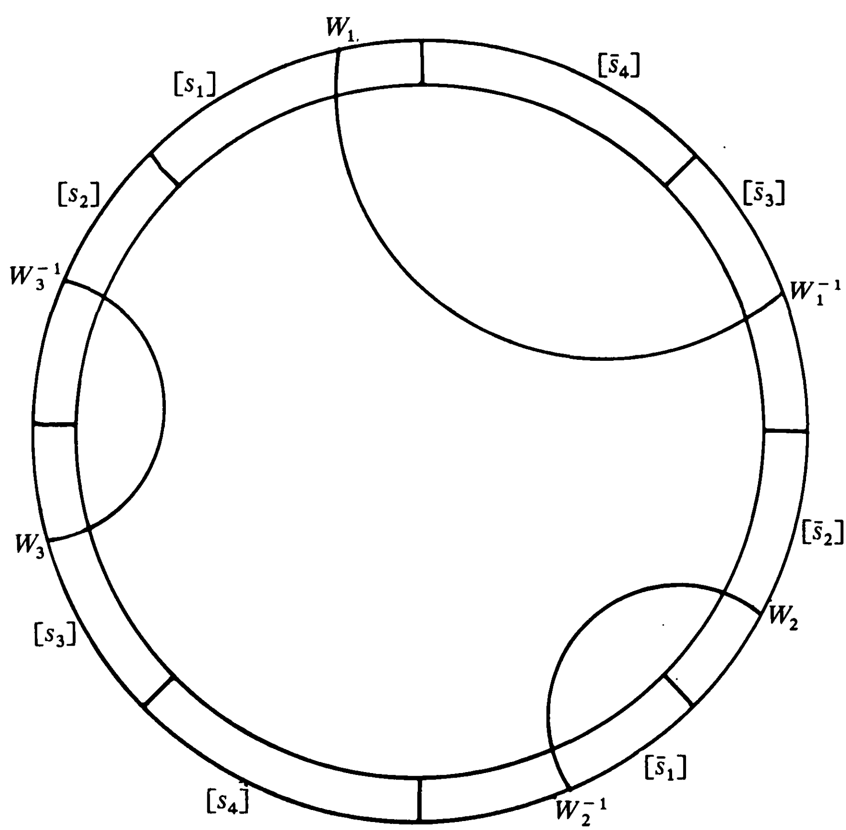

FIG. 4.

in the alphabet A_{s_2}). Since no pair W_j, W_j^{-1} separates any pair W_k, W_k^{-1}, the word W is simple. Only three pairs need be checked because W is a repeating word.

3. Let $\mathbf{W} = \ldots s_1^2 s_2^2 \ldots$. Introduce the cuts $\ldots s_1 \mid s_1 \ldots, \ldots s_2 \mid s_2 \ldots$. Since $s_1, \bar{s}_1$ separate $s_2, \bar{s}_2$ in the alphabet A_{s_1}, the word $\mathbf{W}$ cannot be simple.

Proof of Theorem B. We shall first prove part 2. We then show how it specializes to part 1 by regarding W as a periodic biinfinite word $\mathbf{W} = \ldots WWW \ldots$.

We are given that $\mathbf{W} = \ldots e_{-2}e_{-1}e_0e_1e_2 \ldots$. By (3.1) $\mathbf{W}$ determines a unique geodesic $\gamma(\mathbf{W})$ in $\mathbb{D}$ whose endpoints ε, η on $\partial\mathbb{D}$ are the limits of the sequence $\{e_1e_2 \ldots e_nO,\ \bar{e}_0\bar{e}_{-1} \ldots \bar{e}_{-n}O : n \in \mathbb{N}\}$. Since, by (3.3) the geodesics $\gamma(\sigma^n(\mathbf{W}))$, $n \in \mathbb{Z}$, are all translates of $\gamma(\mathbf{W})$ by Γ, it is clear that if any two of these geodesics intersect, then $\mathbf{W}$ is not simple. The heart of the argument is to prove the converse of this statement.

The sequence $\mathbf{W}$ determines an *edge path* in $\mathbb{D}$ which consists of arcs joining the points $\ldots, \bar{e}_0\bar{e}_{-1}O, \bar{e}_0O, O, e_1O, e_1e_2O, \ldots$ in order. We denote this path by the symbol $\mathbf{W}(O)$. More generally $f\mathbf{W}(O)$, $f \in \Gamma$, denotes the edge path sequence joining the points $\ldots, f\bar{e}_0\bar{e}_{-1}O, f\bar{e}_0O, fO, fe_1O, fe_1e_2O, \ldots$; its limit points on $\partial\mathbb{D}$ are $f\varepsilon, f\eta$, where ε, η are the limit points of the original edge path sequence.

Now, the geodesic $\gamma(\mathbf{W})$ covers a geodesic C on M. If C is *not* simple then there exist distinct geodesics which cover C, say the images of $\gamma(\mathbf{W})$ under $f, h \in \Gamma$, which intersect transversally in $\mathbb{D}$. By our observations above, these geodesics have ideal endpoints on $\partial\mathbb{D}$ which are the images of ε, η under f, h respectively, also they intersect if and only if $f\varepsilon, f\eta$ separate $h\varepsilon, h\eta$ on $\partial\mathbb{D}$. On the other hand, $f\varepsilon, f\eta$ are also limits of the edge path sequence $f\mathbf{W}(O)$, and similarly $h\varepsilon, h\eta$ are limits of the edge path sequence $h\mathbf{W}(O)$, and so C is not simple if and only if for some $f, h \in \Gamma$ the edge paths $f\mathbf{W}(O), h\mathbf{W}(O)$ intersect. But then these edge paths have a common vertex, say kO, $k \in \Gamma$. Therefore the image of $f\mathbf{W}(O)$, $h\mathbf{W}(O)$ under k^{-1} have the common vertex, O, that is there are edge paths $f\mathbf{W}(k^{-1}O)$, $h\mathbf{W}(k^{-1}O)$ which intersect at O. Hence we may assume without loss of generality that $f, h \in \Gamma$ were chosen in the first place so that $f\mathbf{W}(O)$, $h\mathbf{W}(O)$ intersect at O.

Now, the vertices of $f\mathbf{W}(O)$ are the points

$$\{fe_1e_2 \ldots e_kO,\ f(e_{-k} \ldots e_1e_0)^{-1}O : k \in \mathbb{N}\}.$$

That is, $f\mathbf{W}(O)$ must be the path $\sigma^k\mathbf{W}(O)$ for some $k \in \mathbb{Z}$. Similarly $h\mathbf{W}(O)$ must be the path $\sigma^j\mathbf{W}(O)$ for some $j \in \mathbb{Z}$, that is, the family of geodesics $\{\gamma(\sigma^n\mathbf{W})\}_{n \in \mathbb{Z}}$ contains intersecting members. Thus we have proved that *the projection of $\gamma(\mathbf{W})$ on M is non-simple if and only if there are integers $j, k \in \mathbb{Z}$ such that the geodesics $\gamma(\sigma^j\mathbf{W})$, $\gamma(\sigma^k\mathbf{W})$ intersect transversely.*

Now, the geodesics $\gamma(\sigma^j\mathbf{W})$, $\gamma(\sigma^k\mathbf{W})$ intersect transversely in $\mathbb{D}$ if and only if the ideal endpoints of $\gamma(\sigma^j\mathbf{W})$ separate those of $\gamma(\sigma^k\mathbf{W})$ on $\partial\mathbb{D}$. Since $\mathbf{W} = \ldots e_{-2}e_{-1}e_0e_1e_2 \ldots$, the ideal endpoints of $\gamma(\sigma^j\mathbf{W})$ have boundary expansion $W_j = e_je_{j+1} \ldots$ and $W_j^{-1} = \bar{e}_{j-1}\bar{e}_{j-2} \ldots$. These are the words associated to a cut at e_j. Hence $\mathbf{W}$ is non-simple if and only if the points on $\partial\mathbb{D}$ which are determined by W_j, W_j^{-1} separate those determined by W_k, W_k^{-1} for some $j, k \in \mathbb{Z}$. By Theorem A the order of points on $\mathbb{D}$ is determined from their boundary expansions by our cyclic lexicographical ordering rule. Thus the algorithm is valid in the situation of biinfinite words.

To prove part 1 suppose that $\mathbf{W}$ is periodic. Let W be the word which is the minimum period of $\mathbf{W}$. It is only necessary to test finitely many pairs $\sigma^j\mathbf{W}, \sigma^k\mathbf{W}$, and these are in one to one correspondence with the cyclic permutations of W. We are then reduced to ordering the $2r$ words $W_j^\varepsilon, j = 1, \ldots, r, \varepsilon = \pm 1$ as in the statement of the theorem. These determine r geodesics on $\mathbb{D}$, say $\gamma_1, \ldots, \gamma_r$, where γ_j has endpoints W_j, W_j^{-1} with boundary expansions $W_jW_j\ldots, W_j^{-1}W_j^{-1}\ldots$. Figure 4 shows a typical picture of how the geodesics $\gamma_1, \ldots, \gamma_r$ might arrange themselves in D. The condition that no pair γ_j, γ_k, $1 \leqslant j \neq k \leqslant r$, intersect is easily seen to be equivalent to the condition that the word $W_{\mu_1}^{\varepsilon_1}W_{\mu_2}^{\varepsilon_2}\ldots W_{\mu_{2r}}^{\varepsilon_{2r}}$ be freely equal to the empty word. This completes the proof.

Remark. In Theorem A of [1] the authors prove that the number of blocks of length n which can occur in a biinfinite simple word is bounded by a polynomial in n. The proof of this fact in [1] is independent of the work in this paper, although the result is necessarily a consequence of Theorem B above, since Theorem B gives necessary and sufficient conditions for a biinfinite word to be simple. It would be interesting to have a proof of polynomial growth based upon Theorem B.

5. *Applications*

In this section we study the word forms which can occur when W is a simple word in $\Gamma(g,1)$, $g \geqslant 1$. As before, the generating set is $\bar{\Gamma} = \{s_1, \ldots, s_{2g}, \bar{s}_1, \ldots, \bar{s}_{2g}\}$, where by our choice the loop around the single boundary component is represented by

$$(s_1\bar{s}_2s_3\bar{s}_4\ldots\bar{s}_{2g})(\bar{s}_1s_2\bar{s}_3s_4\ldots s_{2g}).$$

Our work will be seen to generalize a recent result of Cohen, Metzler and Zimmerman, who in [5] studied basis elements (that is words which are generators) for the group $\Gamma(1,1)$. Since s_1, s_2 are basis elements, and since every automorphism of $\Gamma(1,1)$ is induced by a homeomorphism of $M(1,1)$, every basis element of $\Gamma(1,1)$ is the homotopy class of some simple loop on $M(1,1)$. Conversely, if C is a simple loop which does not separate $M(1,1)$, then, by classification of surfaces, C is the image of a standard generator under some homeomorphism of $M(1,1)$; thus the homotopy class of C determines a basis element. Finally, since there is only one homotopy class in $\Gamma(1,1)$ which is represented by a separating curve, namely $s_1\bar{s}_2\bar{s}_1s_2$, we have the following restatement of the main result of [5].

Theorem 5.1 [5]. *Up to permutations of the generators which interchange s_1 and s_2, s_1 and $\bar{s}_1$, or s_2 and $\bar{s}_2$ a simple word $W^{\pm1}$ in $\Gamma(1,1)$ is up to cyclic permutations either s_1 or $s_1\bar{s}_2\bar{s}_1s_2$ or has the form*

$$W = s_1^{n_1}s_2s_1^{n_2}s_2\ldots s_1^{n_k}s_2$$

where $\{n_1, n_2, \ldots, n_k\} \subseteq \{n, n+1\}$ for some $n \in \mathbb{Z}^+$.

Remark 5.2. If one defines an automorphism ϕ of $\Gamma(1,1)$ by $\phi(s_1) = s_1$, $\phi(s_2) = \bar{s}_1^n s_2$, then $\phi(W)$ has length strictly less than W and s_1 occurs with exponent ±1. As observed in [5] this fact, in conjunction with the Euclidean algorithm, yields a recursive description which can be used to enumerate all simple words in $\Gamma(1,1)$. Compare this with Corollary 5.3 and Remark 5.4 below.

We now state our generalization of Theorem 5.1.

THEOREM C. *Let* $\mathbf{W}$ *be a reduced biinfinite in the generators of* $\Gamma(g, 1)$, $g \geqslant 1$.

(1) $\mathbf{W}$ *is simple if and only if its image under each of the automorphisms* $\tau^a \rho^b$ *of* Γ *is simple, where* ρ *is the cyclic permutation* $(s_1, \bar{s}_2, s_3, \bar{s}_4, \ldots, \bar{s}_{2g}, \bar{s}_1, s_2, \bar{s}_3, s_4, \ldots, s_{2g})$ *and* τ *is the involution* $(s_1, \bar{s}_1)(s_2, \bar{s}_2) \ldots (s_{2g}, \bar{s}_{2g})$.

(2) $\mathbf{W}$ *is simple only if at most one of the letters* s_j, $1 \leqslant j \leqslant 2g$, *appears in* $\mathbf{W}$ *with exponent* $n \neq \pm 1$.

(3) *Suppose that some letter* s_j *appears in* W *with all exponents* $\neq \pm 1$. *By* (1) *above, we may without loss of generality assume that the generator which appears with exponent* $\neq \pm 1$ *is* s_1. *Let* $n_1, n_2, \ldots, n_k$ *be the set of exponents of* s_1 *in* v. *Then* $\mathbf{W}$ *is simple only if there exists an integer n such that*

$$\{|n_1|, |n_2|, \ldots, |n_k|\} = \{n, n+1\}.$$

(4) *Let* v *be a subword of* $\mathbf{W}$ *which after a permutation of generators as in* (1) *has the form*

$$v = s_1^{n_1} u_1 s_1^{n_2} u_2 \ldots s_1^{n_k} u_k,$$

where s_1 *does not appear in any of the subwords* u_j *and where* $|n_j| > 1$, $1 \leqslant j \leqslant k$. *Then* $\mathbf{W}$ *is simple only if*

(i) *the sequence of exponents* ± 1 *in each* u_j *is alternating;*

(ii) *the last letter of* u_j *has the same exponent as the first letter of* u_{j+1};

(iii) *if* n_j *and* n_{j+1} *have the same (respectively opposite) signs then* u_j *has odd (respectively even) length.*

REMARK. If W is finite, we may test for simplicity by applying the theorem to the biinfinite word $\mathbf{W} = \ldots WWW \ldots$. The example of the simple word $\bar{s}_1 s_3 \bar{s}_1 s_4 s_1^4 \bar{s}_4$ shows that the condition $|n_j| > 1$ for $1 \leqslant j \leqslant k$ in (3) and (4) is necessary.

Proof. (1) Note that the permutation ρ of the generators is an automorphism of $\Gamma(g, 1)$. This automorphism is induced by the isometry of $M(g, 1)$ which is induced by an anticlockwise rotation of $\mathbb{D}$ of $\pi + (2\pi/4g)$ about O. One sees similarly that τ is geometrically induced by rotation through π.

Our algorithm (Theorem B(2)) asserts that $\mathbf{W}$ is simple if and only if for every pair of cuts the endpoint pairs associated to one do not separate the endpoint pairs associated to the other. The criterion for deciding whether this is the case is the rule for cyclic lexicographical ordering.

(2) Suppose that $\mathbf{W}$ has the form $\ldots s_i^{2\varepsilon} \ldots s_j^{2\delta} \ldots$ for some pair $1 \leqslant i \neq j \leqslant 2g$, with $\varepsilon, \delta = \pm 1$, that is that two distinct letters s_i, s_j each appear in $\mathbf{W}$ with exponent $\neq \pm 1$. Introducing cuts $\ldots s_i^{\varepsilon} \mid s_i^{\varepsilon} \ldots s_j^{\delta} \mid s_j^{\delta} \ldots$ we obtain right and left words $s_i \ldots$, $\bar{s}_i \ldots$, $s_j \ldots$, $\bar{s}_j \ldots$. Assume without loss of generality that s_i precedes s_j in the alphabet A_{s_1}. Then since $s_i < s_j < \bar{s}_i < \bar{s}_j$ we see that $\mathbf{W}$ is not simple.

(4i) Suppose that s_1 appears with exponent $n \neq 1$ and that $\mathbf{W}$ also contains a subword $s_i^\varepsilon s_j^\varepsilon$, $\varepsilon = \pm 1$, $2 \leqslant i \neq j \leqslant 2g$. Introduce the cuts $\mathbf{W} = \dots s_1^{\pm 1} \mid s_1^{\pm 1} \dots s_i^\varepsilon \mid s_j^\varepsilon \dots$ (as will be seen it will not matter whether $s_i^\varepsilon s_j^\varepsilon$ precedes or follows s_1^2). Since $s_1, \bar{s}_1$ separate $s_j^\varepsilon, s_i^{-\varepsilon}$ for every admissible choice of i, j, ε it follows that $\mathbf{W}$ is not simple. This proves that the exponents in the words u_j of (4i) must alternate.

(4ii) Suppose that $\mathbf{W}$ contains a subword of the form $v = s_i^\varepsilon s_1^n s_j^{-\varepsilon}$, $\varepsilon = \pm 1, i \neq 1, j \neq 1$. Without loss of generality we may assume that $n > 1$ (if not, permute the generators so that each s_i is interchanged with $\bar{s}_i$). Choosing cuts $\mathbf{W} = \dots s_i^\varepsilon s_1 \mid s_1^{n-1} \mid s_j^{-\varepsilon} \dots$, we obtain right and left words $s_1^{n-1} \dots, \bar{s}_1 s_i^{-\varepsilon} \dots$ from the first cut and $s_j^{-\varepsilon} \dots, \bar{s}_1 \bar{s}_1 \dots$ from the second.

If $\varepsilon = 1$, cyclic lexicographical ordering gives $s_1^{n-1} < \bar{s}_1 \bar{s}_1 < \bar{s}_1 \bar{s}_i < \bar{s}_j$. If $\varepsilon = -1$ we obtain $s_1^{n-1} < s_j < \bar{s}_1 s_i < \bar{s}_1 \bar{s}_1$, so that $\mathbf{W}$ cannot be simple.

(3) Suppose that $\mathbf{W}$ contains subblocks $\dots s_i^\varepsilon s_1^n s_j^\varepsilon \dots$ and $\dots s_k^\varepsilon s_1^m s_q^\delta \dots$, with $\lvert |m| - |n| \rvert \geqslant 2$. We may without loss of generality assume that m and n are both positive, since this depends only on the orientation of arcs on the geodesic determined by $\mathbf{W}$, but the fact that two arcs do or do not intersect is independent of their orientation. We may also assume that $m > n$, by changing names if necessary. If $\varepsilon = 1$ cut at $\dots s_i s_1 \mid s_1^{n-1} s_j \dots$ to get the right and left words $s_1^{n-1} s_j \dots, \bar{s}_1 \bar{s}_i \dots$ and at $\dots s_1^2 \mid s_1^n \dots$ to get the right and left words $s_1^n \dots,\ \bar{s}_1 \bar{s}_1 \dots$. Cyclic lexicographical ordering gives $s_1^n < s_1^{n-1} s_j < \bar{s}_1^2 < \bar{s}_1 \bar{s}_i$ which is impossible for simple $\mathbf{W}$. If $\varepsilon = -1$ we choose our cuts at $\dots \bar{s}_i \bar{s}_1^{n-1} \mid s_1 \bar{s}_j \dots$ and $\dots s_1^n \mid s_1^2 \dots$. Now the ordering gives $s_1 \bar{s}_j < s_1 s_1 < \bar{s}_1^{n-1} s_i < \bar{s}_1^n$ (because in the alphabet $A_{\bar{s}_1}$ the letter $\bar{s}_j$ precedes s_1), and so again $\mathbf{W}$ is not simple.

(4iii) Suppose that $\mathbf{W} = \dots s_i^\varepsilon s_1^{n_j} s_u^\varepsilon \dots s_v^\delta s_1^{n_{j+1}} s_z^\delta \dots$, $\varepsilon, \delta = \pm 1$. We may without loss of generality assume that $|n_{j+1}| \geqslant |n_j|$ (if not, replace $\mathbf{W}$ by its 'inverse' as in (3)). Introducing the permutation $(s_1, \bar{s}_1) \dots (s_{2g}, \bar{s}_{2g})$ if necessary, we may further assume that $n_j \geqslant 2$. We claim that if n_j and n_{j+1} have the same sign, then $\delta = \varepsilon$, while if they have opposite signs then $\delta = -\varepsilon$. For, suppose that $n_{j+1} \geqslant n_j \geqslant 2$ and $\delta = -\varepsilon$. Choose cuts in $\mathbf{W}$

$$\dots s_i^\varepsilon s_1 \mid s_1^{n_j - 1} s_u^\varepsilon \dots s_v^{-\varepsilon} s_1 \mid s_1^{n_{j+1} - 1} s_z^{-\varepsilon} \dots .$$

If $\varepsilon = +1$, $n_{j+1} > n_j$ we have $s_1^{n_{j+1}-1} < s_1^{n_j - 1} s_u < \bar{s}_1 s_v < \bar{s}_1 \bar{s}_i$, so that $\mathbf{W}$ is not simple. If $\varepsilon = +1$, and $n_{j+1} = n_j$, we have $s_1^{n_j - 1} \bar{s}_z < s_1^{n_j - 1} s_u < \bar{s}_1 s_v < \bar{s}_1 \bar{s}_i$ and again $\mathbf{W}$ is not simple. If $\varepsilon = -1$, $n_{j+1} > n_j$, we have $s_1^{n_j - 1} \bar{s}_u < s_1^{n_{j+1} - 1} < \bar{s}_1 s_i < \bar{s}_1 \bar{s}_v$, whereas if $n_{j+1} = n_j$ we have $s_1^{n_j - 1} \bar{s}_u < s_1^{n_j - 1} s_z < \bar{s}_1 s_i < \bar{s}_1 \bar{s}_v$. The case when n_j, n_{j+1} have opposite signs is similar. Since the sequence of exponents which occur in each u_j alternate, by rule (4i), the assertion in (4iii) follows.

COROLLARY 5.3. *Let W be a simple word in Γ in which some letter occurs with all exponents $n \neq \pm 1$. Then there is a canonical automorphism of Γ which strictly reduces the length of W.*

Proof. By Theorem C we may assume that W has the form $v = s_1^{n_1} u_1 s_1^{n_2} u_2 \dots s_1^{n_k} u_k$ where the u_i satisfy conditions (4i) to (4iii). Apply the automorphism ϕ of Γ defined by $\phi(s_1) = s_1$, $\phi(s_j) = s_1^{\pm n} s_j$.

The rules (4) will be seen to be exactly what is needed in order that $\phi(W)$ be shorter than W.

REMARK 5.4. Notice that the rules in Theorem C and the reduction above depend crucially on our choice of generators for Γ, although the algorithm itself works quite generally as explained at the beginning of §2.

REMARK 5.5. Using Corollary 5.3 one can simplify the problem of describing simple words in $\Gamma(g, 1)$ by a reduction analogous to the procedure outlined in Remark 5.2. However, unlike the situation in $\Gamma(1, 1)$, the reduction process is not complete, the principle reason being that there are in general infinitely many non-trivial homotopy classes of separating simple curves on $M(g, 1)$, $g \geqslant 2$, but only one on $M(1, 1)$.

References

1. J. S. BIRMAN and C. SERIES, 'Simple curves have Hausdorff dimension 1', preprint, Warwick University, 1983.
2. D. CHILLINGWORTH, 'Simple closed curves on surfaces', *Bull. London Math. Soc.*, 1 (1969), 310–314.
3. D. CHILLINGWORTH, 'Winding numbers on surfaces I', *Math. Ann.*, 196 (1972), 218–249.
4. D. CHILLINGWORTH, 'Winding numbers on surfaces II', *Math. Ann.*, 199 (1972), 131–153.
5. M. COHEN, W. METZLER and A. ZIMMERMAN, 'What does a basis of $F(a, b)$ look like?', *Math. Ann.*, 257 (1981), 435–445.
6. J. NIELSEN, 'Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen', *Acta Math.*, 50 (1927), 189–358.
7. H. POINCARÉ, 'Analysis situs', *J. Ecole Polytechn.* (2), 1 (1895), 1–121.
8. C. SERIES, 'The infinite word problem and limit sets in Fuchsian groups', *Ergod. Th. and Dyn. Sys.*, 1 (1981), 337–360.
9. H. ZIESCHANG, 'Algorithmen für einfache Kurven auf Flächen', *Math. Scand.*, 17 (1965), 17–40; 25 (1969), 49–58.

Department of Mathematics,
Columbia University,
(Barnard College),
New York 10027,
U.S.A.

Mathematics Institute,
University of Warwick,
Coventry CV4 7AL.